ISSN : 2165-4069(Online)
ISSN : 2165-4050(Print)

IJARAI

International Journal of Advanced Research in Artificial Intelligence

Volume 1 Issue 8

www.ijarai.thesai.org

A Publication of
The Science and Information Organization

IJARAI

INTERNATIONAL JOURNAL OF
ADVANCED RESEARCH IN ARTIFICIAL INTELLIGENCE

Editorial Preface

From the Desk of Managing Editor…

"The question of whether computers can think is like the question of whether submarines can swim." ― Edsger W. Dijkstra, the quote explains the power of Artificial Intelligence in computers with the changing landscape. The renaissance stimulated by the field of Artificial Intelligence is generating multiple formats and channels of creativity and innovation.

This journal is a special track on Artificial Intelligence by The Science and Information Organization and aims to be a leading forum for engineers, researchers and practitioners throughout the world.

The journal reports results achieved; proposals for new ways of looking at AI problems and include demonstrations of effectiveness. Papers describing existing technologies or algorithms integrating multiple systems are welcomed. IJARAI also invites papers on real life applications, which should describe the current scenarios, proposed solution, emphasize its novelty, and present an in-depth evaluation of the AI techniques being exploited. IJARAI focusses on quality and relevance in its publications.

In addition, IJARAI recognizes the importance of international influences on Artificial Intelligence and seeks international input in all aspects of the journal, including content, authorship of papers, readership, paper reviewers, and Editorial Board membership.

The success of authors and the journal is interdependent. While the Journal is in its initial phase, it is not only the Editor whose work is crucial to producing the journal. The editorial board members , the peer reviewers, scholars around the world who assess submissions, students, and institutions who generously give their expertise in factors small and large— their constant encouragement has helped a lot in the progress of the journal and shall help in future to earn credibility amongst all the reader members.

I add a personal thanks to the whole team that has catalysed so much, and I wish everyone who has been connected with the Journal the very best for the future.

Thank you for Sharing Wisdom!

Managing Editor
IJARAI
Volume 1 Issue 8 November 2012
ISSN: 2165-4069(Online)
ISSN: 2165-4050(Print)

Editorial Board

Reviewer Board Members

CONTENTS

The Research of the Relationship between Perceived Stress Level and Times of Vibration of Vocal Folds

Yin Zhigang
Phonetics Laboratory, Institute of Linguistics,
Chinese Academy of Social Sciences
Beijing, China

***Abstract*—Whether a syllable is perceived as stressed or not and whether the stress is strong or weak are hot issues in speech prosody research and speech recognition. A focus of the stress study is on the investigation of the acoustic factors which contribute to the perception of stress level. This study examined all possible acoustic/physiological cues to stress based on data from Annotated Chinese Speech Corpus and proposed that times of vibration of vocal folds (TVVF) reflects stress level best.**

It is traditionally held that pitch and duration are the most important acoustic parameters to stress. But for Chinese which is a tone language and features special strong-weak pattern in prosody, these two parameters might not be the best ones to represent stress degree. This paper proposed that TVVF, reflected as the number of wave pulses of the vocalic part of a syllable, is the ideal parameter to stress level. Since number of pulses is the integral of pitch and duration (Pulse=∫f(pitch)dt), TVVF can embody the effect of stress on both pitch and duration. The analyses revealed that TVVF is most correlated with the grades of stress. Therefore, it can be a more effective parameter indicating stress level.

Keywords- Times of vibration of vocal folds (TVVF); stress level; parameter.

I. INTRODUCTION

An important aspect of the study of natural language processing is the research of speech stress—for natural language generation, the stress manipulation of synthetic speech units will affect the result of speech synthesis; for natural language understanding, stress-location can affect the comprehensions of sentences or some words. For example, in English, *[re'cord]* is a verb but *['record]* denotes a noun. In some of the other languages, such as Chinese, stress plays a more important role in speech understanding.

But what is stress? What factors contribute to the perception of stress levels? The past research shows that stress is the psychological reflection of the prominence of prosodic units and this reflection is influenced by the objective acoustic factors [1]. These factors include pitch (f0), duration, intensity, lexical tones, and prosodic location. Of these factors, some factors are more correlated with stress than others. [2]

In order to estimate stress levels of syllables automatically, the relationship between perceived stress level and attainable acoustic factors should be analyzed. In this study, we investigated all possible acoustic/physiological parameters to stress levels of syllables in order to find out the one that is most correlated. To a certain degree, it can be an effective indicator to stress grade.

II. POSSIBLE PARAMETERS TO THE STRESS LEVEL

A. Duration and Pitch

According to traditional theory, pitch and duration are most closely related to stress level.

Generally, the prosodic unit which has stress will have raised upper limit of pitch range (PitchMax), enlarged pitch range [3] and lengthened duration [4]. More specifically, the phonetic changes caused by stress can be explained as the intensification of articulation. As shown in figure1, this effect will lengthen the duration and make prominent the pitch features by raising pitch when it has [+high] feature and lowering pitch when it has [+low] feature.

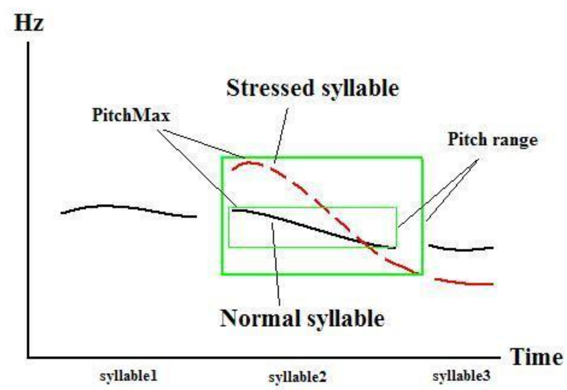


Figure 1. The pitch contours of a normal syllable and a stressed syllable

The above theory suggests that the pattern of pitch change of a stressed syllable will be influenced by its tone type in a tone language. It could be illuminated by the example of Mandarin Chinese. Mandarin is a tone language and the figure2 shows its 4 basic tones—Tone 1(HH), Tone 2(LH), Tone 3(LL) and Tone 4(HL). Of these tones, Tone 1(HH), Tone 2(LH) and Tone 4(HL) all have [+high] feature. When the syllables with these tones are stressed, the [+high] feature will be amplified, with the raised upper limit of pitch range and the enlarged pitch range. The only tone in Mandarin with all [+low] feature is Tone 3(LL). So when the syllables with Tone 3 are stressed, the lower limit of pitch range will be lowered to enlarge pitch range.

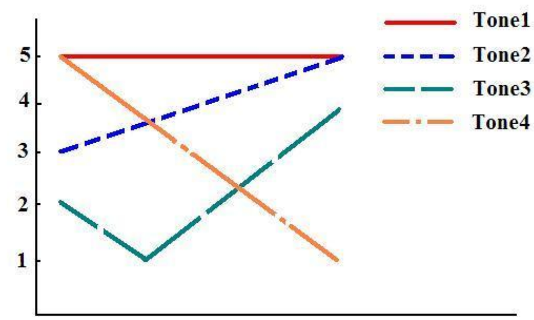


Figure 2. Diagrammatic sketch of Mandarin tones

As for pitch and duration, which one influences the perception of stress most in Mandarin? The issue is controversial. One school of researchers holds that duration contributes most to the perception of stress [5]. The other school thinks that pitch is the acoustic cue to stress. One study [6] found that pitch is most correlated with stress in disyllabic words. Cao [7] studied the phonetic realization of focus stress. His results showed that the most important acoustic cue is the change of pitch register. The change of duration is the secondary cue to stress. Cai [8] proposed an equation to simulate the perception of stress:

$$L=1.5F(pitch)+0.95(duration)+0.65R(pitchRange) \quad (1)$$

The coefficients implied that pitch is the most influential factor. Other studies demonstrated complimentary relationship and interaction effect between them [9]. Even with complimentary relationship, pitch is more important than duration.

B. *Limitation of regarding pitch and duration as acoustic parameters to stress level*

Although pitch-related acoustic parameters (like pitch range and upper limit of pitch range) and duration are important parameters to stress level, they still have many limitations:

1） The parameter like upper limit of pitch range (pitchMax) can not reflect the perception of stress of Tone 3 in Mandarin. The feature of Tone 3(LL) is [+low]. The stress will lower the lower limit of pitch range. However, the effect of stress for the other tones in Mandarin is to raise the upper limit of pitch range.

2） Another limitation of taking pitchMax as the acoustic cue to stress is that the declination of pitch range happening in the large prosodic unit will make the pitch of the syllables in latter position lower than that of the previous syllables. The declination phenomena of pitch curve could be found in figure3.This in turn will weaken the effect of stress. That is, the higher pitch of the previous syllables doesn't mean these syllables are more stressed (local relative pitch differences are more important).

3） The limitation of pitch range is that it cannot represent the range change of Tone 1(HH) under stress, for the pitch range of Tone 1 is very narrow by nature. The change of pitch range is hard to be detected.

4） The limitation of duration as the cue to stress is that the lengthening of syllables can be due not only to the effect of stress but also the effect of prosodic boundaries [3]. If a syllable is at a prosodic boundary, it will be lengthened, but this lengthening doesn't mean the syllable is stressed. The lengthening phenomena before a boundary could also be found in figure3 (the last syllable).

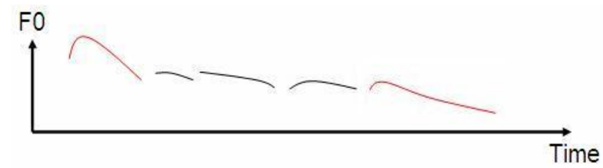


Figure 3. The pitch curve of a sentence

In general, pitch and duration alone cannot effectively cue stress in Mandarin. A new cue which can fully embody stress is needed.

C. *A new parameter to stress level–Times of vibration of vocal folds (TVVF)*

In this research, *Times of vibration of vocal folds* (TVVF) is regarded as an effective cue to stress level. TVVF is a physiological parameter. The times of vibration of vocal folds of a syllable is equal to the number of wave pulses of the voiced part of a syllable (usually, this part is vowel).

The figure below (based on one syllable) shows the relationship among TVVF, pitch and duration on syllable final.

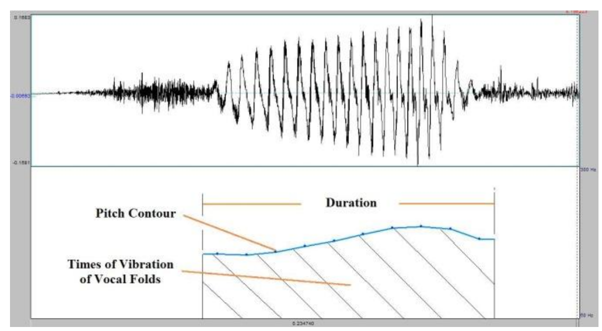


Figure 4. The relationship among TVVF, pitch and duration

In the figure, the blue curve means the pitch contour of a syllable. It could be regarded as a function curve of independent variable (*time*). Because TVVF means the total number of pitch periods in the *duration* of the syllable, it could be represented as the integral of *pitch* and *time* (*duration)* value of the syllable in mathematics. In figure4, the area (signified by diagonal lines) under the blue curve of f0 means TVVF. The relationship among TVVF, *pitch* and *time* could be expressed as the formula:

$$P = \int f(pitch)dt \quad (2)$$

In this formula, P means TVVF（*the times of vibration of vocal folds*）. t means *time* and its maximum is equal to the *duration* value of the syllable. Therefore, the times of vibration of vocal folds (TVVF) is the integral of pitch and duration value of the syllable.

The times of vibration of vocal folds (TVVF) is related to both pitch and duration.

1) The higher or lower of pitch will increase or decrease the times of vibration of vocal folds.

2) The lengthening or shortening of duration will also increase or decrease the times of vibration of vocal folds.

3) Although the TVVF (times of vibration of vocal folds) is a physiological cue, it has an acoustic counterpart, i.e. number of wave pulses. It is convenient to measure.

4) The limitation of using TVVF as a parameter to stress is that like pitch, it refers to the voiced part of the syllable. So TVVF can be used for the vocalic part. It can't be used to analyze the consonant part. But some research [10] demonstrated that the lengthening of the syllable is mainly due to the lengthening of the vowel. Therefore, the effect of stress can be embodied with the lengthening of vowel alone.

III. ANALYSES

According to the above discussion, theoretically, TVVF should be more correlated with perceived stress level. The following analyses will test the assumption by comparing the correlation coefficients between stress levels and all acoustic parameters.

A. Corpus

The data analyzed in this study are from ASCCD Annotated Speech Corpus of Chinese Discourse, made by the phonetic laboratory of the institute of linguistics of CASS (Chinese Academy of Social Sciences). The corpus is made of the recordings of 18 passages of different styles (about 10,000 syllables). These passages were read by 10 Mandarin speakers. In this paper, we only analyzed the data of 2 female speakers (F001, F002) and 2 male speakers (M001, M002). The styles of reading between the 4 speakers are very different.

The recordings were made in stereo, with a sampling frequency of 16 kHz and 16 bits per sample resolution. The corpus is about 1.5GB. All the recordings were annotated by manual and the consistent rate of 6 annotators was about 87.5%. The annotations include orthography, syllable initials, syllable finals, prosodic boundaries and grades of stress. The stress is indexed as 0, 1, 2, 3. 0 means weak, 1 means normal, 2 means strong and 3 means very strong. All the labels of stress levels were annotated based on perception of stress without consulting any acoustic parameters.

The following analyses were made to examine the correlation between the acoustic parameters and stress.

The data analyzed in this part were not normalized. In the following tables showing results, the figures marked by ** mean the correlation is significant at the level of 0.01 (two-tailed).

B. Correlation between TVVF and stress

Correlation analysis was conducted between TVVF and the grades of stress for F001, F002, M001 and M002. The results showed that the correlation coefficients were 0.50, 0.56, 0.55 and 0.58 (significant at the level of 0.01). The mean of 4 correlation coefficients was 0.55.

TABLE I. CORRELATION BETWEEN TVVF AND STRESS LEVEL

Correlation between TVVF and stress level					
		F001	**F002**	**M001**	**M002**
stress level	Pearson Correlation	0.50**	0.56**	0.55**	0.58**
	Sig. (2-tailed)	0.00	0.00	0.00	0.00
	N	8762	8762	8759	8760

C. Correlation between upper limit of pitch range (pitchMax) and stress

The upper limit of pitch range (pitchMax) was the maximum pitch of each syllable.

The following table showed that with all data included (including all the lexical tones, Tone 1, Tone 2, Tone 3, Tone 4 and neutral tone), the 4 speakers' correlation coefficients between pitchMax and stress level are 0.37, 0.45, 0.48 and 0.51 (significant at the level of 0.01), which were both lower than those between TVVF and stress. The mean of this kind of correlation coefficients was about 0.45.

TABLE II. CORRELATION BETWEEN UPPER LIMIT OF PITCH RANGE AND STRESS LEVEL (INCLUDING ALL TONES)

Correlation between pitchMax and stress level (including all tones)					
		F001	**F002**	**M001**	**M002**
stress level	Pearson Correlation	0.37**	0.45**	0.48**	0.51**
	Sig. (2-tailed)	0.00	0.00	0.00	0.00
	N	8762	8762	8759	8760

Since the effect of stress on the syllable with Tone 3 is different from the syllables with other tones (for syllables with

Tone 3, being stressed will enlarge the pitch range by lowering the lower limit of pitch range), the following correlation is conducted when syllables with Tone 3 were excluded.

TABLE III. CORRELATION BETWEEN PITCHMAX AND STRESS LEVEL (EXCLUDING SYLLABLES WITH TONE 3)

Correlation between pitchMax and stress level (excluding Tone 3)					
		F001	**F002**	**M001**	**M002**
stress level	Pearson Correlation	0.43**	0.52**	0.55**	0.56**
	Sig. (2-tailed)	0.00	0.00	0.00	0.00
	N	7273	7274	7268	7270

The results showed that the coefficients of all speakers were higher in different degree and the mean of them was about 0.52. This means that the effect of stress imposed on syllables with Tone 3 is different from those with other tones. Excluding syllables with Tone 3, the correlation between pitchMax and stress is stronger and approaches to the level between TVVF and stress.

D. Correlation between pitch range and stress

The pitch range is the difference between pitch maximum and minimum.

TABLE IV. CORRELATION BETWEEN PITCH RANGE AND STRESS LEVEL (INCLUDING ALL TONES)

Correlation between pitch range and stress level (including all tones)					
		F001	**F002**	**M001**	**M002**
stress level	Pearson Correlation	0.09**	0.25**	0.39**	0.41**
	Sig. (2-tailed)	0.00	0.00	0.00	0.00
	N	8762	8762	8759	8760

The results showed that with all tones included, the correlation coefficients for 4 speakers are 0.09, 0.25, 0.39, 0.41 (significant at the level of 0.01) and the mean of them is 0.29.

The results showed that the correlation between pitch range and stress is not as strong as indicated by the traditional sentential stress theory. To further examine the data, another analysis is conducted between stress and pitch range based on ST (Semitone) scale. f0 is changed into ST through the formula St=12*log2(F0 / F(reference)). The correlation based on St revealed stronger relation, 0.15 for F001, 0.30 for F002, 0.32 for M001, and 0.43 for M002 (significant at the level of 0.01). But still it is not strong enough.

The reason for the weak correlation between pitch range and stress might be due to Tone 1. Tone 1 is with very narrow pitch range by nature. So even the syllable with Tone 1 is stressed, its pitch range will not be enlarged, which in turn weakens the correlation between pitch range and stress. With Tone 1 excluded, the correlation coefficients are 0.12 for F001, 0.35 for F002, 0.50 for M001, and 0.55 for M002. (table V)

TABLE V. CORRELATION BETWEEN PITCH RANGE AND STRESS LEVEL (EXCLUDING TONE 1)

Correlation between pitch range and stress level (excluding Tone 1)					
		F001	**F002**	**M001**	**M002**
stress level	Pearson Correlation	0.12**	0.35**	0.50**	0.55**
	Sig. (2-tailed)	0.00	0.00	0.00	0.00
	N	7094	7096	7088	7092

The results of the above two correlation analyses (based on Hz and ST) are not consistent with the strong correlation between pitch range and stress proposed by the traditional sentential stress theory. The reasons might be:

1. The natural speech has the feature of pitch declination (at the same time with narrowed pitch range). The past research usually made normalization on this effect. All the data analyzed in this study were raw data without normalization.

2. The claim of the strong correlation between pitch range and stress is largely based on the local comparison in some adjoining prosodic units. However, in this study, all the data were compared in the overall instead of focusing on particular type of prosodic units.

3. According to traditional stress theory, the widening effect of stress on pitch range is with reference to larger prosodic units rather than the small unit like syllable.

The reasons mentioned above can explain why the correlation in this study is weaker than that proposed by the traditional theory. At the same time, these reasons also imply that using pitch range as cue to stress is not effective.

E. *Correlation between duration and stress*

The results are shown in the following table.

As shown in the table, the coefficients for F001, F002, M001, M002 were 0.46, 0.50, 0.54 and 0.57 (significant at the level of 0.01), which were good results but still lower than those between TVVF and stress.

TABLE VI. CORRELATION BETWEEN DURATION AND STRESS

Correlation between duration and stress level					
		F001	**F002**	**M001**	**M002**
stress level	Pearson Correlation	0.46**	0.53**	0.54**	0.57**
	Sig. (2-tailed)	0.00	0.00	0.00	0.00
	N	8762	8762	8759	8760

F. *Correlation between intensity and stress*

For speech, the intensity isn't generally acknowledged a dominant parameter to stress, compared to pitch and duration.

TABLE VII. CORRELATION BETWEEN INTENSITY AND STRESS

Correlation between intensity(Max) and stress level					
		F001	**F002**	**M001**	**M002**
stress level	Pearson Correlation	0.06**	0.05**	0.05**	0.08**
	Sig. (2-tailed)	0.00	0.00	0.00	0.00
	N	8762	8762	8759	8760

The correlation between maximum of intensity and stress were analyzed. In general, intensity were not strongly correlated with stress. For F001, F002, M001and M002, the correlation coefficients between stress grades and intensities were 0.06, 0.05, 0.05 and 0.08. The reason why intensity was only weakly correlated with stress was that it was hard to measure intensity under the standard condition and intensity will be influenced easily by factors other than acoustic ones. For example, the distance from the mouth to the microphone and the physical and emotionally status of the speaker will cause the change of intensity.

G. *Correlation between energy and stress*

TABLE VIII. CORRELATION BETWEEN ENERGY AND STRESS LEVEL

Correlation between energy and stress level					
		F001	**F002**	**M001**	**M002**
stress level	Pearson Correlation	0.05**	0.04**	0.02**	0.06**
	Sig. (2-tailed)	0.00	0.00	0.00	0.00
	N	8762	8762	8759	8760

As shown in the table VIII, the correlation coefficients for 4 speakers were 0.005, 0.04, 0.02 and 0.06 (significant at the level of 0.01). These coefficients meant that energy was only weakly correlated with stress. The reason for the weak correlation was that like intensity, energy will be affected by other factors (recording volume, distance, etc.) rather than acoustic ones. Therefore, it is not reliable to take energy as the parameter to stress.

IV. CONCLUSION AND DISCUSSION

Taken all the analyses together, based on the correlation coefficients derived, the acoustic parameters mentioned above are ranked as follows:

TABLE IX. THE RANKS OF THE ACOUSTIC PARAMETERS IN ACCORDANCE WITH THE CORRELATION COEFFICIENTS

Correlation ranks (stress level)	**F001**	**F002**	**M001**	**M002**	**Mean**
TVVF	0.50	0.56	0.55	0.58	**0.55**
Duartion	0.46	0.53	0.54	0.57	**0.53**
pitchMax (excluding Tone3)	0.43	0.52	0.55	0.56	**0.52**
pitchMax(all tones)	0.37	0.45	0.48	0.51	**0.45**
pitch range (excluding Tone 1)	0.12	0.35	0.50	0.55	**0.38**
pitch range (ST scale, all tones)	0.15	0.30	0.30	0.43	**0.30**
pitch range (all tones)	0.09	0.25	0.39	0.41	**0.29**
intensity	0.06	0.05	0.05	0.08	**0.06**
energy	0.05	0.04	0.02	0.06	**0.04**

As shown in the table, TVVF is most correlated with stress level while duration, upper limit of pitch range and pitch range are less correlated. The weakest correlation is between intensity and energy, and stress.

Though TVVF is not that strongly correlated with stress as indicated by the coefficient (less than 0.6), considering all data analyzed are not normalized, TVVF can be taken as a good parameter to stress.

Normalizing data can make the correlation stronger. Generally, the effects of types of syllable initials, syllable finals and prosodic location should be normalized. For a tone language, the type of syllable tones is another influencing factor and should be normalized too. Table X gives the correlation coefficients between stress level and 4 normalized global acoustic parameters-*TVVF, duration, pitchMax(all tones) and pitch range(all tones)*.

TABLE X. THE CORRELATION COEFFICIENTS OF NORMALIZED DATA

Correlation ranks (stress level)	**Mean (original data)**	**Mean (normalized data)**
TVVF	0.55	0.62
Duartion	0.53	0.61
pitchMax(all tones)	0.45	0.54
pitch range (all tones)	0.29	0.37

The new result shows that the normalization processing could raise the correlation coefficients between stress level and the acoustic data, and TVVF was still the best parameter to stress.

All in all, according to the analyses of this study, TVVF is most correlated with perceived stress level. That means the stronger the stress is, the more of TVVF. It can be concluded that times of vibration of vocal folds (TVVF) is a sensitive parameter to stress in Mandarin Chinese.

It should be pointed out that the study is based on a Mandarin corpus but the same conclusion applies for other languages in theory. In the future, more work will be carried out with other language corpus to check the validity of the conclusion.

ACKNOWLEDGMENT

This work has been supported by Phonetic Laboratory, Institute of Linguistics, Chinese Academy of Social Sciences.

REFERENCES

[1] Ye jun, “The Research of Rhythm of Modern Chinese”, Shanghai Century Press, 40-49, 2008

[2] Yin Zhigang, “The Research of Rhythm of Standard Chinese”, 61-71, PhD Academic Dissertation of CASS,2011

[3] Cao Jianfen, The relationship of Tone and Intonation, China Language, (3) 195-202, 2002

[4] Lin Maocan, Yan Jingzhu, Sun Guohua, The Stress Pattern of Foot in Beijing Dialect, Dialect (1), 1984

[5] Zhong Xiaobo, et al., The Stress Perception of Standard Chinese, Psychological Sinica, (6)vol31, 2001

[6] Wang Yunjia, Chu Min, He Lin, A preliminary study of Focus Stress and semantic Stress in Chinese, Chinese Courses for Foreigners, (2)86-98,2006

[7] Cao Wen, The Prosody of Chinese Focus, 5-10, Beijing Language Press, 2010

[8] Cai Lianhong, Wu Zongji, Tao Jianhua, The Computability Research of Chinese Prosody Characters, The Modern Phonetics of New Century, Tsinghua University Press, 2001

[9] Xu Jieping, Chuming, He Lin, Lv Shinan, The Infection of Utterance Stress to Pitch and Duration, Acoustics Journal, Vol25(4),2000

[10] Mei Xiao, Analysis of Duration of Mandarin Prosodic Structures, Journal of Chinese Information Processing, Vol.24, No.4, Jul 2010

AUTHORS PROFILE

PhD Yin Zhigang is an assistant professor of Phonetic Laboratory, Institute of Linguistics, Chinese Academy of Social Sciences. His research interests include Phonetics, Phonology, Speech Database, Multi-modal, EEG/ERP, Eye-tracking research and Semantics.

Clustering Method Based on Messy Genetic Algorithm: GA for Remote Sensing Satellite Image Classifications

Kohei Arai [1]
Graduate School of Science and Engineering
Saga University
Saga City, Japan

***Abstract*— Clustering method for remote sensing satellite image classification based on Messy Genetic Algorithm: GA is proposed. Through simulation study and experiments with real remote sensing satellite images, the proposed method is validated in comparison to the conventional simple GA. It is also found that the proposed clustering method is useful for preprocessing of the classifications.**

Keywords- clustering; classification; Genetic Algorityhm: GA; Messy GA; simple GA.

I. INTRODUCTION

Although there are conventional clustering methods such as k-means, ISODATA, etc., these cannot ensure the optimum clustering results at all [1]. Also Genetic Algorithm: GA based clustering method is proposed as an optimum combination problem solving method. It still cannot ensure the optimum clustering result [2]. GA based clustering uses probabilistic searching method for optimum combination through learning processes. The conventional GA based clustering is referred to "Simple GA". Simple GA has problem such as relatively long schema used to be broken, fitness of schema may change in the defined chromosome in GA algorithm. Therefore, simple GA cannot ensure optimum clustering result at all.

In order to overcome the problems of simple GA based clustering method, Messy GA based clustering method is proposed in the paper. Messy GA utilize "Codon" which is defined as variable length of list structure of chromosome representation [3]. Therefore, relatively long schema used to be maintained and fitness of schema may not been changed in the defined chromosome in GA algorithm. Convergence processes are discussed [4]. Also Modified ISODATA clustering is proposed for acceleration of convergence processes [5]. GA based ISODATA is also proposed for improvement of clustering performance [6]. On the other hands, online clustering utilizing learning automata and pursuit reinforcement competitive learning is proposed [7],[8].

The following section describes the proposed Messy GA based clustering method followed by simulation study and some experimental study with remote sensing satellite images. Then conclusion is described together with some discussions.

II. PROPOSED METHOD

A. GA Clustering

Genetic Algorithm: GA based clustering is defined as follows,

(1) The image in concern is defined as two dimensional array of pixels which is shown in Figure 1.

(2) Chromosome is defined as a pair of pixel number and cluster number.

(3) Fitness function is defined as between cluster variance

(4) Cross over, mutation, and selection of the chromosome is repeated until the finish condition is satisfied though the processes that higher fitness function of chromosome is remained (This is the typical GA processes)

(5) Thus all the pixels are assigned the most appropriate cluster number (cluster results)

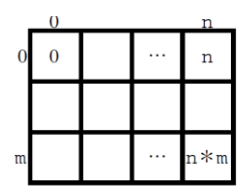
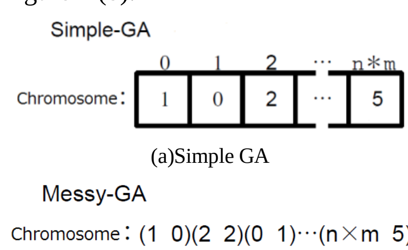


Figure 1 Definition of image (two dimensional array of pixels) in concern for clustering

In this process, chromosome is illustrated in Figure 2 (a) which is referred to Simple GA while Messy GA chromosome is shown in Figure 2 (b).



(a)Simple GA

(b)Messy GA

Figure 2 Chromosome definitions for Simple and Messy GA

Messy GA is composed with the following three phases,

(1) Initial phase:

All the schema of which chromosome with optimum scheme is generated referencing to the maximum schema length

(2) Primordial phase:

Selection of chromosome until the cross over process can be applied to chromosome through the following processes, 1) the maximum length of schema, k is determined, 2) all the possible schema is generated (The number of possible combination of schema is ${}_kC_m 2^m$), 3) through fitness function of chromosome is evaluated, then 4) selection is made by tournament selection method

(3) Juxtapositional phase:

Choose the arbitrary pair of chromosome for cut off (Cut) and connect (Splice). The probability of the cut depends on the chromosome length. Meanwhile, splice is made independently the chromosome length. Therefore, splice is used to be made for the shorter chromosome because the probability of splice for shorter chromosome is less than that for long chromosome.

Figure 3 shows the process flow of the Messy GA.

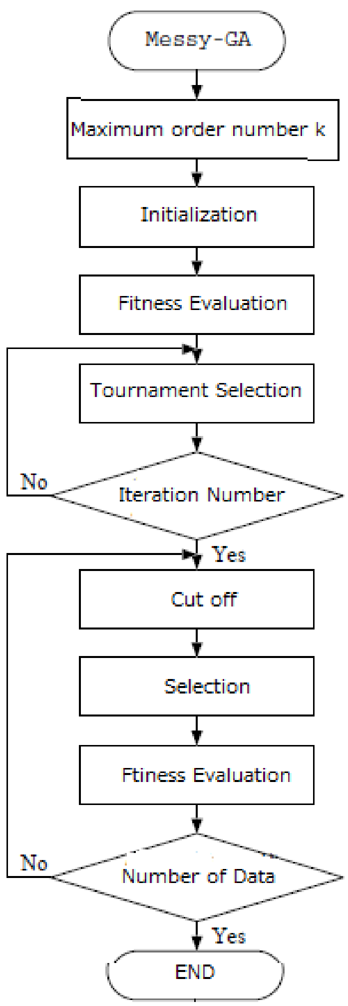


Figure 3 Process flow of Messy GA

B. *Simple GA Clustering*

The process of simple GA clustering is as follows,

(1) Pair of pixel number (ranges from 0 to n-1) and cluster number (ranges from 0 to k-1)is set as an initial stage

(2) Fitness function is evaluated with the between cluster variance

(3) Expectation strategy with uniformly distributed random number is used together with elite preserve strategy for selection of chromosome which has high score of the fitness function

(4) Multi point cross over, then applied to the previously preserved chromosome and the current chromosome with cross over probability

(5) The cross over processed chromosome is preserved only if the fitness function is greater than the previous before cross over is applied

(6) Mutation is applied with uniformly distributed random number with mutation probability

(7) Finish condition is set at the number of GA process is exceeded at 3000.

C. *Messy GA Clustering*

The schema length of the Simple GA is fixed. Therefore, relatively long schema which is effective for cross over is used to be broken. Consequently, it is difficult to find the most appropriate solution of chromosome. Meanwhile, cross over is much effective for Messy GA due to the fact that all the possible chromosome of maximum length can be prepared because the chromosome length is variable together with list structural representation of chromosome.

(1) Coding of chromosome, then initial pair of pixels number and cluster number is set

(2) Fitness function evaluation

(3) Initialization

(4) Primordial phase

(5) Juxtapositional phase

When the iteration number and the data number is exceed the threshold, all the pixels are assigned to cluster number.

III. EXPERIMENTS

A. *Simulation Results*

Simulation study is conducted with the following parameters, The Initial chromosome assignment number=50, Cross over probability=0.75, Finalized generation number (iteration number)=3000, Mutation probability=0.03 (this is only for Simple GA), Standard deviation of cluster=0.04, Between cluster variance (distance between clusters)=0.16, The number of data=100, The number of bands of the data=2, The number of cluster=2.

Figure 4 shows generated simulation dataset of band 1 and 2. There are two clusters, cluster # 1 (left) and # 2 (right). Figure 5 shows the clustering results from the Simple GA clustering while Figure 6 shows the clustering results from the Messy GA clustering.

(a)Band 1

(b)Band 2

Figure 4 Simulation dataset used

(a)Simulation dataset #1

(b)Simulation dataset #2

Figure 5 Clustering results from the Simple GA clustering

(a)Simulation dataset #1

(b)Simulation dataset #2

Figure 6 Clustering results from Messy GA clustering

Such this simulation is repeated for 900 times. Then the clustering performance is evaluated. The convergence iteration number, between cluster variance, and percent correct clustering is evaluated as cluster performance.

These are shown in Table 1.

TABLE I. CLUSTERING PERFORMANCE OF THE SIMULATION STUDY

	Converged Iteration Number	Between Cluster Variance	Percent Correct Cluster
Simple GA	2866	171330	0.89
Messy GA	780	230947	0.99

All the evaluation items for Messy GA clustering is superior to those for Simple GA clustering. Significance test is also conducted for the clustering performance with 5% of significance level.

With the number of samples of 900, confidence interval at the confidence level of 95% is evaluated. The results show that the clustering results in Table 1 is significant.

Between cluster variance of fitness function is also evaluated as a function of iteration number. Figure 7 shows an example of the results from the between cluster variance evaluation.

Messy GA clustering (Dotted line in the Figure 7) converged faster than Simple GA clustering (Straight line in the Figure 7).

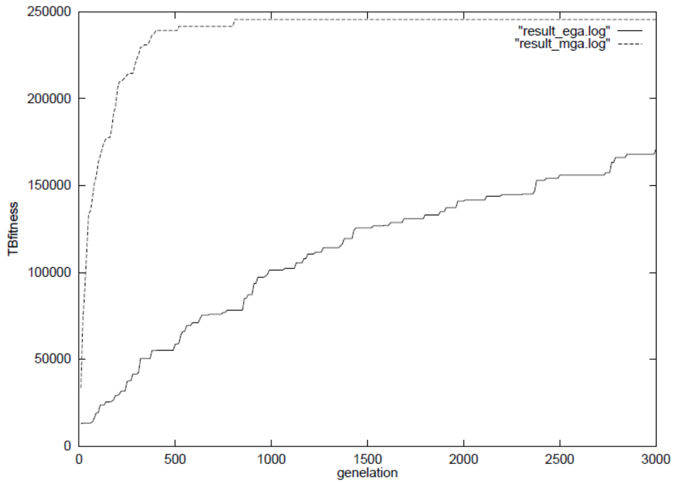
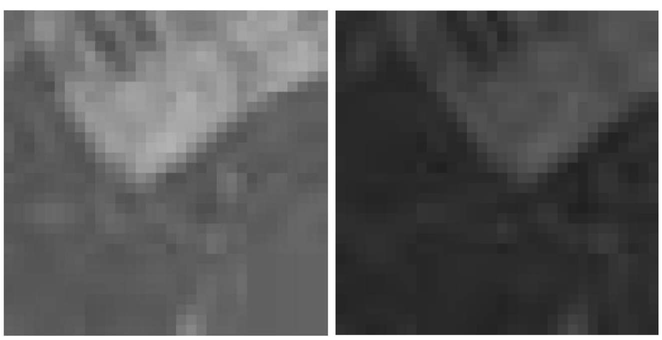


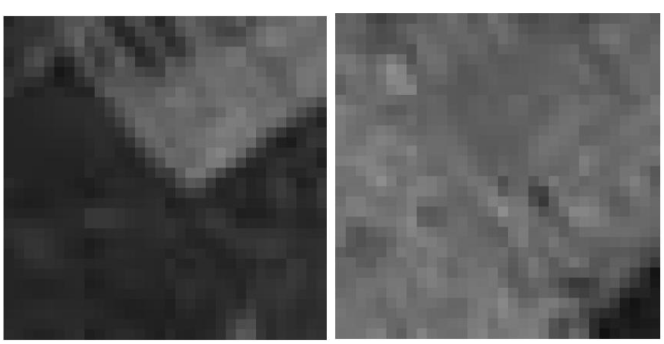
Figure 7 Example of the results from the between cluster variance evaluation.

B. Experiemntal study

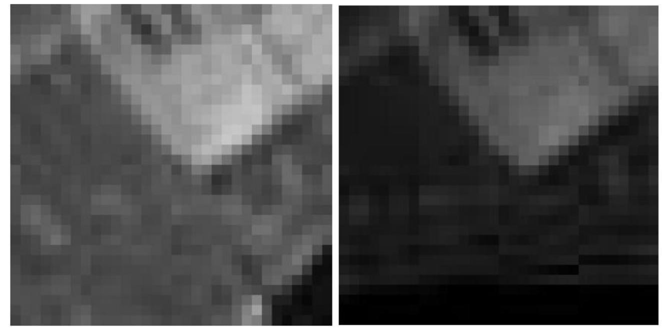
Small portion of LANDSAT-5/TM data (32 x 32 pixels) of the south-west of Saga city which is situated in the northern Kyushu, Japan which is acquired on May 28 1986 is used for the experiments. TM data consists of 6 bands, blue, green, red, near-infrared and two shortwave infrared bands with 30 m of Instantaneous Field of View: IFOV. Other than these, there is thermal infrared band with 120 m of IFOV. The thermal infrared band data is not used for the experiments. Figure 8 shows the images for the aforementioned 6 bands.



(a)Band 1 (b)Band 2



(c)Band 3 (d)Band 4



(e)Band 5 (f)Band 7

Figure 8 LANDSAT-5/TM Images of Saga city, Kyushu, Japan acquired on May 28 1986

GA parameters are set as follows,

The number of the initial chromosome=50,

Cross over probability=0.75,

Finished generation number=30000,

Mutation probability (only for Simple GA)=0.03,

The number of final clusters=5.

In order to evaluate clustering performance, maximum likelihood classification method is applied to the data with the following five classes, artificial structure, road, water body, paddy field, and bare soil. Classified results are shown in Figure 9. Also clustered results for Simple GA and Messy GA are shown in Figure 10 (a) and (b), respectively. The results are quite obvious that the clustered result from Messy GA is superior to that of Simple GA.

Clustering performance of Percent Correct Clustering: PCC and the number of correct clustering pixels are shown in Table 2 together with the converged iteration number.

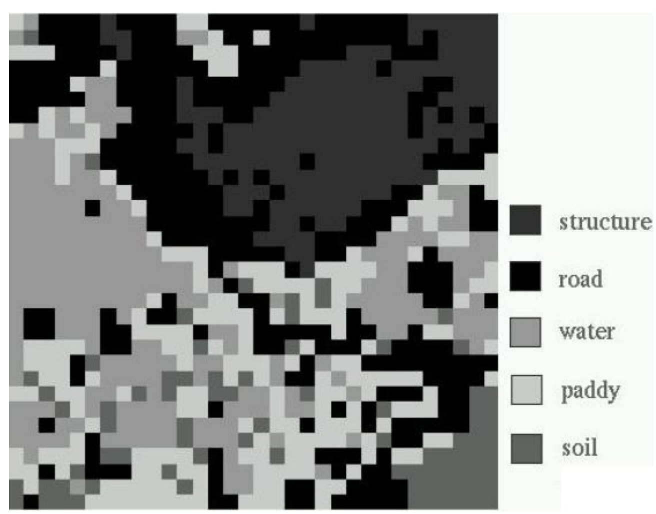
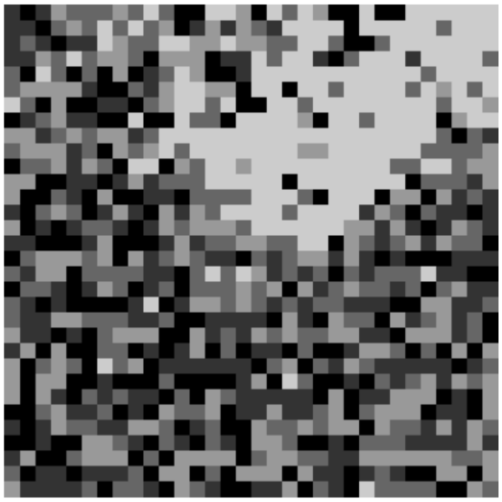
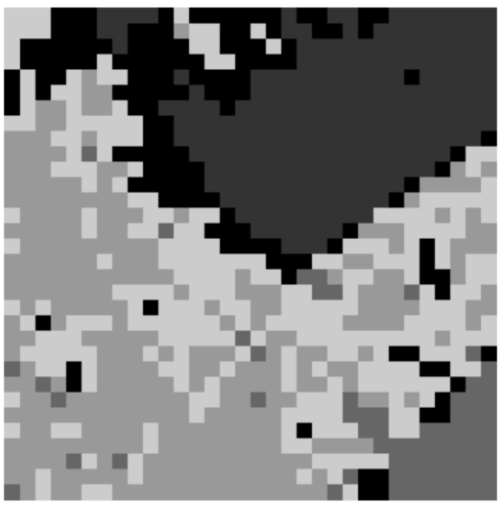


Figure 9 Classified results



(a)Simple GA



(b)Messy GA

Figure 10 Clustered Results

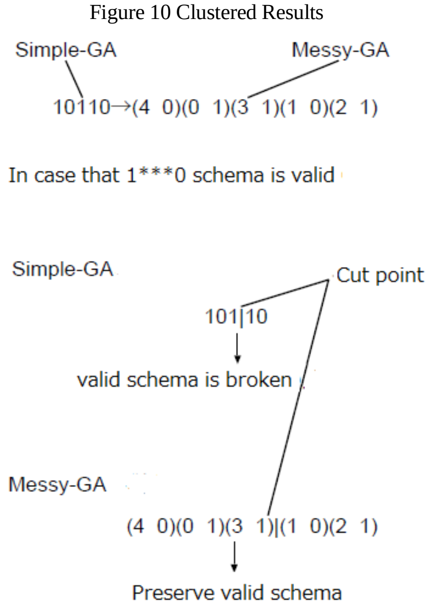


Figure 11 Reason for which the proposed Messy GA clustering is superior to the Simple GA clustering

TABLE II. CLUSTERING PERFORMANCE

Method	Converged iteration number	Number of correct clustering pixels	Percent Correct Clustering: PCC
Simple GA	29238	335	32.71
Messy GA	2998	543	53.03

From Table 2, the converged iteration number of the Simple GA clustering is greater than that of the Messy GA clustering. It takes ten time of processing time is required for the Simple GA clustering in comparison to that of the Messy GA clustering. Meanwhile, the number of correct clustering pixels of the Simple GA clustering is less than that of the Messy GA clustering. This is same thing for percent correct clustering. Therefore, it may say that the proposed Messy GA clustering is superior to the conventional Simple GA clustering. The reason for improvement of convergence performance is that useful building block of chromosome can be created in the initial phase. Through iteration, such useful schema can be preserved and updated efficiently and effectively. On the other hands, relatively long useful schema is broken during the cross over processes for the Simple GA clustering. Therefore, Percent Correct Clustering: PCC of the Simple GA clustering is not good enough. In contrast, the probability of schema broken is reduced for the Messy GA clustering. Figure 11 shows an example for such mechanism. If the cut point for cross over process is situated at the last two digits of the chromosome, and if the 1***0 of chromosome is useful and valid, then valid chromosome of 10110 is broken by cross over for the Simple GA clustering while it is maintained or preserved for the Messy GA clustering.

IV. CONCLUSION

Clustering method for remote sensing satellite image classification based on Messy Genetic Algorithm: GA is proposed. Through simulation study and experiments with real remote sensing satellite images, the proposed method is validated in comparison to the conventional simple GA. It is also found that the proposed clustering method is useful for preprocessing of the classifications.

The converged iteration number of the Simple GA clustering is greater than that of the Messy GA clustering. It takes ten time of processing time is required for the Simple GA clustering in comparison to that of the Messy GA clustering. Meanwhile, the number of correct clustering pixels of the Simple GA clustering is less than that of the Messy GA clustering.

ACKNOWLEDGMENT

The author would like to thank Mr. Koichi Tateno for his effort to conduct simulation and experiments.

REFERENCES

[1] M. Takagi and H.Shimoda Edt. K.Arai et al., Image Analysis Handbook, The University of Tokyo Publishing Co. Ltd., 2001.

[2] A.Yoshizawa, K.Arai, GA based clustering utilizing spectral and spatial context information, Journal of Image and Electronics Engineering Society of Japan, Vol.31, No.2, 202-209, 2003.

[3] K. Arai, A. Yoshizawa, K. Tateno, Messy GA algorith utilizing clustering method for remote sensing satellite inmage clustering, Journal

of Japan Society for Photogrammetry and Remote Sensing, Vol.41, No.5, 34－41, 2003.

[4] Kohei Arai, Learning processes of image clustering method with density maps derived from Self-Organizing Mapping(SOM), ournal of Japan Society for Photogrammetry and Remote Sensing, Vol.43、5、62-67 (2004)

[5] Kohei Arai, Non-linear merge and split method for image clustering, ournal of Japan Society for Photogrammetry and Remote Sensing, Vol.43、5、68-73 (2004)

[6] K. Arai, Quang Bu, GA based ISODATA clustering taking into account concave shape of probability density function, ournal of Japan Society for Photogrammetry and Remote Sensing, Vol.47、1、17一25、2008

[7] K.Arai, Quang Bu, Pursuit reinforcement competitive learning based online clustering and its application of image portion retrievals, Journal of Image and Electronics Engineering Society of Japan, Vol.39,3,301-309,2010

[8] K.Arai, Quang Bu, Accerelation of convergence of the pursuit reinforcement competitive learning based online clustering with learning automaton and its application of evacuatio simulations, Journal of Image and Electronics Engineering Society of Japan, Vol.40, 2, 361-168, 2011.

AUTHORS PROFILE

Kohei Arai, He received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science, and Technology of the University of Tokyo from 1974 to 1978 also was with National Space Development Agency of Japan (current JAXA) from 1979 to 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post Doctoral Fellow of National Science and Engineering Research Council of Canada. He was appointed professor at Department of Information Science, Saga University in 1990. He was appointed councilor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was also appointed councilor of Saga University from 2002 and 2003 followed by an executive councilor of the Remote Sensing Society of Japan for 2003 to 2005. He is an adjunct professor of University of Arizona, USA since 1998. He also was appointed vice chairman of the Commission "A" of ICSU/COSPAR in 2008. He wrote 30 books and published 332 journal papers.

Visualization of Link Structures and URL Retrievals Utilizing Internal Structure of URLs Based on Brunch and Bound Algorithms

Personal Information Collection Robot

Kohei Arai[1]
Graduate School of Science and Engineering
Saga University
Saga City, Japan

***Abstract*— Method for visualization of URL link structure and URL retrievals using internal structure of URLs based on brunch and bound method is proposed. Twisting link structure of URLs can be solved by the proposed visualization method. Also some improvements are observed for the proposed brunch and bound based method in comparison to the conventional URL retrieval methods.**

Keywords- visualization of link structure; URL retrieval; brunch and bound method; serch engine; information collection robot.

I. INTRODUCTION

Search engines, Google, Yahoo!, Baidu, Bing, Yandex, Ask, and AOL (market share order) are widely available for URL retrievals [1]-[3]. Information collection robots are working for creation of URL database which allows for smart information retrievals, in particular, URL retrievals. The conventional search engines can be divided into four categories, Directory types, Robot types, Meta types, and Hybrid types. Yahoo search engine is one of Hybrid types (Robot type and manmade classification) while Google search engine is one of Robot types. Any of search engines classify newly created web sites into some classes in the database by human and/or robots. Thousands of hits are obtained when users try to search the most preferable web site with a single keyword or frequent keywords. On the other hands, not appropriate web sites are hit by using simple keywords. To make URL or information search much efficient in a comprehensive manner, the methods for visualization and for search with brunch and bound method are proposed. Furthermore, the proposed information collection robot is a personalized by users and does work together with any kinds of search engines.

In the database, all the URLs are linked each other, in general. Visualization tools for URL link structure are very useful for not only manual URL search but also direct understanding the relations among URLs. Natto View is one of those of visualization tools. It is sometime hard to understand the relations among URLs for the reason that those are linked each other in twisted structure. The proposed visualization tool allows represent the twisted structure of the relations among URLs.

By using the relations among URL links, it can be accelerate information retrievals, or URL retrievals based on brunch and bound method. Through experiments with internal URL links in Saga University, it is verified that the proposed URL retrieval method is superior to the conventional URL retrieval methods.

The following section describes the proposed URL link structure visualization tool and URL retrieval method based on brunch and bound method followed by some experiments. Then conclusion is described together with some discussions.

II. PROPOSED METHOD

A. Search Engine

Configuration of the conventional search engines with information collection robot are shown in Figure 1.

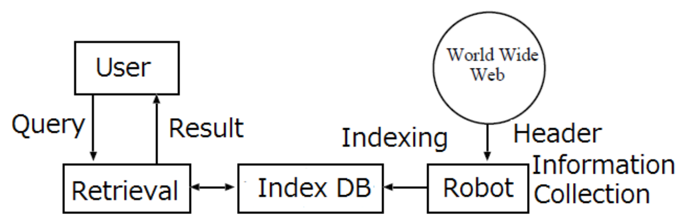


Figure 1 Configuration of the conventional search engines with information collection robot

Users send their query to the retrieval system. In parallel, information collection robot always collect header information from the web sites then create index for search in the index database (DB). Therefore, retrieval system send back the retrieved results by referencing the index DB.

B. Process Flow

Figure 2 shows procedure of the proposed URL search engine. The basic idea of the proposed search procedure is personalized information collection robot. Every time users search web site, information collection robot gathers preferable information of web sites and update users own personal database. Through learning processes, the personal information collection robot is getting craver. The other proposed method is visualization tool for a comprehensive representation of link relations among web sites. Then brunch and bound method based search is described.

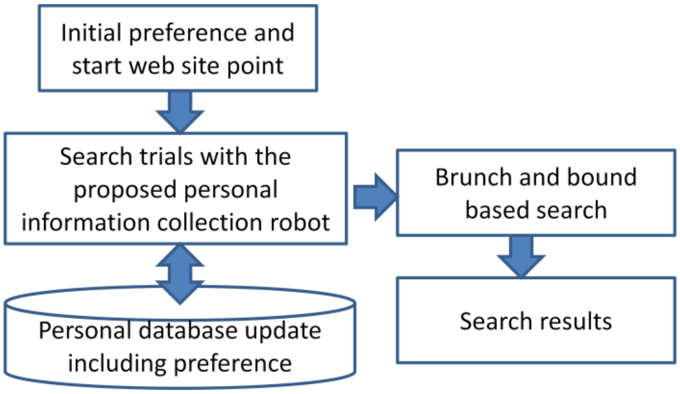


Figure 2 Process flow of the proposed URL search engine

C. URL Link Structure Visualization Method

General information collecting robots start from the original URL, and get the header information which contains all the information required for links as well as keyword, descriptions and so on as tag information. Then the robot gathers the information required for links to the other URLs from the found URLs. Thus the gathered information required for reaching to the other related URLs are stored in the database of search engine.

There are some related research works on visualization of the relations among URLs, Kinematic model, Natto view model, and corn tree model. In the Kinematic model, there are the following conditions,

(1) The distance between spheres (nodes) has to be greater than the acceptable distance,

(2) Summation of the arcs (links) has to be minimized

(3) Avoid any overlap between arc and the different nearby nodes.

On the other hands, Natto view allows manipulate nodes and links manually. Magnification, translation, rotation are available to improve visibility. Meanwhile, corn tree model allows representation of hierarchical structure of the URLs using corn shaped three dimensional space. At the top of the corn, there is a parent URL followed by children and grandchildren and so on. Thus all the previously proposed visualization tools allow representation of focusing URLs in concern precisely.

A concept of the visualization of the relations among URLs is shown in Figure 3. All the URLs are illustrated with circles while the relations are represented with lines between URLs.

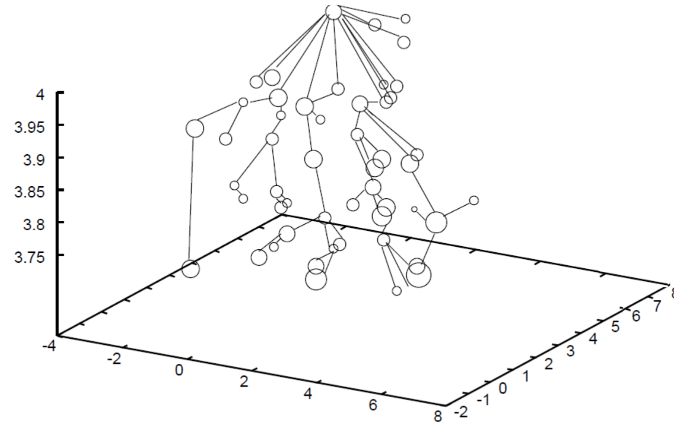


Figure 3 Concept of the visualization of the relations among URLs

In order to visualize the relations among URLs, the size and the distance are defined. The size defined as the number of links of the URL in concern. Deepening on the size, radius of node circle is determined. The distance between node circles is determined by the number of identical keywords and the words in the Meta tag of the header information of URL which must be highly correlated to the relation between URLs. The location of URL in X, Y plane is determined with the relations of web servers. If the URLs are in the same web server, then Z axis is determined while the URLs are in the different web server, then X,Y coordinates are determined as shown in Figure 4.

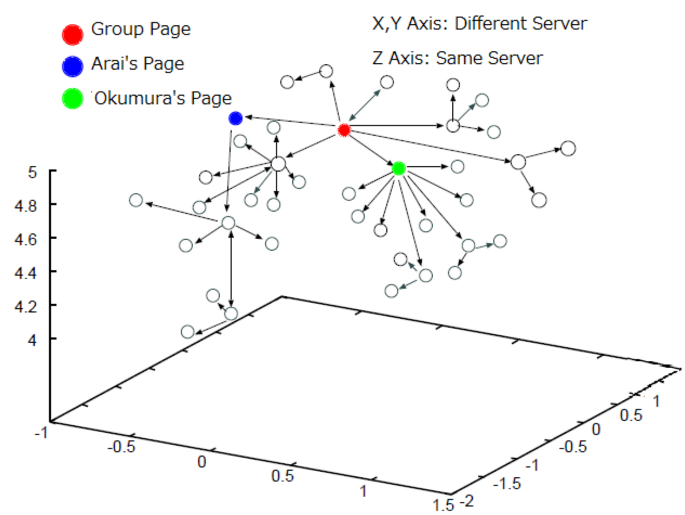


Figure 4 Determination of the location of URL node circles.

Then search begins from the red circle as shown in Figure 5, for instance. Information collection robot works next in the same time at which search engine is working.

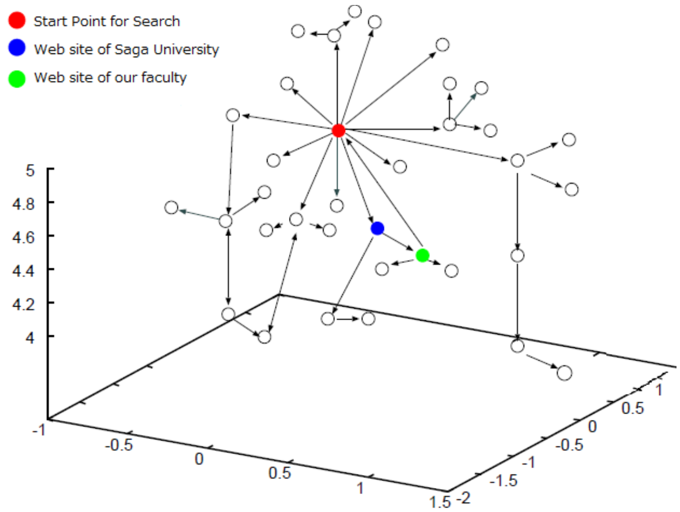
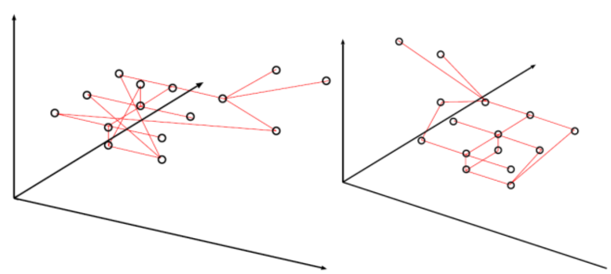


Figure 5 Information collection robot gather the information required for link and the next search

There is a problem on the twisted link representations as shown in Figure 6 (a). It is really hard to understand the links for the twisted links. By using Z axis allowance. untwisted representation can be done as shown in Figure 6 (b)



(a)Twisted links (b)Untwisted links

Figure 6 Method for untwisted representation of links among URL nodes

D. Information Collection Robot

The proposed information collection robot gathers header information together with checking whether or not information contains users' preferable information. Keywords and the frequency of the keywords are defined as users' preferences. Therefore, users can reach the most preferable web site efficiently by keying these preferences. Keywords and their frequencies are linked together. Therefore, users can omit key-in the frequency. Also information collection robot has learning capability. Therefore, only thing users have to do is just key-in their preferable keywords.

A Client-Server system is assumed for information collection robot. With the socket interface, server and client send and receive information. Hand shake procedure is shown below,

```
new IO::Socket::INET(
     [LocalAddr => 'hostname',]
     LocalPort => 'port',
     Proto => 'protocol',
     Listen => listen-limit,
     Reuse => reuse-number )
```

This is for the server side and the following is the client side,

```
new IO::Socket::INET (
     PeerAddr => 'hostname',
     PeerPort => 'port',
     Proto => 'protocol',
     TimeOut => 'timeout-second')
```

E. Brunch and Bound Search Method

URL search has to be done in an efficient manner. Blue force type of search method (try to search all available URLs) is not efficient. There is brunch bound search method as one of optimization methods which allows bound the brunch at which no appropriate brunch exists further below as shown in Figure 7.

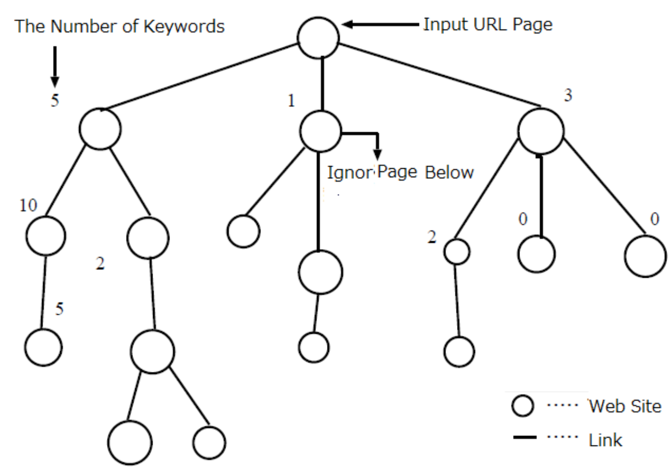


Figure 7 Brunch and bound search method

Namely, the procedure of the proposed brunch and bound is as follows,

(1) Client connect to URLs which are included in the designated web site through socket interface and then Web page contents are saved in the DB

(2) Check the frequency of the corresponding keywords from the saved web page contents in the DB

(3) If the frequency is not exceed the certain number, then the URL and URLs which are existing below the URL are omitted for search

(4) (1) to (3) are performed for new query of search

(5) Retrieved results are output when the search is finished then the result file is converted to html file.

III. Implementation And Experiments

A. Web Design

Web pages are created with Perl programming language. Perl is interpreter language which allows input text and output report as well as text files manipulations and system management. Figure 8 shows an example of web page which allows input keywords for URL collection. Only thing users have to do is just key-in their own preferable keywords in the boxes.

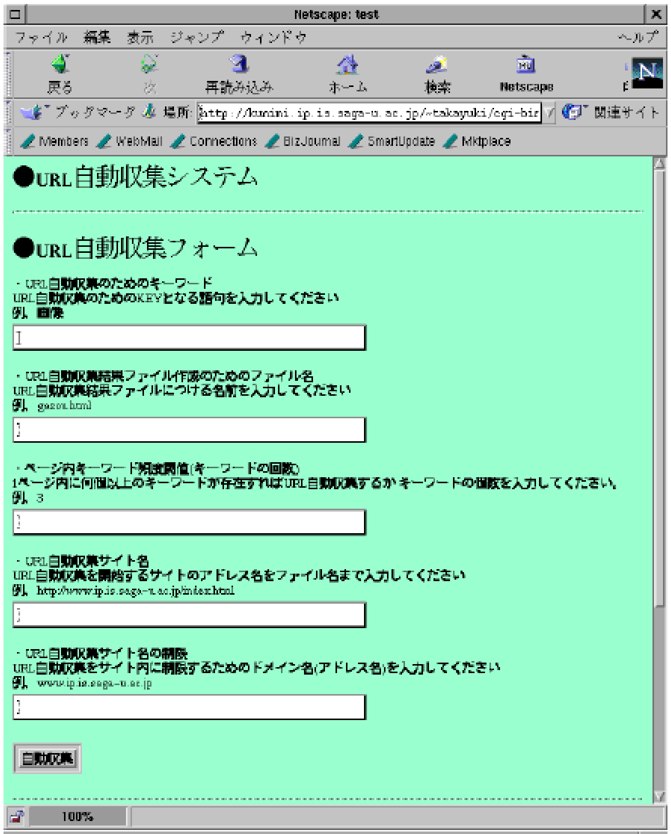


Figure 8 Example of web page which allows input keywords for URL collection.

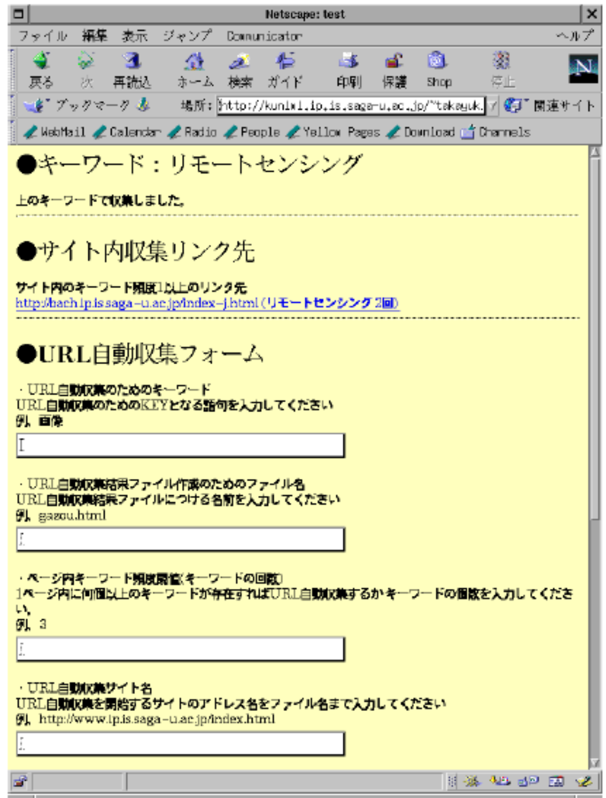
B. Example of Search Results

Figure 9 and 10 shows the retrieved results with the keywords, “Remote sensing” and “Image”, respectively. All the existing URLs in the laboratory are listed as the results. For the search results of “Remote sensing”, there are eight URLs as the search results while that for the keyword of “Image”, there are 24 of URLs as the search results. The order of the search results of URLs are sorted by the frequency.

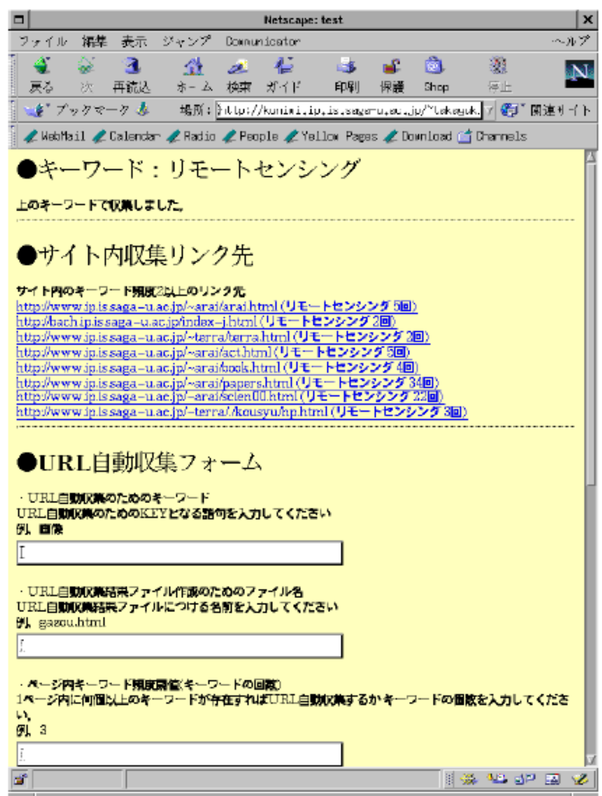
If users start their search with just “Remote Sensing” then users received Figure 9 (a) of search results. Then users selected the first URL, after that, users received Figure 9 (b) of search results. These are same thing for the keyword “Image”. Thus users can search deeper and deeper as much as they could.

C. Effect of Brunch and Bound Search Method

In order to confirm the effect of the brunch and bound search method, man-machine time, CPU time, and elapsed time are evaluated for the search with the keywords “Remote Sensing” and “Image”. The results are summarized in the Table1. It is quite obvious that the proposed brunch and bound search method is superior to the blue force type of search which search all the available URLs.

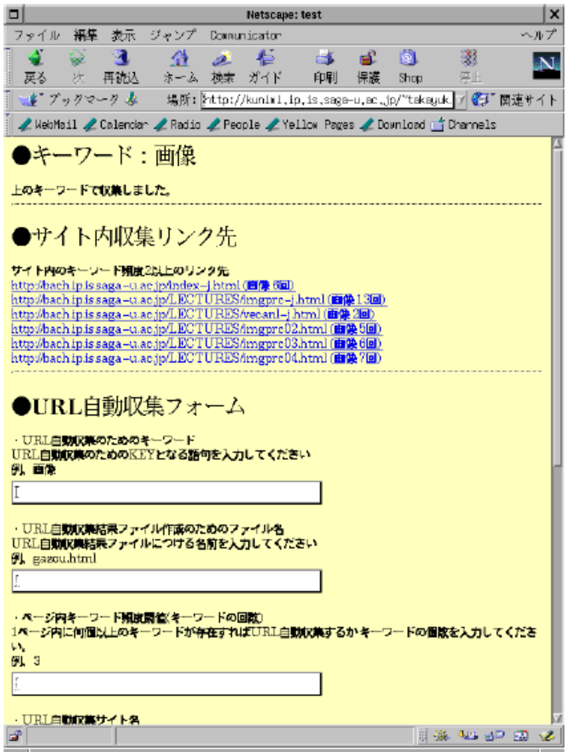


(a)

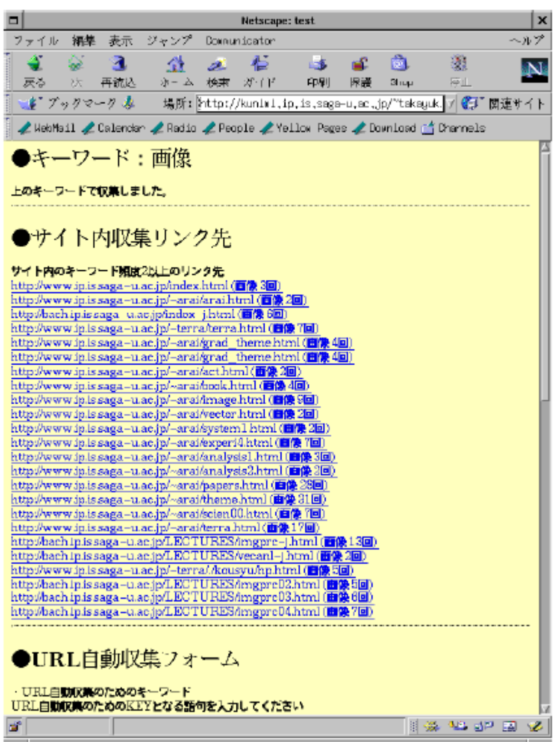


(b)

Figure 9 Retrieved results with the keyword “Remote Sensing”



(a)



(b)

Figure 10 Retrieved results with the keyword "Image"

D. *Effect of the Starting URL Point*

The effect of the search starting URL points is also evaluated with the keywords "Remote Sensing" and "Image". The results are shown in Table 2. The results show that efficient search can be done by starting the search with the closest URL to the desired URL.

TABLE I. TABLE TYPE STYLES

	Brunch and Bound	All URL Search
Man-Machine Time(min)	0.58	1.74
CPU Time (sec)	0.41	1.34
Elapsed Time (min)	1.07	3.32

TABLE II. TABLE TYPE STYLES

	Number of Keywords	
Keyword	Image	Remote Sensing
From Top of Laboratory	18	9
From Okumura's Page	2	1

It is easy to say that. It is not so easy to search. By looking at the link structure of the URLs using the proposed visualization method, users can find the starting URL point for the designated search purposes easily.

IV. CONCLUSION

Method for visualization of URL link structure and URL retrievals using internal structure of URLs based on brunch and bound method are proposed. Twisting link structure of URLs can be solved by the proposed visualization method. Also some improvements are observed for the proposed brunch and bound based method in comparison to the conventional URL retrieval methods

ACKNOWLEDGMENT

The author would like to thank Mr. Takayuki Sera for his effort to conduct experiments.

REFERENCES

[1] Pavliva, Halia (2012-04-02). "Yandex Internet Search Share Gains, Google Steady: Liveinternet". Bloomberg.com. http://www.bloomberg.com/news/2012-04-02/yandex-internet-search-share-gains-google-steady-liveinternet.html. Retrieved 2012-05-14.

[2] Segev, Elad (2010). Google and the Digital Divide: The Biases of Online Knowledge, Oxford: Chandos Publishing.

[3] Vaughan, L. & Thelwall, M. (2004). Search engine coverage bias: evidence and possible causes, Information Processing & Management, 40(4), 693-707.

AUTHORS PROFILE

Kohei Arai, He received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science, and Technology of the University of Tokyo from 1974 to 1978 also was with National Space Development Agency of Japan (current JAXA) from 1979 to 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post Doctoral Fellow of National Science and Engineering Research Council of Canada. He was appointed professor at Department of Information Science, Saga University in 1990. He was appointed councilor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was also appointed councilor of Saga University from 2002 and 2003 followed by an executive councilor of the Remote Sensing Society of Japan for 2003 to 2005. He is an adjunct professor of University of Arizona, USA since 1998. He also was appointed vice chairman of the Commission "A" of ICSU/COSPAR in 2008. He wrote 30 books and published 332 journal papers.

An Exploration on Brain Computer Interface and Its Recent Trends

T. Kameswara Rao
Assoc. Prof., Dept of CSE
Visvodaya Technical Academy
Kavali, AP, India

M. Rajyalakshmi
Assoc. Prof., Dept of CSE
Visvodaya Technical Academy
Kavali, AP, India

Dr. T. V. Prasad
Dean of Computing Sciences
Visvodaya Technical Academy
Kavali, AP, India

***Abstract*— Detailed exploration on Brain Computer Interface (BCI) and its recent trends has been done in this paper. Work is being done to identify objects, images, videos and their color compositions. Efforts are on the way in understanding speech, words, emotions, feelings and moods. When humans watch the surrounding environment, visual data is processed by the brain, and it is possible to reconstruct the same on the screen with some appreciable accuracy by analyzing the physiological data. This data is acquired by using one of the non-invasive techniques like electroencephalography (EEG) in BCI. The acquired signal is to be translated to produce the image on to the screen. This paper also lays suitable directions for future work.**

Keywords- BCI; EEG; brain image reconstruction.

I. INTRODUCTION

In general, all living beings can connect with the surrounding environment with their five senses, but major role is played by the visual data, which is percepted by eyes. Vision begins with light passing through the cornea. All the living beings can see the surroundings, along with some animals and birds, humans also can identify the colors due to their eye structure. Unlike some other living beings, human eye contains three types of cones (so named because of their shape) that are sensitive to red, green and blue colors, and identify the colors under suitable lighting conditions.

Rods (so named because of their shape) also are the part of the eye structure along with the cones, and are involved in identifying objects under dim lighting conditions because they are more sensitive to dim light than cones and do not sense the color, produce grey scale data of the objects. Rods are common in all living beings' eye structure. Figure 1 shows the arrangement of rods and cones. The human eye's 125 million visual receptors (composed of rods and cones) turns light into electrical signals and send to the brain through the optical nerve [1].

II. WORKING OF BRAIN

Brain is completely covered by a network of different types of neurons, which are the processing units of the data. Figure 2 shows the different types of neurons according to their structure. Most neurons have four functionalities in common, viz., input, trigger, conductile and output. These neurons percepts the data from synaptic terminals and it is processed in the cell body. Neurons can perform two types of actions namely fire and inhibit. They fire when the severity of the data is more than the threshold value which is set based on the experience/ training or they inhibit if severity is below the threshold value [1].

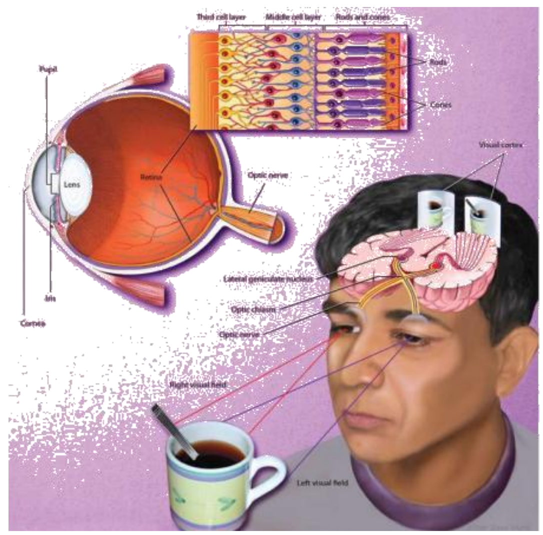


Figure 1Identification of objects and their colors by brain

Brain is assumed to be divided into 52 discrete local areas by Korbinian Brodmann, and named as cyto-architectural map [3], each is processing specific type of data. For example visual data is gathered to visual cortex areas 17, 18, 19. These areas are depicted in Figure 3.

When the visual data is processed by the neurons which are in the areas 17, 18 and 19 [3], there, the neurons of that area, generate electric pulses or signals and magnetic fields [2] to perform actions by actuators of the body. Usually, the electric signals of the brain are generated by pumping the ions like sodium (Na^+), potassium (K^+), calcium (Ca^{++}), chlorine (Cl^-), through the neuron membranes in the direction ruled by the membrane potential[14].

III. PHASES IN BRAIN COMPUTER INTERFACING

There are many phases in Brain Computer Interfacing. The major phases are as follows:

1. Signal Acquisition
2. Signal Pre-Processing
3. Signal Classification
4. Computer Interaction

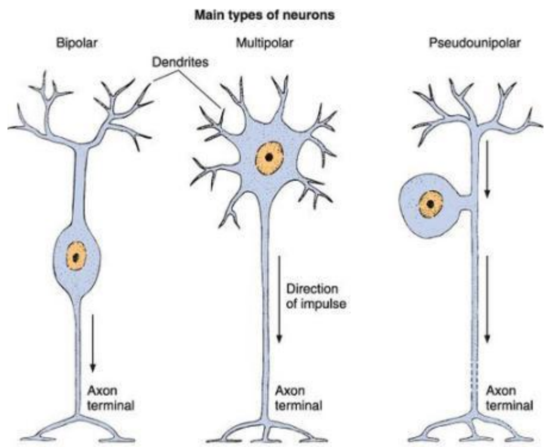


Figure 2 Structural types of Neurons

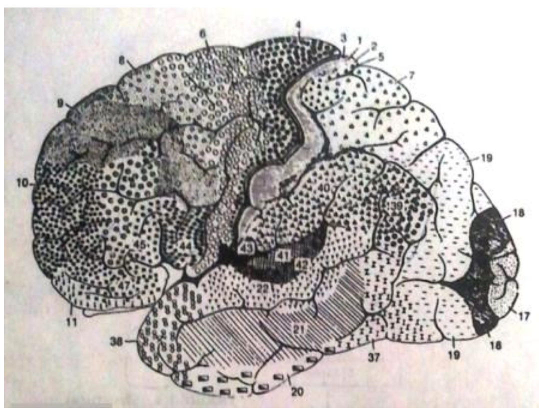


Figure 3 Cytoarchitectural map of cerebral cortex.

A. Signal Acquisition:

The electric signals generated by the neurons are acquired and processed by the signal acquisition and processing techniques and devices. In general, there are two types of brain signal acquisition techniques [4][7]:

a) Invasive acquisition

These techniques are used to capture the brain signals (also referred as electrophysiological signals) using implanted electrodes in brain tissue directly from the cerebral cortex, as shown in the Figure 4.

This Invasive signal acquisition should require surgery to implant the sensors or electrodes. These electrodes are implanted by opening the skull through a proper surgical procedure called craniotomy [5]. The electrodes are placed on the cortex; the signals acquired from these electrodes are called electro-corticogram (ECoG) [5]. Invasive signal acquisition techniques give excellent quality of signals.

But, even though the Invasive signals, can provide fast and potentially information-rich information, there are drawbacks in this technique. Some of them are, Invasive technique must require a surgery which is an ethical controversy. That's why Invasive techniques are almost exclusively investigated in animal models. At the same time, the threshold for their use will be higher.

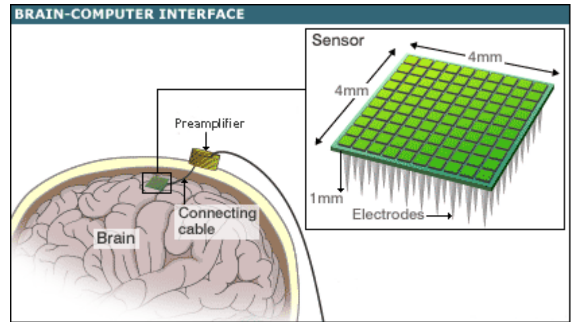


Figure: 4. Invasive data acquisition

Electrodes are piece of wire or metal pins, sometimes they may not work properly in the brain.

b) Non-Invasive acquisition

These techniques are used to capture the signals or electrophysiological signals from the scalp, as seen in the Figure 5, by using the technologies like electroencephalogram (EEG), functional magnetic resonance imaging (fMRI), magneto-encephalogram (MEG), P-300 based BCI etc. [5]. Among non-invasive Brain Computer Interfaces (BCIs), electroencephalogram (EEG) has been the most commonly used for them because EEG is advantageous in terms of its simplicity and ease of use, which meets BCI specifications when considering practical use.

Figure: 5. Non-Invasive data Acquisition.

In general, EEG signals (EEGs) can be classified into two categories, spontaneous EEGs and stimulus evoked EEGs. Focusing on stimulus evoked EEG signals called P300 and Visual Evoked Potentials (VEPs) are often utilized for BCIs [17]. Both types of BCIs extract the intention of users. While P300 signals are thought to be derived from the thoughts of users, VEPs are simply derived from physical reaction to visual stimulation. In that sense, VEP-based BCIs are thus known as the simplest BCIs [19]. In this paper, only electroencephalogram is concentrated as the signal capturing technology.

These captured signals are used for BCI and are too weak, about 100 μV [6], that's why they have to be amplified. The capturing devices are wired or wireless. A wired device and a wireless device are shown in Figure 5 and 6 respectively.

The signals are amplified to 10000 times. A sample EEG wave is shown in Figure 7 along with a Frequency-Amplitude graph.

Figure: 6. A wireless non-invasive signal capturing device

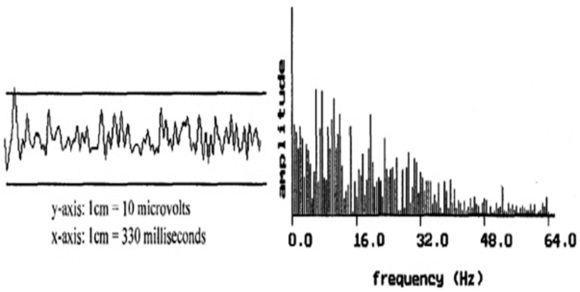


Figure 7. An EEG wave form recorded by a fore head electrode and its spectrum

B. Signal Pre-Processing

Whatever the technique we follow, the unwanted signal i.e., noise in the signal is inevitable. EEG recordings typically not only contain electrical signals from the brain, but also several unwanted signals like

- interference from electronic equipment, as for example the 50 or 60Hz, power supply signals,
- electromyography (EMG) signals evoked by muscular activity,
- ocular artifacts, due to eye movement or blinking.

Those unwanted components may bias the analysis of the EEG, and may lead to wrong conclusions. So the signal must be pre-processed to remove the noise. There are several preprocessing techniques to remove unwanted signals from EEG [18].

- **Basic Filtering:** The spurious 50 or 60Hz power supply signals are typically removed by a band-stop filter, which is a filter that passes most frequencies unaltered, but attenuates those in a specific range (e.g., at 50 or 60Hz) to very low levels.
  However, other artifacts such as electromyogram (EMG) signals and ocular artifacts typically affect a large frequency band and their spectrum may vary over time. Therefore, bandstop filters are usually not effective to eliminate such artifacts. One is often interested in specific frequency bands in the EEG, such as 4–8Hz (theta), 8–10Hz (alpha 1), 10–12Hz (alpha 2), 12–30Hz (beta), and 30–100Hz (gamma). Such frequency bands are usually extracted by a band-pass filter, which is a filter that passes frequencies within a certain range and rejects (attenuates) frequencies outside that range.
- **Adaptive Filtering:** The spectrum of artifacts is often a priori unknown. Therefore, applying a fixed filter to EEG data would not be effective to remove artifacts. The filter needs to adapt to the spectrum of the recorded EEG: it should attenuate the recorded EEG in frequency ranges that mostly contain artifacts. For instance, instead of using an online notch filter centered at a fixed frequency, one may apply an offline notch filter whose characteristics are determined by the spectrum of the recorded EEG. One may additionally use EOG (electro-oculography) or EMG (electromyography) measurements to design the adaptive filter, since those measurements are usually strongly correlated with artifacts.
- **Blind Source Separation:** An alternative approach, known as "blind source separation" (BSS), starts from the assumption that EEG signals can be described, to a good approximation, by a finite set of sources, located within the brain; each of those sources generate certain components of the EEG. Besides EEG, one sometimes also incorporates EOG and EMG signals into the analysis. In the context of artifact rejection, one makes the additional assumption that artifacts are generated by a subset of the extracted sources; one removes those sources, and next reconstructs the EEG from the remaining "clean" sources. Once the signals are acquired, it is inevitable to avoid noise. So the noise must be eliminated by using the algorithms like Adaptive filtering algorithm or RLS algorithm [13].

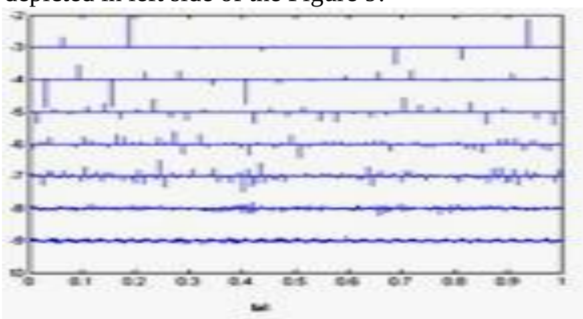
The recent technologies like merging of Translation Invariant Wavelet and ICA giving better results in noise filtering [15], shown in Figure 8.

C. Signal Classification

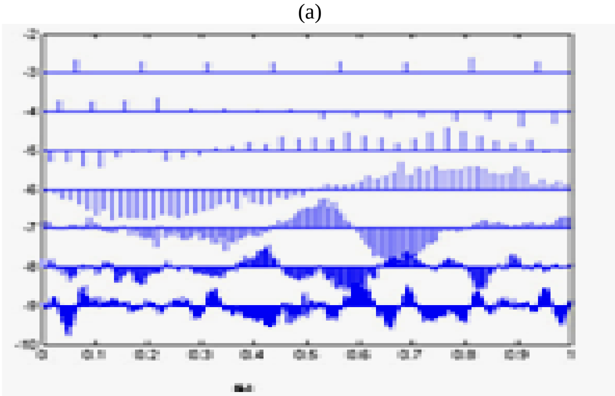
Since the brain signals or EEG are continuously captured by the capturing devices that have numerous electrodes, simultaneously capturing the signals in large amounts, it is not possible to clearly classify the waves. The waves are sometimes classified on both frequency and on their shape. There are six types of important signals [4], [8-12].

1) **Beta waves:** Frequency of these waves is between 13 and 30 Hz and the voltage or amplitude is very low about 5 to 30 µV. Beta waves are released when brain is active, in thinking, focusing on the problem solving etc. These waves' frequency can reach 50 Hz, during brain's intense activity. These waves are depicted in right side of the Figure 9.
2) **Alpha waves:** Frequency of these waves is between 8 and 13 Hz and the voltage or amplitude is about 30 to 50 µV. these waves are released when brain is relaxed or at inattention from the occipital and frontal cortexes. Alpha wave can reach a frequency of 20 Hz, which is a beta range, but has the characteristics of alpha state rather than beta. Alpha alone indicates a mindless state or empty state rather than relaxed of passive states. These waves are reduced by opening eyes or ears or by creating anxiety or

tension mental concentration etc. These waves are depicted in left side of the Figure 9.



(a)



(b)

Figure 8 (a) wave coefficient before de-noising (b) wave coefficient after de-noising [14]

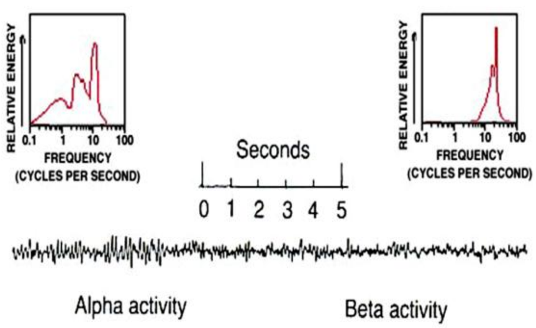


Figure 9. Alpha and Beta waves

3) **Theta waves:** Frequency of these waves is between 4 and 7 Hz and the voltage or amplitude is about 20 µV. Theta waves are generated when the brain is under emotional tensions, stress, frustration, disappointment etc. Theta waves are also released in unconsciousness or deep meditation. The peak frequency of theta waves is 7 Hz. These waves are depicted Figure 10.

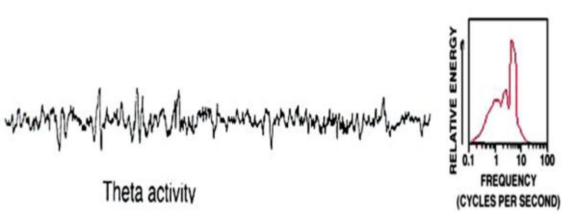


Figure 10. Theta waves

Delta waves: Frequency of these waves is between 0.5 and 4 Hz and the voltage or amplitude is varying. These waves are released deep sleep or when physical defects are there in the brain. These waves are depicted in Figure 11.

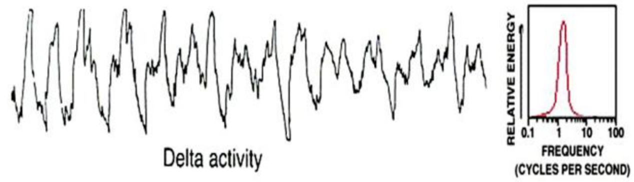


Figure 11. Delta waves

4) **Gamma waves:** Frequency of these waves is 35Hz or high. These waves are released when the brain is under a stream of consciousness.

5) **Mu waves:** Frequency of these waves is between 8 and 12 Hz. These waves are released with spontaneous nature of the brain like motor activities etc. These waves looks like alpha waves but alpha waves are recorded at occipital cortex and mu waves are recorded at motor cortex. These waves are depicted in Figure 12.

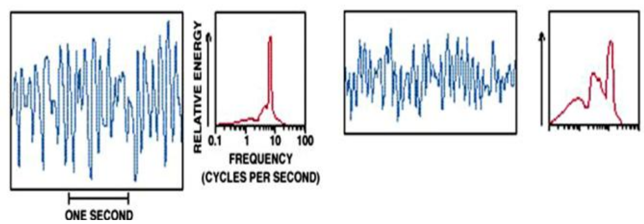


Figure 12. Mu and Alpha waves

Once the signals are cleaned, they will be processed and classified to find out which kind of mental task the subject is performing.

By using different BCI principles, the subjects participating in the P300 study had to spell a 5 character word with only 5minutes of training.

EEG data were acquired to train the system while the subject looked at Recent Advances in Brain-Computer Interface Systems, a 36 character matrix to spell the word WATER. During the real-time phase of the experiment, the subject spelled the word LUCAS.

For the P300 system 72.8 % were able to spell with 100 % accuracy and less than 3 % did not spell any character correctly as shown in Table 1.

Interesting is also that the Row-Column Speller reached a higher mean accuracy compared to the single character speller which produces higher P300 responses. This can be explained by the longer selection time per character for the SC speller [11].

D. Computer Interaction

Once the signals are classified, they will be used by an appropriate algorithm for the development of a certain application.

TABLE 1. CLASSIFICATION ACCURACY OF P300 EXPERIMENTS [11]

Classification Accuracy [%]	Row-Column Speller: Percentage of sessions (N=81)	Single Character Speller: Percentage of Sessions (N=38)
100	72.8	55.3
80-100	88.9	76.3
60-79	6.2	10.6
40-59	3.7	7.9
20-39	0.0	2.6
0-19	1.2	2.6
Average Accuracy of all subjects	91.0	82.0
Mean of subjects f participated in RC and SC (N=19)	85.3	77.9

IV. LIMITATIONS

Some limitations in implementing the BCI system are variability in the acquired EEG signals. Different types of signals are acquired from the same person in different sessions, different signals are acquired when many people are performing the same mental task and EEG signals are affected by the person's eye blinks, muscular movements, suddenly hearing sound and interference from electronic devices etc [7].

Based on these different types of applications like, character recognition, phrase recognition, object movement etc., it is observed that it is also possible to reconstruct the images and videos on to the screen from brain signals [20], [16]. This is depicted in Figure 13 and 14.

When the human watches the environment, the brain starts analyzing the image or video. It generates the signals according to the situation and information it received.

All the time, it is not possible to reconstruct the accurate image or video on to the screen because of the limitations in BCI, but it can be obtained some abstract image on to the screen as the Figure 13 depicts.

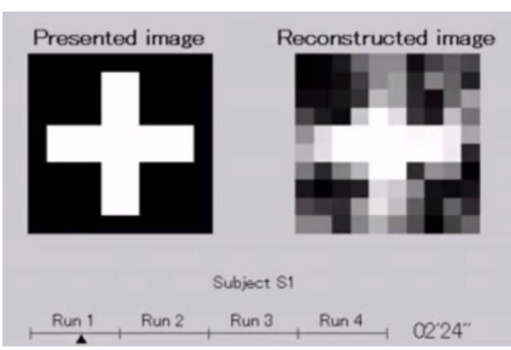
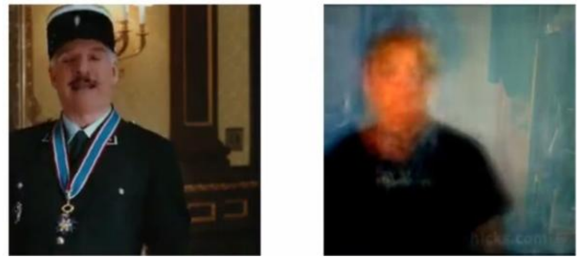


Figure 13 Binary image re-construction from brain signals [20].

This can be enhanced in such a way that, any color or object or image or video is seen, that can be produced on to the computer by properly capturing the signals of the brain. Not only binary images, but color images also reconstruct able.

Most of the binary image is reconstructed, but color image is not since the composition of colors in the brain is not as easy as binary.

So, to produce the color image or video on to the screen accurately from the brain signals, further much more great brain signal acquisition and analysis has to be taken place as future enhancement.



(a) (b)

Figure 14. Color image reconstruction from brain signals [16]: a) Presented Clip b) Clip reconstructed from brain activity

V. CONCLUSION AND FUTURE DIRECTIONS

Many complex processes and system would operate on the basis of thought in the future. Currently, the field of BCI is in infancy stage and would require deeper insights on how to capture the right signals and then process them suitably. The advancements are limited to recognition of certain words, expressions, moods, etc.

Efforts are being made to recognize the objects as they are seen by the brain. These efforts will bring in newer dimensions in the understanding of brain functioning, damage and repair. It is possible to recognize the thoughts of the human brain by capturing the right signals from the brain in future.

Research work can be taken up on the signal processing and analysis to tap the thoughts of the human. But present operating systems and interfaces are not suitable for working with thought based system. They have to be modified accordingly.

Processing and understanding thoughts for different purposes can lead to security and privacy issues. Ethical and standardization issues can also come up, which needs to be resolved earliest.

REFERENCES

[1] S. Haykin, "Neural Networks – A Comprehensive Foundation", Pearson Education, Asia, Second Edition.

[2] M. Junghofer and P. Peyk, "Analyzing electrical activity and magnetic fields in the brain", University of Constance, Germany.

[3] E. R. Kandel, J. H. Schwartz and T. M. Jessel, "Principles of Neural Sciences". Center for Neurobiology and Behavior, College of Physicians & Surgeons of Columbia University and The Howard Hughes Medical Institute, Fourth Edition

[4] J. Baztarrica Ochoa, "EEG Signal Classification for Brain Computer Interface Applications", 2002.

[5] B. Graimann, B. Allison, and G. Pfurtscheller, "Brain–Computer Interfaces: A Gentle Introduction".

[6] Md. Mahfuzur Rahman, Md. Mahmudul Hasan, Md. Mizanur Rahman, Md. Siam Hasan, Md. Mansurul Alam, "Brain computer interface", Bangladesh University of Engineering and Technology.

[7] S. Soman, S. Sen Gupta, P. Govind Raj, "Non invasive Brain Computer Interface for controlling user desktop", available at *www.cdacnoida.in/ASCNT-2012/ ASCNT- 2012/UC/6.pdf*

D. A. Torse, R. R. Maggavi, S. A. Pujari, “Non-Linear blind source separation for EEG signal pre processing in BCI system for epilepsy”, Int. J. of Computer Applications (0975 – 8887) Vol. 50, No. 14, 2012

[8] O. A. Rosso, S. Blanco, J. Yordanova, V. Kolev, A. Figliola, M. Schurmann and E. Basar, “Wavelet entropy: a new tool for analysis of short duration brain electrical signals”, J. of Neuroscience Methods, Vol. 105, pp. 65–75, 2001

[9] A. Erfanian, F. Oveisi and A. Shadvar, “Feature Extraction by Mutual Information Based on Minimal-Redundancy-Maximal-Relevance Criterion and Its Application to Classifying EEG Signal for Brain-Computer Interfaces”, Iran University of Science and Technology, Iran.

[10] R. Fazel-Rezai, “Recent Advances in Brain-Computer Interface Systems”, InTech, Janeza Trdine 9, Croatia

[11] L. r. Hochberg and J. P. Donoghue, “Sensors for Brain-Computer Interfaces- Options for Turning Thought into Action”, IEEE Engineering in Medicine and Biology Magazine, Sep/Oct 2006.

[12] C. Chandrakar and M. K. Kowar, “Denoising ECG signals using adaptive filter algorithm”, Int. J. of Soft Computing and Engineering (IJSCE), Vol. 2, Issue 1, 2012.

[13] H. L. Attwood and W. A. MacKay, Essentials of Neurophysiology, B. C. Decker, Hamilton, Canada, 1989.

[14] J. Walters-Williams and Y. Li, “A New Approach to Denoising EEG Signals - Merger of Translation Invariant Wavelet and ICA”, Int. J. of Biometrics and Bioinformatics, Vol. 5, Issue 2, 2011.

[15] Visual image recognition from human brain, available at http://www.youtube.com/watch?v=daY7uO0eftA

[16] J. R. Wolpaw, N. Birbaumer, D. J. McFarlanda, G. Pfurtschellere, Theresa M. Vaughan, “ Invited review Brain–computer interfaces for communication and control”, Elsevier Science Ireland Ltd., 2002

[17] J. Dauwels and F. Vialatte, “Topics in Brain Signal Processing”, available at *www.dauwels.com/Papers/Review_BSP.pdf*

[18] N. Yoshimura and N. Itakura “Usability of Transient VEPs in BCIs”, Tokyo Institute of Technology, The University of Electro-Communications, Japan

[19] Reconstruction from brain activity, available at http://www.youtube.com/watch?v=nsjDnYxJ0bo

AUTHORS PROFILE

Mr. T. Kameswara Rao received his Masters in Computer Applications from University of Madras in 2004 and Masters of Engineering from Satyabhama University in 2007. He is currently associated with the Dept. of Comp. Sc. & Engg. at Visvodaya Technical Academy, Kavali, AP, India, as Associate Professor. He has over 7 years of teaching experience at under graduate and graduate level. His areas of interest are biometrics, brain computer interface, artificial intelligence, artificial neural networks, psychology etc.

Ms. M. Rajyalakshmi received her Masters in Computer Science and Engineering from Jawaharlal Nehru Technological University, Anantapur in 2011. She is currently associated with the Dept. of Comp. Sc. & Engg. at Visvodaya Technical Academy, Kavali, AP, India, as Associate Professor. She has over 5 years of teaching experience at under graduate and graduate level. Her areas of interest are language processing, brain computer interface, automata theory, artificial neural networks, etc.

Dr. T. V. Prasad has over 17 years of experience in industry and academics. He has received his graduate and master’s degree in Computer Science from Nagarjuna University, AP, India. He was with the Bureau of Indian Standards, New Delhi for 11 years as Scientist/Deputy Director. He earned PhD from Jamia Millia Islamia University, New Delhi in the area of computer sciences/ bioinformatics. He has worked as Head of the Department of Computer Science & Engineering, Dean of R&D and Industrial Consultancy and then as Dean of Academic Affairs at Lingaya’s University, Faridabad. He is with Visvodaya Technical Academy, Kavali as Dean of Computing Sciences. He has lectured at various international and national forums on subjects related to computing. Prof. Prasad is a member of IEEE, IAENG, Computer Society of India (CSI), and life member of Indian Society of Remote Sensing (ISRS) and APBioNet. His research interests include bioinformatics, consciousness studies, artificial intelligence (natural language processing, swarm intelligence, robotics, BCI, knowledge representation and retrieval). He has over 73 papers in different journals and conferences, and has authored six books and two chapters.

An Interval-Based Context Reasoning Approach

Tao Lu, Xiaoling Liu, Xiaogang Liu, Shaokun Zhang
System Engineering Institute
Dalian University of Technology
Dalian, China

***Abstract*— Context-aware computing is an emerging computing paradigm that provides intelligent context-aware application. Context reasoning is an important aspect in context awareness, by which high level context can be derived from low-level context data. In this paper, we focus on the situation in mobile workspace, where a worker performs a set of activities to archive defined goals. The main part of being aware is to be able to answer the question of "what is going on". Therefore high level context we need to derive is current activity and its state. The approach we propose is knowledge-driven technique. Temporal relations as well as semantic relations are integrated into the context model of activity, and the recognition is performed based on the model. We first define the context model of activity, and then we analyze the characteristics of context change and propose a method of context reasoning.**

Keywords- context-aware; context reasoning; interval-based; activity recognition.

I. INTRODUCTION

Context-aware computing is an emerging computing paradigm that provides intelligent context-aware application. According to well-known definition proposed by Dey, context is any information that can be used to characterize the situation of an entity. An entity is a person, place or object that is considered relevant to the interaction between a user and an application, including the user and application themselves [1]. Context-awareness means exploiting context information to provide adaptive information or service, or reducing the interaction between user and application.

Context-awareness is related to the manipulation of context information pertaining to certain entities. Information from physical sensors are called low-level context. High level context, also called situation sometimes, can be derived from low-level context by proper interpreting. This process is context reasoning, or context interpretation. Situations are semantic abstractions from low-level contexts cues. So the relationship between low level context and situation must be integrated into a context model, which represents human knowledge about the world [2].

This can either be done by specification, i.e. human defines situations and their relationship based on his /her knowledge, or the model are learned automatically using machine learning techniques [2]. The two approaches are also called knowledge-driven approach and data-driven approach [3].

Data-driven techniques are based on the machine learning methods and are well suited for recognizing simple activities and gestures from raw sensor data or video data [4]. A wide range of algorithms and models include Hidden Markov Models [5], dynamic and naïve Bayes networks [6], and decision trees [7] and so on. Knowledge-driven techniques, concerning knowledge representation as well as reasoning with them, is closely related with classical topics in artificial intelligence [8]. Ontology-based method is one of frequently used technique. For example, Chen has proposed a technique to recognize activities through ontological reasoning [9]. Knowledge-driven techniques can also be used to recognizing complex situations based on the recognized simple context [10].

In this paper, we focus on the situation in mobile workspace, where a worker performs a set of activities to archive defined goals. The main part of being aware is to be able to answer the question of "what is going on". Therefore the situation, or high level context, we need to derive is activity and its state.

Determining the activity and its state cannot only be completed by observing the context but also need to draw conclusion from the observations. Knowledge about the relation of context and activity should be integrated into the model. Compared with most of knowledge-driven techniques, the approach we propose considers the temporal relation between activities and contexts.

Activity is seemed as process with duration, and context may change in the process. Each activity context model is a set of semantic relations and temporal constraints with respect to the activity and the context. If some observed contexts match the defined patterns, or if their times of occurrence meet the specified constraints, then an instance of this situation occurs.

We have developed an activity recognition approach taking the temporal relation into account in previous work [11]. This paper considers more temporal relations. Moreover semantic relations are also integrated into the model. By this means, more types of situation can be recognized.

The rest of the paper is organized as follows. Section 2 provides the context model of activity. Section 3 proposes the activity recognition approach. Case study is described in section 4 and section 5 concludes the paper.

II. CONTEXT MODEL OF ACTIVITY

Suppose the application is related to a set of contexts which are denoted as $c_1,c_2,...c_n$, and the domain of c_i is D_i. At given time t, the value of context c_i is denoted as $c_i(t)(c_i(t)\in D_i)$, $i=1,2,\ldots n$. The values of context $c_1,c_2,...c_n$ at time t is denoted as $C(t)$, $C(t)=(c_1(t),c_2(t),\ldots c_n(t))$.

Suppose $d_j^{(i)} \subseteq D_i$, $(c_i, d_j^{(i)})$ is called a context pattern. The context pattern$(c_i, d_j^{(i)})$ holds at time t, if and only if $c_i(t) \in d_j^{(i)}$, denoted as $hold\_at((c_i,d_j^{(i)}),t)$, i.e. $hold\_at((c_i,d_j^{(i)}),t) \Leftrightarrow c_i(t) \in d_j^{(i)}$.

Apparently $hold\_at((c_i,d_j^{(i)}),t) \Leftrightarrow \neg hold\_at((c_i,D_i-d_j^{(i)}),t)$, $(c_i,D_i-d_j^{(i)})$ is called the negative pattern of $(c_i,d_j^{(i)})$, denoted as $\neg(c_i,d_j^{(i)})$.

We define a special context pattern time-elapse(t_s,T), $hold\_at(time-elapse(t_s,T),t) \Leftrightarrow t = t_s + T$. t_s is the time to start timing. The context pattern holds after T. For simplicity, we use time-elapse instead of time-elapse(t_s,T). Initializing time-elapse in this paper means setting the time length and start timing.

A. Temporal relations of context pattern and activity

Context pattern may hold over time interval. We consider time as a linearly ordered discrete set of instants, a time interval is represented as an ordered pair of time points representing starting and ending time. Context pattern may hold in one or more time intervals.

We use $I(p)$ to denote the interval set in which the context pattern p holds, i.e.

$$I(p)=\{[t_s,t_e] \mid \forall t\in[t_s,t_e],\ hold\_at(p,t) \text{ and } \neg\exists\ [t_m,t_n],\ t_m \le t_s, t_n \ge t_e, [t_m,t_n] \ne [t_s,t_e],\ \forall t\in[t_m,t_n],\ hold\_at(p,t)\} \quad (1)$$

In same way, we can use an ordered pair to represent the time interval in which an activity is conducted. The starting time and ending time of activity a are denoted as *a.start* and *a.end*.

Allen has suggested a well-known temporal model based on relationships among intervals [12]. Here we use these predicates to model the temporal relations between context patterns and activities.

Let P be a context pattern set and A an activity set. *Equal*, *During*, *Start*, *Finish*, *Before*, *Overlap* are predicates representing relations between context patterns and activities. The meaning of the predicate is defined in TABLE I.

In TABLE I, we only list the relations usually used in our model. Other relations are not difficult to deduce from the meaning.

TABLE I. *TEMPORAL RELATION OF CONTEXT PATTERN AND ACTIVITY*

Relation	Definition
Equal(p,a)	$\exists\ [t_{ps},t_{pe}] \in I(p)$, $t_{ps}=$ *a.start*, $t_{pe}=$ *a.end*
During(p,a)	$\exists\ [t_{ps},t_{pe}] \in I(p)$, $t_{ps} \ge$ *a.start*, $t_{pe} \le$ *a.end*
During(a,p)	$\exists\ [t_{ps},t_{pe}] \in I(p)$, $t_{ps} \le$ *a.start*, $t_{pe} \ge$ *a.end*
Start(p,a)	$\exists\ [t_{ps},t_{pe}] \in I(p)$, $t_{ps}=$ *a.start*, $t_{pe}<$ *a.end*
Start(a,p)	$\exists\ [t_{ps},t_{pe}] \in I(p)$, $t_{ps}=$ *a.start*, $t_{pe}>$ *a.end*
Finish(p,a)	$\exists\ [t_{ps},t_{pe}] \in I(p)$, $t_{ps}>$ *a.start*, $t_{pe}=$*a.end*
Finish(a,p)	$\exists\ [t_{ps},t_{pe}] \in I(p)$, $t_{ps}<$ *a.start*, $t_{pe}=$*a.end*
Overlap(p,a)	$\exists\ [t_{ps},t_{pe}] \in I(p)$, $t_{ps}<$ *a.start*, $t_{pe}<$ *a.end*
Overlap(a,p)	$\exists\ [t_{ps},t_{pe}] \in I(p)$, $t_{ps}>$ *a.start*, $t_{pe}>$*a.end*
Before (p,a)	$\exists\ [t_{ps},t_{pe}] \in I(p)$, $t_{ps}<$ *a.start*

B. Semantic relations of context pattern and activity

Besides temporal relations, there are also semantic relations existing between context patterns and activities. COND, CONSEQ, PREM, ACCOMP are semantic relations defined to illustrate logic connection in context patterns and activities. The meaning and corresponding temporal relations are listed in TABLE II.

TABLE II. *SEMANTIC RELATION CONTEXT PATTERN AND ACTIVITY*

Relation	Meaning	Corresponding temporal realtion
COND(p,a)	Context pattern p is a condition of conducting activity a. If it is not satisfied as its corresponding temporal pattern, the activity is abnormally conducted.	*Before*(p,a), *During*(a,p), *Overlap*(p,a)
PREM(p,a)	Context pattern p is a premise of conducting activity a. If it is not satisfied as its corresponding temporal pattern, the activity is not started or interrupted.	*Overlap*(p,a), *During*(a,p), *Finish*(a,p)
ACCOMP(p,a)	Context pattern p will occur as its corresponding temporal pattern if the activity a is conducted normally	*Start*(p,a), *Equal*(p,a), *During*(p,a), *Finish*(p,a), *Overlap*(a,p)
CONSEQ(p,a)	Context pattern p is a consequence of activity a.	*Overlap*(a,p)

If COND(p,a), p is called a condition context pattern of a. In same way, for PREM(p,a), ACCOMP(p,a), CONSEQ(p,a), p are called premise context pattern, accompanying context pattern and consequence context pattern respectively.

C. Modeling activity with semantic relation and temporal relation

We can model the relation of activity and context pattern with semantic relation and temporal relation. These relations show the context changing rules when the activity is performed normally without interruption.

For real world problem, some semantic relations are not easily to be distinguished. For example, condition and premise are very similar. Here the difference between condition context pattern and premise context pattern is that the condition context pattern can be controlled by human, while the premise is objective. That is if a condition context pattern does not hold in proper time as defined, the activity can be performed but not normally. However, if a premise context pattern does not hold, it can be deduced that the activity has not been started.

The difference between consequence and accompanying context pattern is that the consequence is related with the goal of the activity, and the pattern holding means the activity will end or has ended, while the accompanying context pattern holding may means the activity has been started and if it does not occurred as defined, there may be interruption or abnormal situation.

TABLE II shows the possible temporal relation of every type semantic relation. The temporal relation should be determined when defining an activity. The different temporal relation is used differently.

When context interpretation is performed, the occurring context is compared with the model, and current high level context, i.e. the activity and its state is deduced.

III. Recognizing Activity By Context Reason

Generally in context-aware system, sensors (either virtual sensors or environmental sensors) are used to acquire the raw context. The acquired low-level context data are dealt with by many context-aware middleware or infrastructure and higher level context information are available and represented in formal format. Our work is to derive context in much higher level based on these context information, i.e. recognize the activity and its current state.

The input of our work is sequence of context values in different time, $c(t_1),c(t_2),\ldots c(t_n)$. Assume that every meaningful context change can be detected, and the values are omitted if there is no change happening. So if we have $c(t_i)$ and $c(t_{i+1})$, for any t, $t_i \le t < t_{i+1}$, $c(t)=c(t_i)$. We need to identify the activity and its state according to these context values and activity model.

There are two opposite strategies of processing context data. The first one analyzes previously received context, matching them with related temporal pattern on the basis of their semantic meanings and give the result. This strategy requires less memory but may introduce a higher processing delay. The second approach processes data incrementally, recognizing and stores partial consequence in the form of automata as soon as they are detected. This may need more memory (memorizing different state) but speed up the processing. We adopt the second strategy.

A. State and internal state

In activity models, temporal relations define the rules of context pattern occurring. Combined with their semantic meanings, these rules can be used to interpret “what is going on”.

Definition 1 Let $a \in A$, *s-pattern*(a) is called **start pattern** of activity a, if and only if:

$$s\text{-}pattern(a)=\{p \mid \mathrm{PREM}(p,a) \vee (\mathrm{ACCOMP}(p,a) \wedge (start(p,a) \vee equal(p,a) \vee start(a,p))\} \quad (2)$$

Starting pattern of an activity is a context pattern set, the pattern in which should hold when the activity is started. Therefore according to generally underlying assumption of context reason, it can also be used to determine the start of the activity.

When an activity a ends normally, its consequence context pattern should hold. Otherwise, if there is an accompanying or premise context pattern p, $equal(p,a)$, or $finish(p,a)$, p should cease holding, i.e. $\neg p$ holds, when the activity ends.

Definition 2 Let $a \in A$, *e-pattern*(a) is called **end pattern** of activity a, if and only if:

$$e\text{-}pattern(a)=\{p \mid \mathrm{CONSEQ}(p,a) \vee (\mathrm{PREM}(\neg p,a) \wedge finish(a, \neg p)) \vee (\mathrm{ACCOMP}(\neg p,a) \wedge (finish(\neg p,a) \vee equal(\neg p,a)))\} \quad (3)$$

Let $e\text{-}pattern(a,t)=\{p \mid p \in e\text{-}pattern(a) \wedge hold\_at(p,t)\}$, apparently $e\text{-}pattern(a,t)$ is a subset of $e\text{-}pattern(a)$, which can be used to measure how close to the end for the activity process. If $e\text{-}pattern(a,t_{i-1}) \subset e\text{-}pattern(a,t_i)$, the state at time t_i is closer to activity end than last moment. This is useful when the activity is judged to be ended, but meaningless when the activity is still proceeding normally.

When an activity is being conducted without interruption, its accompanying context pattern should always hold if its temporal relations with the activity is *equal*, so should the premise context pattern if the relation is *during* (i.e. activity during context pattern). If for an accompanying context pattern p, $finish(p,a)$, or overlap(a,p), p should be hold once it occurs until or after the activity end. Therefore, when the activity is being conducted without interruption, the conditions that the context should satisfy are changing.

Definition 3 Let $a \in A$, *o-pattern*(a) is called an **on pattern** of a, if and only if :

$$o\text{-}pattern(a) \subseteq \{p \mid (\mathrm{PREM}(p,a) \wedge During(a,p)) \vee (\mathrm{ACCOMP}(p,a) \wedge (Equal(p,a) \vee Finish(p,a) \vee Overlap(a,p)))\}$$
$$\text{and } o\text{-}pattern(a) \supseteq \{p \mid (\mathrm{PREM}(p,a) \wedge During(a,p)) \vee (\mathrm{ACCOMP}(p,a) \wedge Equal(p,a))\} \quad (4)$$

An activity may have more than one on patterns. Among them, if:

$$o\text{-}pattern(a)=\{p \mid (\mathrm{PREM}(p,a) \wedge During(a,p)) \vee (\mathrm{ACCOMP}(p,a) \wedge (Equal(p,a) \vee Finish(p,a) \vee Overlap(a,p)))\} \quad (5)$$

Any on pattern of a is the subset of *o-pattern*(a), denoted as *max-on-pattern*(a).

Proposition 1 Let $o\text{-}pattern(a,t) = \{p \mid p \in max\text{-}on\text{-}pattern(a) \wedge hold\_at(p,t)\}$, if the activity a is performing at time *t*, and there is no context change until next moment *t*+1, then the activity *a* is performing without interruption if and only if: *o-pattern*($a,t+1$) $\supseteq$ *o-pattern*(a,t).

Since the above proposition is easy to be proved according to the definitions, proof is omitted here.

However, context value at *t* satisfies the start pattern or end pattern, i.e. all patterns in *s-pattern* or *e-pattern* hold, does not mean the activity starts or ends normally. There are other condition should be checked. Similarly, if the context values over the interval satisfy the on patterns of an activity does not mean the activity is performed normally, even if no interruption happens. We will illustrate this afterwards.

There are 4 types of state to be recognized, called output state. The states are: *waiting* (for activity beginning), *on* (one activity is conducted), *suspend* (one activity is interrupted), *abnormal*. The recognizing result is denoted as 2-tuple:(activity name, state). The state transition is shown in Figure 1.

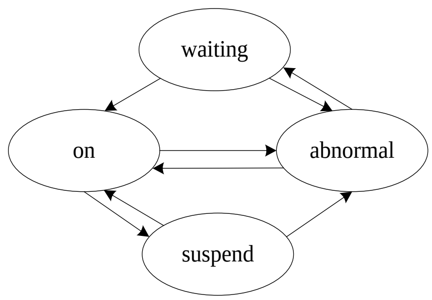


Figure 1. Output state transition.

In addition to above states, we define other 9 states which are used for system control, called internal states. The concept of internal state is actually state of automate, which has process memory itself when it is reached. The states are: *waiting*, *starting*, *on*, *suspend*, *ending*, *end*, *start abnormal*, *process abnormal*, *end abnormal*.

Among them, *waiting*, *on*, *suspend* have similar meanings with output states of same name. State *starting* means stage that all premise conditions are satisfied and part of accompanying context patterns have occurred but the start pattern is not satisfied. State *ending* means consequence context has occurred, or for the activity without consequence context pattern interruption lasts too long, hypothesis of activity ending need to be proved. If end pattern is satisfied, or time is longer enough, system converts to end state and check the condition to determine if it is normal. Therefore state *end* is actually a checking point without duration and there is no corresponding output state. Three abnormal states corresponds *abnormal* output state. Internal state *starting* is *start* when output. Since mostly there is no a time point in which all the start patterns are satisfied at same time, setting an internal state *starting* can avoid taking this situation as abnormal. Internal state transition is shown in Figure 2.

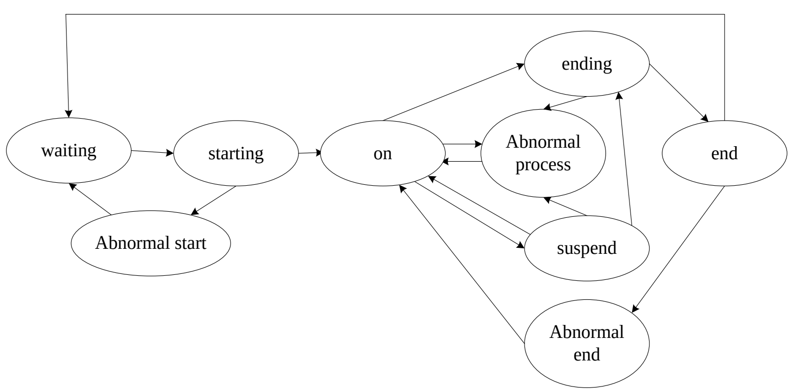


Figure 2. Output state transition.

Since internal state has more process information, analyzing previous context is avoided by this mean. This can be shown below. Internal state is related with the recognizing process, so we focus on the internal state in this paper and *waiting* (*on*, *suspend*) means internal state unless it is specifically explained.

B. Memorizing occurred context

For some of temporal relation, such as *Before* and *During*, determining whether they are satisfied depends on the previous context information. For example, if *During*(p,a), and p does not hold at time t, we can't determine whether it is normal.

Only when the activity ends and *p* has never occurred, it is determined that activity is not performed normally. In order to avoid checking previous context which may introduces high processing delay, we use some flag to memorize some special occurred context.

We use a set *b*-COND(*a*) to record the context patterns which need to hold before activity *a* is started. The initial value of *b*-COND(*a*) is *b*-COND(*a*) $=\{p \mid p \in P$, *before*(*p*,*a*) and COND(*p*,*a*)}. From the very beginning, whenever a context value input, the patterns in *b*-COND(*a*) are checked and holding patterns are deleted from the set. Normally the set is empty when the activity is stated. So the when activity *a* is recognized to start, it is abnormal if *b*-COND(*a*) is not empty.

Similarly, we use a set *d*-ACCOMP(*a*) to record the context patterns which need to hold during the time that activity *a* is performed. Let *d*-ACCOMP(*a*)$=\{p \mid p \in P$, *during*(*p*,*a*) and ACCOMPA(*p*,*a*)}. This is the initial value of *d*-ACCOMP(a). When a is recognized to be started, the patterns in the set are checked and holding patterns are deleted. When activity *a* is recognized to end, it is determined to be an abnormal end if *d*-ACCOMP is not empty.

C. Recognizing activity and its state by state transition

The recognizing process depends on the state and state transition. The basic idea is as following.

If all the premise conditions are satisfied and accompanying context patterns start to hold, it is recognized that the activity has started. Considering there may be more than one accompanying context should hold at the beginning (*Equal* or *Start* relation) and there is no real time point that these context patterns change to hold, we use state *starting* to solve the problem. When all the premise context patterns hold and at least one accompanying pattern hold, the state is converted to *starting* from *waiting*. When the start pattern is satisfied, condition checking is performed and the state is converted to *on* or *abnormal start*.

As the activity is conducted, condition is checked when any context changes and the state is converted to *abnormal process* when any of them are not satisfied. Moreover, comparing current on pattern with last moment can determine whether the activity is interrupted. If interruption occurred, the state is converted to *suspend*.

Activity end can be judged by consequence context. If consequence context patterns occurred, the state is converted to *ending* and on pattern is no longer to be checked. If there is no consequence context pattern for an activity, activity end can be judged by premise and accompanying context pattern. If some of premise and accompanying context pattern are not satisfied, state will change to *suspend* and if the state lasts long enough, it will be converted to *ending* for the activity with no consequence context pattern, and to *abnormal process* for others.

In ending state, if *e-pattern* is satisfied, the state is converted to *end*, or else, *e-pattern*(*a*,*t*)is calculated according to the current context *c*(*t*). *e-pattern*(*a*,*t*) is compared with *e-pattern*(*a*,*t*-1) to determine whether the activity is close to end. If more context patterns in end pattern occurred, i.e. *e-pattern*(*a*,*t-1*)$\subset$*e-pattern*(*a*,*t*), initialize *time-elapse* context pattern to make the state lasting for longer time. If *time-elapse* is true, it means that the state lasts too long and state is converted to abnormal process.

In end state, *d*-ACCOMP(*a*) is checked and if the set is empty, the state is converted to *waiting*(for next activity). Otherwise there exists not happened context pattern that should hold during the activity process, the state is converted to *abnormal end.*

For three types of *abnormal*, the state transition rule is waiting for a certain time and converted to *waiting*, *on* respectively.

TABLE III shows the transition rules in every state.

IV. Case Study

We use the approaches presented above in single-crystal X-ray diffraction experiment support. We focus on the context reason part, i.e. recognize the activity and its state. This is important in support environment, because the reaction of the support system, i.e. provide alerting and guiding information depends on the correctly recognized the situation.

The single-crystal X-ray diffraction procedure consists of three phases, in which activities are conducted in three places, such as Room 101, Room 102 and Room 103 in our example. The three phases are: selecting a crystal which is carried out in Room 101, analyzing the crystal which are conducted in Room 103 and structural determination which is accomplished in Room 103. There are eight activities in these phases and they should be performed as a certain sequence.

To perform every activity, different tools are needed. The location of user, location of tools as well as their states can be sensed and detected, which are used to determine the activity. There are 19 types of contexts related to the activities in experiment. TABLE IV is the context model of every activity.

Huang et al has designed same kind of system and analyze its effect in education [13]. However technology is not an important issue in that paper and activity recognition is simply realized by rule-based reason, only considering context at one time point.

Our work focuses on the technical aspects. We have developed an activity model taking the temporal logic into account, as in [11]. However, in [11], there are only equal and during relation, and some of abnormal situation cannot be recognized. These problems are partly solved in this paper.

V. Conclusion

Context interpretation in most of research is based on the assumption that high level context may result in different sensor readings or low level context, where the relation and knowledge is integrated into context model. Therefore the context interpretation is actually the hypothesis of “cause” from “consequence” which is regarded as true unless it is denied. The condition of deducing high level context is generally not sufficient. Our method is also based on this assumption and more complex because of composition of temporal relation with semantic relations. The underlying assumption of our approach is consistent with the research in the domain and consistent with intuition.

TABLE III. *STATE TRANSITION RULES*

State	state transition rule
Waiting (for activity *a*)	Input:*C*(*t*) IF $\forall p_i \in \{p \mid PREM(p,a)\}, hold\_at(p_i,t) \wedge \exists p_i \in \{p \mid ACCOMP(p,a)\}, hold\_at(p_i,t)$ THEN current state is (*a*,*starting*) ELSE current state is (*a*,*waiting*)
starting	Input:*C*(*t*) IF $\forall p_i \in s - pattern(a), hold\_at(p_i,t)$ THEN IF (*b*-COND(*a*)= ϕ) $\wedge(\forall p_i \in \{p \mid COND(p,a) \wedge \neg Before(p,a)\}, hold\_at(p_i,t))$ THEN current state is (*a*,*on*) ELSE current state is (*a*, *abnormal start*) ELSE current state is (*a*,*waiting*)
on	Input:*C*(*t*), on pattern *o-pattern*(*a*) in last "on" moment IF $\exists p_j \in \{p \mid CONSEQ(p,a)\}, hold\_at(p_j,t)$ THEN current state is (*a*,*ending*) ELSE IF $\exists p_j \in \{p \mid (During(a,p) \wedge COND(p,a))\}, \neg hold\_at(p_j,t)$ THEN current state is (*a*, *abnormal process*) ELSE BEGIN set *o-pattern*(*a*,*t*) IF *o-pattern*(*a*,*t*)⊇ *o-pattern*(*a*) THEN current state is (*a*,*on*) *o-pattern*(*a*)= *o-pattern*(*a*,*t*) ELSE current state is (*a*,*suspend*) END
suspend	Input:*C*(*t*), on pattern *o-pattern*(*a*) in last "on" moment IF $\forall p_i \in o - pattern(a), hold\_at(p_i,t)$ THEN current state is (*a*,*on*) ELSE current state is (*a*,*suspend*) Input: current time *t* IF *hold_at*(*time-elapse*,*t*) THEN IF $\{p \mid CONSEQ(p,a)\} = \varphi$ current state is (*a*,*ending*) ELSE current state is (*a*, *abnormal process*)
ending	Input:*C*(*t*), *e-pattern*(*a*,t_{i-1}) Set *e-pattern*(*a*,*t*) IF *e-pattern*(*a*,*t*)= *e-pattern*(*a*) THEN current state is (*a*,*end*) ELSE IF *e-pattern*(*a*,*t*)⊇ *e-pattern*(*a*,t_{i-1}) Initialize *time-elapse* Input: current time *t* IF *hold_at*(*time-elapse*,*t*) THEN current state is (*a*, *abnormal process*)
end	Input:*C*(*t*) THEN IF (*d*-ACCOMP(*a*)= ϕ) THEN current state is ((next(*a*),*waiting*) // next(*a*) is next possible activity of *a* ELSE current state is (*a*, *abnormal end*)
abnormal start	Input: current time *t* IF *hold_at*(*time-elapse*,*t*) THEN current state is (*a*,*waiting*)
abnormal process	Input: current time *t* IF *hold_at*(*time-elapse*,*t*) THEN current state is (*a*,*on*)
abnormal end	Input: current time *t* IF *hold_at*(*time-elapse*,*t*) THEN current state is (*a*,*on*)

TABLE IV. CONTEXT MODEL OF ACTIVITY IN SINGLE-CRYSTAL X-RAY DIFFRACTION EXPERIMENT

symbol	Activity	Temporal relation	Semantic relation
a_1	Prepare Fiber	*during*(a_1, (**User_Loc**,*Room 101*)) *start*((**Copperbar_Loc**,*Cuttingmat*),a_1) *during*((**Slicer**,*On*),a_1) *overlap*(a_1, (**Copperbar_Loc**,*Styrofoam*))	PREM((**User_Loc**,*Room 101*),a_1) ACCOMP ((**Copperbar_Loc**,*Cuttingmat*),a_1) ACCOMP ((**Slicer**,*On*),a_1) CONSEQ(a_1, (**Copperbar_Loc**,*Styrofoam*))
a_2	Check Crystal Transparency	*during*(a_2, (**User_Loc**,*Room 101*)) *during*(a_2, (**Microscope**,*On*)) *during*(a_2,(**OcularlensI_Loc**,*Microscope*),) *during*((**Adjustmentscrew**,*On*), a_2) *overlap*((**OcularlensI_Loc**,*Ocularlensbox*), a_2)	PREM((**User_Loc**,*Room 101*),a_2) PREM((**Microscope**,*On*),a_2) PREM((**OcularlensI_Loc**,*Microscope*), a_2) ACCOMP((**Adjustmentscrew**,*On*), a_2) CONSEQ((**OcularlensI_Loc**,*Ocularlensbox*) a_2)
a_3	Check Crystal Size	*during*(a_3, (**User_Loc**,*Room 101*)) *during*(a_3, (**Microscope**,*On*)) *during*(a_3 , (**OcularlensII_Loc**,*Microscope*)) *during*((**Adjustmentscrew**,*On*), a_3) *overlap*((**OcularlensI_Loc**,*Ocularlensbox*) a_3)	PREM((**User_Loc**,*Room101*),a_3) PREM((**OcularlensII_Loc**,*Microscope*) ,a_3) PREM((**Microscope**,*On*),a_2) ACCOMP((**Adjustmentscrew**,*On*), a_3) CONSEQ((**OcularlensII_Loc,** *Ocularlensbox*), a_3)
a_4	Place Crystal in Diffractometer	*during*(a_4,(**User_Loc**,*Room 102*)) *equal*((**PCI**,*On*), a_4) *during*(a_4 , (**Temperature**,[*10*,*25*)) *during*(a_4,(**Voltage**,*50KV*)) *equal*((**Copperbar_Loc**,*Diffractometer*),a_4) *during*((**Controller**,*On*), a_4) *overlap*((**Crystalcenter**,*Screencenter*),a_4)	PREM((**User_Loc**,*Room 102*),a_4) ACCOMP((**PCI**,*On*), a_4) COND((**Temperature**,[*10*,*25*) , a_4) COND((**Voltage**,*50KV*) , a_4) ACCOMP ((**Copperbar_Loc**,*Diffractometer*),a_4) ACCOMP((**Controller**,*On*), a_4) CONSEQ((**Crystalcenter**,*Screencenter*),a_4)
a_5	Diffractometer Spot Check	*during*(a_5,(**Diffractometer**,*On*)) *during*(a_5,(**User_Loc**,*Room102*)) *during*(a_5,(**Voltage**,*50KV*)) *equal*((**Copperbar_Loc**,*Diffractometer*),a_5) *overlap*(a_5,**SpotResult_Bool**,*True*))	PREM((**Diffractometer**,*On*),a_5) PREM((**User_Loc**,*Room102*),a_5) COND((**Voltage**,*50KV*) , a_5) ACCOMP ((**Copperbar_Loc**,*Diffractometer*),a_5) CONSEQ(**SpotResult_Bool**,*True*),a_5)
a_6	Lattice Constant Collection	*during*(a_6,(**Matrix**,*On*)) *during*(a_6,(**User_Loc**,*Room102*)) *during*(a_6,(***Voltage***,*20KV*)) *overlap* (a_6,(**ConstantResult_Bool**,*True*))	PREM((**Matrix**,*On*),a_6) PREM((**User_Loc**,*Room102*),a_6) COND((***Voltage***,*20KV*),a_6) CONSEQ((**ConstantResult_Bool**,*True*),a_6)
a_7	Crystal Lattice Regeneration and Revise	*during*(a_7,(**User_Loc**,*Room102*)) *equal*((**Saint**,*On*), a_7) *overlap*(a_7,(**RevisedResult_Bool**,*True*),)	PREM(**User_Loc**,*Room 102*),a_7) ACCOMP((**Saint**,*On*),a_7) CONSEQ((**RevisedResult_Bool**,*True*),a_7)
a_8	Crystal Structure Analyze	*during*(a_8,(**User_Loc**,*Room 103*)) *equal*((**Shelxt**,*On*),a_8) *overlap*((**AnalyzeResult_Bool**,*True*),a_8)	PREM((**User_Loc**,*Room 103*),a_8) ACCOMP((**Shelxt**,*On*),a_8) CONSEQ((**AnalyzeResult_Bool**,*True*),a_8)

The approach proposed in this paper is interval-based, considering the context change over a time period. This can recognized more situation than the method based on the context data in time point, but introduce high complexity in process. We define 6 temporal relations and 4 semantic relations; these are not enough to represent the relation in real world. The approach need to be further studied in future research work.

ACKNOWLEDGMENT

This research work was supported by the National Natural Science Foundation of China (Grant No. 70771017).

REFERENCES

[1] K. Dey Anind, D. Abowd Gregory and S. Daniel, "A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications", Human-Computer Interaction, vol.16, no.2, pp.97-166, 2001.

[2] W.Loke Seng, "Representing and reasoning with situations for context-aware pervasive computing: a logic programming perspective", The Knowledge Engineering Review vol.19, no.3, pp.213–233, 2005.

[3] D. Riboni, and C. Bettini, "OWL 2 modeling and reasoning with complex human activities," Pervasive and Mobile Computing, vol.7, no.3, pp.379-395, 2011.

[4] J. Lester, T. Choudhury, and N. Kern, "A hybrid discriminative /generative approach for modeling human activities," In Proceedings of the 19th International Joint Conference on Artificial Intelligence, Professional Book Center, Acapulco,Mexico, pp. 766-772, 2005.

[5] L. Liao, D. Fox, and H. Kautz, "Location-based activity recognition using relational Markov networks," In Proceedings of the 19th International Joint Conference on Artificial Intelligence, Acapulco, Mexico, pp.773-778, 2005.

[6] D.W Albrecht, and I. Zukerman, "Bayesian Models for Keyhole Plan Recognition in an Adventure Game," User Modeling and User-Adapted Interaction, vol.8, pp.5–47,1998.

[7] E.M Tapia, and S. Instille, "Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate

monitor," the 11th IEEE International Symposium on Wearable Computers, Boston,USA, pp. 37-40, October 2007.

[8] M.S. Ryoo, and J.K. Aggarwal, "Recognition of composite human activities through context-free grammar based representation," IEEE Conference on Computer Vision and Pattern Recognition, New York, USA, pp.1709-1718, June 2006.

[9] L. Chen, and C.D. Nugent, "Ontology-based activity recognition in intelligent pervasive environments," International Journal of Web Information Systems, vol.5, no.4, pp.410-430, 2009.

[10] L. Chen, C. Nugent, M. Mulvenna, D. Finaly, X. Hong, and M. Poland, "Using event calculus for behaviour reasoning and assistance in a smart home," in Proceedings of the 6th International Conference on Smart Homes and Health Telematics, IA, USA, vol. 5120, pp. 81–89, June 2008.

[11] T. Lu, S.K. Zhang and Q. Hao, "Activity Recognition in Ubiquitous Learning Environment," Journal of Advances in Information Technology, vol. 3,no.1, pp29-35, 2012.

[12] J.F. Allen, "Towards a general theory of action and time", Artificial Intelligence, vol.23, pp123-154,1984.

[13] G.J. Hwang, T.C. Yang, and C.C. Tsa, "A context-aware ubiquitous learning environment for conducting complex science experiments," Computers and Education, vol.53, no.2 , pp.402-413, 2009.

[14] M. Weiser, "The computer for 21st century," Scientific American, vol.261, no.30, pp.94-104, 1991.

[15] H.C. Chu, G.J. Hwang, S.X. Huang, and T.T. Wu, "A knowledge engineering approach to developing e-libraries for mobile learning," Electronic Library, vol.26, no.3,pp.303-317, 2008.

[16] C. Zhu, and W.H. Sheng, "Motion- and location-based online human daily activity recognition," Pervasive and Mobile Computing, vol.7, no.2, pp.256-269, 2011.

[17] H. Kautz, "A Formal Theory of Plan Recognition and its Implementation," Reasoning About Plans, San Francisco: Morgan Kaufmann, 1991, pp.69-125.

[18] R. Kowalski, and M. Sergot, "A logic-based calculus of events," New Generation Computing, vol.4, no.1, pp.67-95, 1986.

[19] O. Brdiczka, J.L. Crowley, and P. Reignier, "Learning situation models for providing context-aware services," Ambient Interaction 4th International Conference on Universal Access in Human-Computer Interaction, Beijing,China, pp.23-32, July 2007.

[20] T. Springer and A.Y. Turhan, "Employing Description Logics in Ambient Intelligence for Modeling and Reasoning about Complex Situations," Journal of Ambient Intelligence and Smart Environments, vol.1, no.3, pp.235-259, 2009.

[21] D.Q. Zhang, M.Y. Guo and J.Y. Zhou, "Context Reasoning Using Extended Evidence Theory in Pervasive Computing Environments," Future Generation Computer Systems, vol.26, no.2, pp.207-216,2010.

[22] A. Mannini and A.M. Sabatini, "Machine Learning Methods for Classifying Human Physical Activity from On-Body Accelerometers," Sensors, vol.10, no.2, pp.1154-1175, 2010.

[23] T. Gu, Z. Wu, X.P. Tao and H.K Pung, "An Emerging Patterns based Approach to Sequential, Interleaved and Concurrent Activity Recognition," In Proceeding of the 7th Annual IEEE International Conference on Pervasive Computing and Communications, Galveston, TX, pp.1-9, March 2009.

AUTHORS PROFILE

Tao Lu

She is currently an associate professor in School of Management Science and Engineering, Dalian University of Technology, China. Her main research interest is information system, knowledge-based system and pervasive computing. Email: lutao@dlut.edu.cn

Xiaoling Liu

She is currently a graduate in School of Management Science and Engineering, Dalian University of Technology, China. Her main research interest is pervasive computing. Email: liuxiaoling@mail.dlut.edu.cn

Hybrid Systems for Knowledge Representation in Artificial Intelligence

Rajeswari P. V. N.
Assoc. Professor, Dept of Comp. Sc. & Engg.
Visvodaya Technical Academy
Kavali, AP, India

Dr. T. V. Prasad
Dean of Computing Sciences
Visvodaya Technical Academy
Kavali, AP, India

***Abstract*— There are few knowledge representation (KR) techniques available for efficiently representing knowledge. However, with the increase in complexity, better methods are needed. Some researchers came up with hybrid mechanisms by combining two or more methods. In an effort to construct an intelligent computer system, a primary consideration is to represent large amounts of knowledge in a way that allows effective use and efficiently organizing information to facilitate making the recommended inferences. There are merits and demerits of combinations, and standardized method of KR is needed. In this paper, various hybrid schemes of KR were explored at length and details presented.**

Keywords- Knowledge representation; hybrid system; hybrid schema structure.

I. INTRODUCTION

An expert (knowledge based) system is a problem solving and decision making system based on knowledge of its task and logical rules or procedures for using knowledge. Both the knowledge and the logic is obtained from the experience of a specialist in the area (Business Expert). An expert system emulates the interaction a user might have with a human expert to solve a problem. The end user provides input by selecting one or more answers from a list or by entering data. The program will ask questions until it has reached a logical conclusion.

A. Knowledge Engineering

As described in [1], KR is the process of designing an expert system. It consists of three stages:

- Knowledge acquisition: The process of obtaining the knowledge from experts (by interviewing and/or observing human experts, reading specific books, etc).
- Knowledge representation: Selecting the most appropriate structures to represent the knowledge (lists, sets, scripts, decision trees, object-attribute value triplets, etc).
- Knowledge validation: Testing that the knowledge of ES is correct and complete.

B. Types of Knowledge

- Declarative: It describes what is known about a problem. This includes simple statements which are asserted to be either true or false.
- Procedural: Describes how a problem is solved. It contains rules, strategies, agendas and procedures.
- Heuristic: It describes a rule-of-thumb that helps to guides the reasoning process.
- Meta knowledge: Describes knowledge about knowledge for improve the efficiency of problem solving.
- Structural knowledge: It describes about knowledge structures. It contains rule sets, concept relationships and concept to object relationships. [2]
- Factual Knowledge: It is verifiable through experiments and formal methods,
- Tacit knowledge: It is implicit, unconscious knowledge that can be difficult to express in words or other representations form.
- Priori/Prior knowledge: It is independent of the experience or empirical evidence e.g. "everybody born before 1990 is older than 15 years"
- Posteriori/Posterior knowledge: dependent of experience or empirical evidence, as "X was born in 1990".

C. The Knowledge Representation

It is an area of AI research which is aimed at representing knowledge in symbols to facilitate inference from those knowledge elements, creating new elements of knowledge, whereas knowledge (is a progression from data to information, from information to knowledge and knowledge to wisdom) and representation (is a combination of syntax, semantics and reasoning) [3].

There are two basic components of KR i.e. reasoning and inference. In cognitive science it is concerned with how people store and process information and in AI the objective is to store knowledge so that programs can process it. [4]

D. Knowledge Representation Issues

The following are the issues to be considered regarding the knowledge representation

- Grain Size – Resolution Detail
- Scope
- Modularity
- Understandability
- Explicit Vs. Implicit Knowledge
- Procedural Vs. Declarative knowledge.

II. KNOWLEDGE REPRESENTATION TECHNIQUES

Many of the problems in AI require extensive knowledge about the world. Objects, properties, categories and relations between objects, situations, events, states and time, causes and effects are the things that AI needs to represents. KR provides the way to represent all the above defined things [5]. KR techniques are divided in to two major categories that are declarative representation and procedural representation. The declarative representation techniques are used to represents objects, facts, relations. Whereas the procedural representation are used to represent the action performed by the objects. Some of the techniques for knowledge representation are

- Bayesian Network
- Facts and Production Rules
- Semantic nets
- Conceptual Dependency
- CYC
- Frames
- Scripts
- Neural Networks

Hybrid Representation

III. HYBRID SYSTEMS

A hybrid KR system is an implementation of a hybrid KR formalism consisting of two or more different sub formalisms. These sub formalism should be integrated through (i) a representational theory, which explains what knowledge is to be-represented by what formalism, and (ii) a common semantics for the overall formalism, explaining in a semantic sound manner the relationship between expressions of different sub formalisms.[6] The generalized architecture for a hybrid system is given in Fig 1.

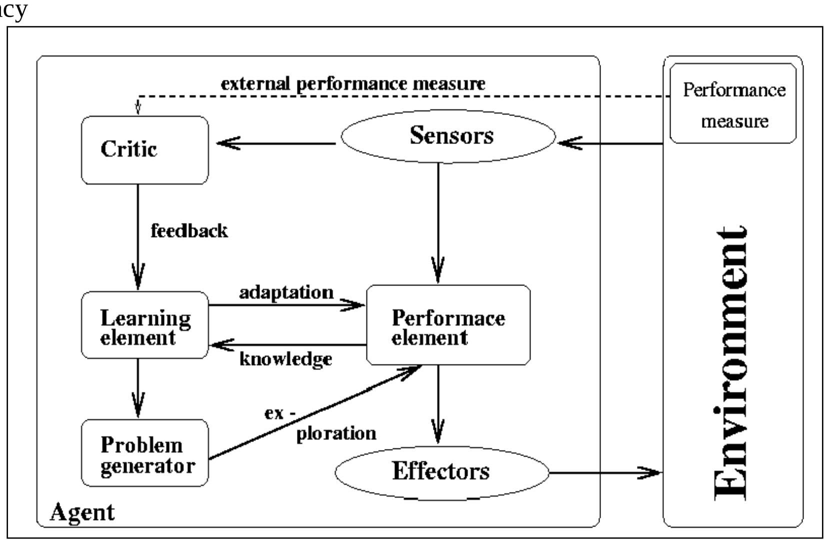


Fig 1 Generalized architecture of Hybrid system

In general these systems consist of two different kinds of knowledge: The terminological knowledge, consisting of a set of concepts and roles defining a terminology, and the assertional knowledge, consisting of some logical formalism suited to represent general assertions.

A. KRYPTON

The system consists of two modules: the Terminological Box and the Assertional Box. The terminological box, or module, is based on the KL-ONE language -a representation system based on semantic networks and frames [7]. The KRYPTON has been developed mainly from the work of KL-ONE. The difficulties in representing assertional knowledge using KL-ONE gives the idea of the integration of a theorem-prover and a KL-ONE-like language into a hybrid system. It is basically like a "tell-ask" module. All interactions between a user and a KRYPTON knowledge base are mediated by TELL and ASK operations shown in Fig 2.

The most important feature introduced by KRYPTON is the notion of a Functional Approach to knowledge representation [8]: KRYPTON is provided with a clear, implementation independent, description of what services are provided to the user. This Knowledge Level [9] description is presented in the form of a formal definition of the syntax and semantics of the languages provided by the two modules along with the interaction between these two modules.

The set of primitives of the KRYPTON language vary from one presentation to another presentation of the language. In the complete form, the terminological box includes primitives for: Concept conjunction, value and number restriction on concepts, primitive sub-concept, concept decomposition, role differentiation, role chain, primitive subrole and role decomposition. And the assertional box provides a complete first-order logic language including the usual operators: Not, and, or, exists and for all.

B. KANDOR

The basic units of KANDOR are individuals and frames. Knowledge model for KANDOR is given in Fig 3. Individuals are associated to objects in the real world and frames are associated to sets of these individuals. These units are manipulated through the standard representational structures of frames, slots, restrictions, and slot fillers common to most frame-based systems. Each slot maps individuals into sets of values, called slot fillers, Elements of these sets can be other individuals, strings, or numbers.

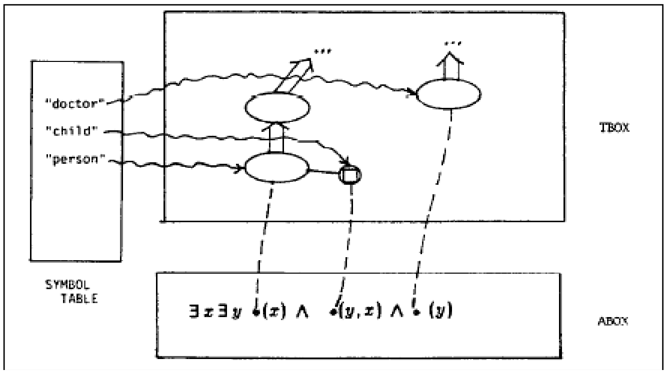


Fig. 2 Overview of KRYPTON

Frames in KANDOR have no assertion import; they look simply as descriptions of some set of individuals. There are two types of frames: Primitive and defined. To be an instance of a primitive frame, an individual must be explicitly specified as an instance of the frame when it is created.

To be an instance of a defined frame an individual must satisfy the conditions associated to the frame definition. There two types of conditions: Super-frames and restrictions. A super-frame is just another frame, and a restriction is a condition on a set of slots fillers for some slot. An individual satisfies the restriction if its slots fillers for that slot satisfy the condition.

KANDOR provides two main operations that require inferences to be made: Given an individual and a frame, determine whether the individual is an instance of the frame, arid, given two frames, it determines whether one frame is subset of another frame.

KANDOR has been used as the knowledge representation component of ARGON [10], which is an interactive information retrieval system which is designed to be used by non experts for retrieval purpose over a large, heterogeneous knowledge bases, possibly taken from a large number of sources or repositories.

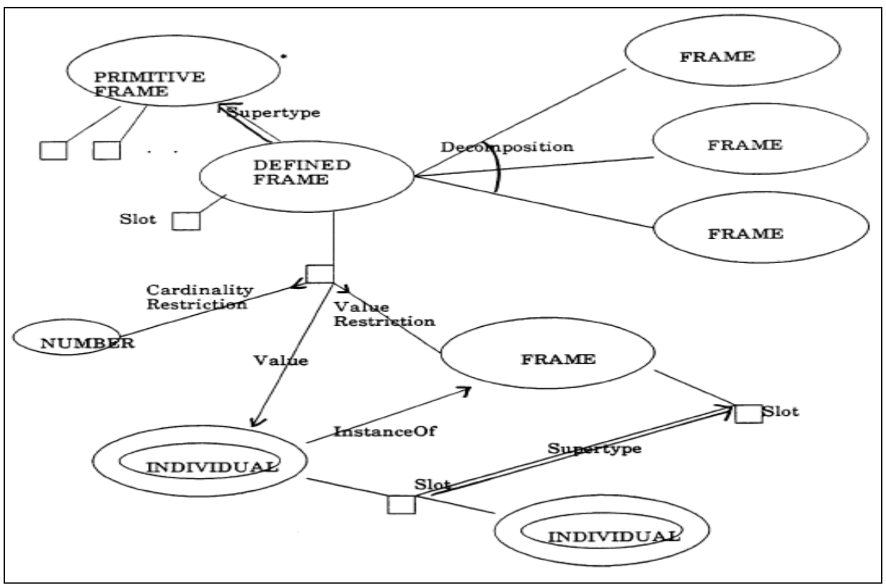


Fig 3 Knowledge model of KANDOR [11]

C. BACK

The structure of a BACK represents as the same structure of KRYPTON, which contains an terminological box and an assertional box. One main aspect in the BACK implementation is the Balancedness of the formalisms involved. Although the fact that the reasoning in hybrid systems is frequently incomplete (because of efficiency requirements) sometimes leads to situations where one formalism allows to express something which obviously should have some impact on another formalism according to the semantics of the system, the incompleteness of the reasoning precludes this impact to be realized by the system. The formalisms of this type of systems are said to be “unbalanced”.

The main criteria taken into account in the development of the BACK system [16] are the following: (i) The sub formalisms of the system should be balanced, (ii) the formalism should permit tractable inference algorithms covering almost all possible inferences, (iii) the assertional box formalism should be able to represent incomplete knowledge in a limited manner, (iv) the system should allow for extending the knowledge base incrementally (retractions are not considered) and (v) the system should reject assertional box entries which are inconsistent. The terminological language of BACK is more powerful than that of KRYPTON.

D. KL-TWO

The KL-TWO system is composed by two sub formalisms: PENNI, a modified version of the RUP (Reasoning Utility Package) system) and NIKL (New implementation of KL-ONE), a terminological reasoner in the KL-ONE [7] tradition. These two formalisms are complementary: PENNI is able to represent propositional assertions without any quantification and NIKL allows the representation of a simple class of universally quantified sentences. These sentences can be applied in PENNI to extend its propositional language with a limited form of quantification. Fig 4 shows the architecture of KL-TWO [12] .

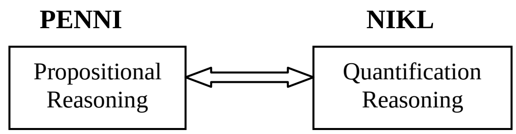


Fig 4 Architecture of KL-TWO

The PENNI formalism consists of a database of propositional assertions, more specifically, a data base of ground sentences of first-order logic without quantifiers. This database permits incremental assertions and retractions. Underlying the deductive mechanism of PENNI is a Truth Maintenance System (TMS) allowing all the useful operations that have been associated with such systems.

And the NIKL terminological reasoner allows the definition of composite concepts and roles through the use of structuring primitives and primitive concepts and roles. The primitives available in NIKL include: Concept conjunction, statement of the minimal number of role fillers, concept value restriction and role differentiation. The inference provided by NIKL is basically the sub assumption relation between concepts. It has been proved recently that the subsumption problem in NIKL is undecidable.

Two forms of hybrid reasoning are performed by the KL-TWO system: The forward reasoning, which is used to classify new assertions according to the concepts already defined in the NIKL knowledge base, and the backward reasoning, used to answer queries. Both mechanisms combine the inferences mechanism of PENNI and NIKL to perform their tasks

E. CAKE

The CAKE system was developed as a knowledge representation and reasoning facility for the Programmer's Apprentice project [13] different from the previously presented hybrid systems. CAKE does not present complementary representation formalisms in which different types of knowledge are represented, but it uses its two formalisms to represent the same knowledge.

The two formalisms present in CAKE are: A predicate calculus package which is based on the RUP (Reasoning Utility Package) system and a specialized, semantic network like formalism which is used to represent the structure of programs. This last formalism, called Plan Diagrams or simply Plans, was developed without any special concern about formal semantics but was only designed to fit the requirements of the program representation problem.

The current architecture of CAKE consists of eight layers: The bottom five layers forming the predicate calculus level and the top three layers corresponding to the Plan level. The predicate calculus layers, from bottom to top, and their functions are the following: (i) Truth Maintenance, unit propositional resolution, retraction and explanation, (ii) Equality, uniqueness of terms, congruence closure, (iii) Demons, pattern directed invocation, priority queues, (iv) Algebraic, commutativity, associativity, etc, lattices, Boolean algebras, (v) Types, type inheritance and functionality. The Plan layers are the following: (i) Plan Calculus, data and control flow graphs, abstract data types, (ii) Plan Recognition, flow graph parsing and recognition heuristics, (iii) Plan Synthesis, search of refinement space and synthesis heuristics.

F. MANTRA [14]

Developed by J. Calmet, I. A. Tjandra and G. Bittencourt in 1991, it is combination of four different knowledge representation techniques. First order logic, terminological language, semantic networks and Production systems. All algorithm used for inference are decidable because this representation used the four value logic. Mantra is a three layers architecture model. It consist the epistemological level, the logical level, Heuristic level.

Example [1]:- Ex of operation in logic level
1 command::= tell (know1edge base, Fact).
2 ask (knowledge base, Query)
3 to-frames (frame-def)
4 to-met (snet-den)
5 Fact: = to-logic (formu1a)
6 Query: = from logic (formula)
Ex of operation on terminological box
frame - def ::= identifier : c = concept | identifier:
r = relation
Concept::= (concept) | concept .

Advantages: 1 An intelligent, graphical user interface would help in building knowledge bases. 2 Support procedural knowledge. 3 A graph editor can be used t o visualize, for instance, hierarchies or terminologies that would help the user for representing expert's knowledge.

Disadvantages: Less expressive, only applicable for symbolic computation (mathematical model**).**

G. FRORL

The acronym for FRORL is Frame-and-Rule Oriented Requirement specification Language [14] which was developed by Jeffrey J. P. Tsai, Thomas Weigert and Hung-Chin Jang in 1992, and this FRORL is based on the concepts of frames and production rules which is mainly designed for software requirement and specification analysis. Six main steps for processing purpose are as follows:

1) Identify subject and themes
2) Define object frames.
3) Define object abstract inheritance relation
4) Define object attributes.
5) Identify activity frames.
6) Define actions and communication

There are two types of frames, i.e., Object frame and Activity frames. Object frames are used to represent the real world entity not limited to physical entity. These frames will act as a data structure. Each activity in FRORL are represented by activity frame to represent the changes in the world. Activity, precondition and action are reserved word not to be used in specification. Language for FRORL consists of Horn clause of predicate logic.

Advantages: 1 Modularity. 2 Incremental development. 3 Reusability. 4 Prototyping.

Disadvantages: Only limited for building prototype model for software.

H. Other Hybrid Systems

Other hybrid systems adopt a restricted version of first-order logic in their assertional module. Examples of some latest hybrid systems are

- The LOOM system is a very ambitious project developed at USC (California, USA). It includes a tern classifier, instance matcher, truth maintenance for both TBox and ABox, default reasoning, full-first-order retrieval language, pattern-driven methods, a pattern classifier, and automatic detection of inconsistency. The system can also interface with a rule-based system. Its semantics uses a three-valued semantics that is extended to seven values when defaults are included.
- The QUARK system developed at the University of Hamburg (FRG), includes a Hom clauses interpreter, and a terminological reasoner in the KL-ONE tradition. Its semantics is defined using the four-valued approach. The system is organized around nodes called Denotasional Entities (DE), and the set of facts associated to these DEs, called aggregates. The aggregates correspond to the frame entities in other systems, and are organized into a network.
- The CLASSIC system is a direct descendent of the KANDOR system, and shares all functionalities with this system. The goal of a CLASSIC hybrid system is to extend the expressive power of KANDOR's terminological language while remaining tractable. Along with the functionalities of the KANDOR system, it includes
  (a) a construct to allow equalities between role fillers,
  (b) a set construct to allow one to say a slot is filled by an individual of one of a set of different frames,
  (c) a test-defined construct which allows one to test membership in classes by a user defined test, and
  (d) a limited form of rules which allow one to say that once something is found to be an instance of one concept, then it is an instance of another concept. The system also allows host concepts, such as integers, strings, and all Common Lisp structures.

The Comparison between different hybrid systems [15] is presented in Table 1.

IV. CONCLUSION

Different KR schemes are used in AI, which differ in terms of semantics, structure and flexibility in level of power of expression. Combination of two or more representation schemes, which is known as Hybrid Systems may be used for making the system more efficient and improving the knowledge representation. Different hybrid systems are discussed with their corresponding architectures and also presented a comparative data in terms of modules (Assertional, Terminology), Formal semantics and Domain of Applications.

REFERENCES

[1] Dr. Roman V Belavkin, “Knowledge Representation”, Lecture notes in Computer Science, BIS3226.2002

[2] Dr. Habil.sc.ing, “Knowledge representation and networked schemes”, a Lecture notes of Fundamentals of artificial intelligence, 2008

[3] Dr. Tarek Helmy El-Basuny, “Knowledge Representation Methods”. Principles of Artifiicial Intelligence, ICS-381.

[4] Tanwar Poonam, Prasad T.V., Aswal Mahendra. S., “Comparative Study of Three Declarative Knowledge representation Techniques”, Int’l Jour. on Comp. Sc. & Engg. (IJCSE), Vol. 02, No. 07, pp. 2274-2281, 2010

[5] Tanwar Poonam, Prasad T.V., Aswal Mahendra. S., “Comparative Study of Three Declarative Knowledge Representation Techniques”, Int. J. on Comp. Sc. and Engg, Vol. 02, No. 07, 2010

[6] Dan Geiger, David Heckerman, “Knowledge representation and inference in similarity networks and Bayesian multinets”, Artificial Intelligence 82, 45-74, 1996.

[7] Brachman, R.J. and Schmolze, J.G.,” An overview of the KL-ONE Knowledge Representation System”, Cognitive Science, Vol. 9, No.2, pp. 171-216, 1985

[8] Levesque, H.J., “A Logic of Implicit and Explicit Belief.” Proceedings of AAAI 84, pp. 198-202, 1984

[9] Newell, “A Physical Symbol Systems”, Cognitive Science”, Vol. 4, pp. 135-183, 1980

[10] Patel-Schneider, P.F., “Small Can Be Beautiful in Knowledge Representation”, Proc. of the IEEE Workshop on Principles of Knowledge-Based Systems", pp. 11-19, Denver, 1984

[11] Jintae Lee, “Knowledge Base Integration: What we can learn from Database Integration Research”, A.I Memo No:1011, 1988.

[12] Marc Vilain, “The Restricted Language Architecture for Hybrid Representation System”, BBN Laboratories, Proc. of the 9th Int. Conf. on Artificial Intelligence, 1985

[13] Rich, C., Shrobr, H.E. and Waters, R.C., “An Overview of the Programmer's Apprentice”, Proc. of IJCAI 6, 1979.

[14] Tanwar Poonam, Prasad T.V., Datta Kamlesh, “An Effective Knowledge base system Architecture and issues in representation techniques”, Int. Jour. of Advancements in Technology, Vol 2, No 3, July 2011

[15] Zur Erlangung des akademischen Grades eines', “An Architecture for Hybrid Knowledge Representation”, Tag der mundlichen Priifung, 1990

[16] Nebel, B. and Luck, K. Von, “Issues of Integration and Balancing in Hybrid Knowledge Representation Systems”, Proceedings of the 11th German Workshop on Artificial Intelligence, K. Morik (Editor), Geseke, pp. 115-123, 1985.

AUTHORS PROFILE

Ms. P. V. N. Rajeswari received her Masters degree in Computer Sc. & Engg. from Sam Higgin Bothams University, Allahabad in 2006. She is currently associated with the Dept. of Comp. Sc. & Engg. at Visvodaya Technical Academy, Kavali, AP, India as Associate Professor. She has about 7 years of teaching experience at under graduate and graduate level. Her areas of interest are network security, artificial intelligence, mobile ad-hoc networks, data mining, etc.

Dr. T. V. Prasad has over 17 years of experience in industry and academics. He has received his graduate and master's degree in Computer Science from Nagarjuna University, AP, India. He was with the Bureau of Indian Standards, New Delhi for 11 years as Scientist/Deputy Director. He earned PhD from Jamia Millia Islamia University, New Delhi in the area of computer sciences/ bioinformatics. He has worked as Head of the Department of Computer Science & Engineering, Dean of R&D and Industrial Consultancy and then as Dean of Academic Affairs at Lingaya's University, Faridabad. He is with Visvodaya Technical Academy, Kavali as Dean of Computing Sciences. He has lectured at various international and national forums on subjects related to computing.

Prof. Prasad is a member of IEEE, IAENG, Computer Society of India (CSI), and life member of Indian Society of Remote Sensing (ISRS) and APBioNet. His research interests include bioinformatics, consciousness studies, artificial intelligence (natural language processing, swarm intelligence, robotics, BCI, knowledge representation and retrieval). He has over 73 papers in different journals and conferences, and has authored six books and two chapters.

TABLE 1 COMPARISON OF HYBRID SYSTEMS

	Assertional Module	**Terminology Module**	**Other Modules**	**Formal Semantics**	**Domain of Application**
KRYPTON	Full first order predicate logic	KL-ONE like	-	Model Theoretic	Natural Language
KANDOR	Frame like schema	KL-ONE like	-	Model Theoretic	-
BACK	Object oriented language	KL-ONE like	-	Model Theoretic	Natural Language
KL-1WO	Variable free predicate logic	KL-ONE like	-	Model Theoretic	-
CAKE	Truth maintenance system	-	Plan diagrams	Mapping into logic	Programming Language
LOOM	Truth maintenance system	Variable-free algebra	Production systems	Seven-valued semantics	Natural Language
QUARK	Horn classes	KL-ONE like	-	Four-valued semantic	Natural Language
CLASSIC	Frame like schema	KL-ONE like	-	Model Theoretic	Prototype application : Wine Choice
DRL	Prolog	Many sorted theory	-	Prolog semantics	Logic Programming
KRAPFEN	Network of propositions	KL-ONE like	Proto type module	Not provided	Natural Language
MANTRA	Decidable first order logic	KL-ONE like	Semantic net and production systems	Four-valued semantic	Mathematical Knowledge

Local Feature based Gender Independent Bangla ASR

Bulbul Ahamed
Senior Lecturer
Northern University Bangladesh,
Bangladesh

Khaled Mahmud
Lecturer Institute of Business
Administration, University of Dhaka
Bangladesh

B.K.M. Mizanur Rahman
Lecturer United International
University Dhaka,
Bangladesh

Foyzul Hassan
Senior S Q A Engineer
Enosis Solutions
Dhaka, Bangladesh

Rasel Ahmed
Lecturer Dhaka Residential
Model College
Bangladesh

Mohammad Nurul Huda
Associate Professor United
International University Dhaka,
Bangladesh

Abstract— **This paper presents an automatic speech recognition (ASR) for Bangla (widely used as Bengali) by suppressing the speaker gender types based on local features extracted from an input speech. Speaker-specific characteristics play an important role on the performance of Bangla automatic speech recognition (ASR). Gender factor shows adverse effect in the classifier while recognizing a speech by an opposite gender, such as, training a classifier by male but testing is done by female or vice-versa. To obtain a robust ASR system in practice it is necessary to invent a system that incorporates gender independent effect for particular gender. In this paper, we have proposed a Gender-Independent technique for ASR that focused on a gender factor. The proposed method trains the classifier with the both types of gender, male and female, and evaluates the classifier for the male and female. For the experiments, we have designed a medium size Bangla (widely known as Bengali) speech corpus for both the male and female.The proposed system has showed a significant improvement of word correct rates, word accuracies and sentence correct rates in comparison with the method that suffers from gender effects using. Moreover, it provides the highest level recognition performance by taking a fewer mixture component in hidden Markov model (HMMs).**

Keywords- Automatic speech recognition; Local featues; gender factor; word correct rates; word accuracies; sentence correct rates; hidden Markov model.

I. Introduction

Various methods were proposed to obtain robust automatic speech recognition (ASR) system; however, the ASR system that shows enough performance at any time and everywhere could not be realized now. One of the reasons is that the acoustic models (AMs) of an HMM-based classifier include many hidden factors such as speaker-specific characteristics that include gender types and speaking styles [1]-[3]. It is difficult to recognize speech affected by these factors, especially when an ASR system comprises only a classifier that made its training by a single type of gender.

One solution is to employ a acoustic model for both types of gender. Though the robustness of this acoustic model by utilizing the both gender specific characteristicis limited, but it resolves the gender effects more precisely.

On the other hand, only a very few works have been done in ASR for Bangla (can also be termed as Bengali) in spite of one of the largely spoken languages in the world. More than 220 million people speak in Bangla as their native language. It is ranked seventh based on the number of speakers [4]. A major difficulty to research in Bangla ASR is the lack of proper speech corpus. Some efforts are made to develop Bangla speech corpus to build a Bangla text to speech system [5]

However, this effort is a part of developing speech databases for Indian Languages, where Bangla is one of the parts and it is spoken in the eastern area of India (West Bengal and Kolkata as its capital). But most of the natives of Bangla (more than two thirds) reside in Bangladesh, where it is the official language. Although the written characters of Standard Bangla in both the countries are same, there are some sounds that are produced variably in different pronunciations of Standard Bangla, in addition to the myriad of phonological variations in non-standard dialects [6]. Therefore, there is a need to do research on the main stream of Bangla, which is spoken in Bangladesh, ASR. Some developments on Bangla speech processing or Bangla ASR can be found in [7]-[14]. For example, Bangla vowel characterization is done in [7]; isolated and continuous Bangla speech recognition on a small dataset using hidden Markov models (HMMs) is described in [8]. Again, Bangla digit recognition was found in [15]. Before us, there was no Bangla ASR system that incorporates gender specific characteristics, but our proposed method was based on Standard mel frequency cepstral coefficients (MFCCs) and consequently, it suffers from lower performance in the recognition stage [16].

In this paper, we have constructeda Gender-Independent (GI) ASR by utilizing the acoustic features [17], local features for suppressing the gender-factor up to a particular level. Here, the proposed technique trains the classifier with the both types of gender, male and female, and evaluates the classifier for the male and female. For the experiments, we have designed a medium size Bangla speech corpus for both the male and female.The proposed system has showed a significant improvement of word correct rates, word accuracies and sentence correct rates in comparison with the method that suffers from gender effects. Since the local features incorporate frequency and time domain information, it shows significant improvement of recognition performance over the method based on MFCCs at fewer mixture components. Moreover, it requires a fewer mixture component in hidden Markov model (HMMs) and hence, computation time.

This paper is organized as follows. Sections II discusses Bangla phoneme schemes, Bangla speech corpus and triphone model. On the other hand, Section III and IV outline mel frequency cepstral coefficients (MFCCs) and Local features (LFs) extraction procedure, respectively and Section V explains the proposed GI-based technique. Section VI describes an experimental setup, and section VII analyzes experimental results. Finally, section VIII concludes the paper with some future remarks.

II. BANGLA PHONEME SCHEMES,TRIPHONE DESIGN AND BANGLA SPEECH CORPUS

Bangla phonetic scheme and IPA (International Phonetic Alphabet) for Bangla were described in [16]. The paper [16] also showed characteristics of some Bangla words by using the spectrogra and triphone model based on HMM were also analyzed for Bangla words

At present, a real problem to do experiment on Bangla phoneme ASR is the lack of proper Bangla speech corpus. In fact, such a corpus is not available or at least not referenced in any of the existing literature. Therefore, we develop a medium size Bangla speech corpus, which is described below.

Hundred sentences from the Bengali newspaper “Prothom Alo” [18] are uttered by 30 male speakers of different regions of Bangladesh. These sentences (30x100) are used as male training corpus (D1). On the other hand, 3000 same sentences uttered by 30 female speakers are used as female training corpus (D2).

On the other hand, different 100 sentences from the same newspaper uttered by 10 different male speakers and by 10 different female speakers are used as male test corpus (D3) and female test corpus (D4), respectively. All of the speakers are Bangladeshi nationals and native speakers of Bangla. The age of the speakers ranges from 20 to 40 years. We have chosen the speakers from a wide area of Bangladesh: Dhaka (central region), Comilla – Noakhali (East region), Rajshahi (West region), Dinajpur – Rangpur (North-West region), Khulna (South-West region), Mymensingh and Sylhet (North-East region). Though all of them speak in standard Bangla, they are not free from their regional accent.

Recording was done in a quiet room located at United International University (UIU), Dhaka, Bangladesh. A desktop was used to record the voices using a head mounted close-talking microphone. We record the voice in a place, where ceiling fan and air conditioner were switched on and some low level street or corridor noise could be heard.

Jet Audio 7.1.1.3101 software was used to record the voices. The speech was sampled at 16 kHz and quantized to 16 bit stereo coding without any compression and no filter is used on the recorded voice.

III. MFCC FEATURE EXTRACTOR

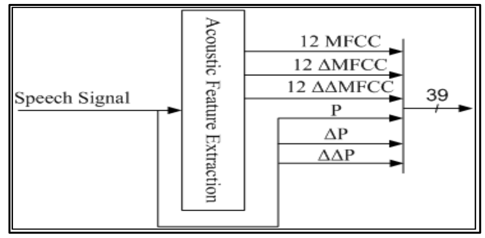


Figure 1. MFCC feature extraction

Conventional approach of ASR systems uses MFCCof 39 dimensions (12-MFCC, 12-ΔMFCC, 12-ΔΔMFCC, P, ΔP and ΔΔP, where P stands for raw energy of the input speech signal) and the procedure of MFCC feature extraction is shown in Fig.1. Here, hamming window of 25 ms is used for extracting the feature. The value of pre-emphasis factor is 0.97.

IV. LOCAL FEATURE EXTRACTOR

At the acoustic feature extraction stage, the input speech is first converted into LFs that represent a variation in spectrum along the time and frequency axes. Two LFs, which are shown in Fig. 2, are then extracted by applying three-point linear regression (LR) along the time (t) and frequency (f) axes on a time spectrum pattern (TS), respectively. Fig. 3 exhibits an example of LFs for an input utterance. After compressing these two LFs with 24 dimensions into LFs with 12 dimensions using discrete cosine transform (DCT), a 25-dimensional (12 Δt, 12 Δf, and ΔP, where P stands for the log power of a raw speech signal) feature vector called LF is extracted.

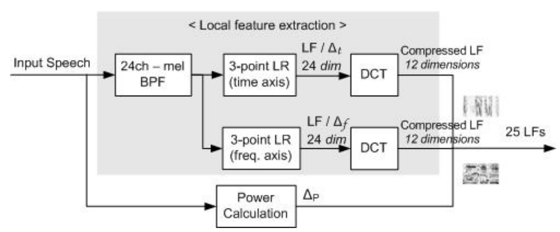


Fig.2 LFs extraction procedure.

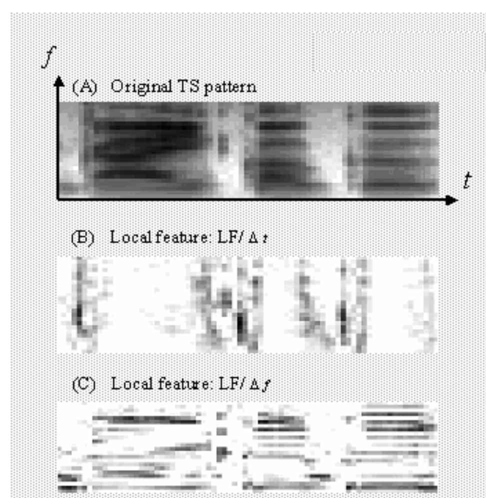


Fig. 3 Examples of LFs.

V. Proposed LF-Based GI ASR System

Fig. 4 shows the proposed LF-based GI ASR system for Bangla Language. Here, an input speech is converted into LFs of 25 dimensions (12 Δt, 12 Δf, and ΔP, where P stands for the log power of a raw speech signal) at the acoustic feature extraction stage, which is described in Section IV. Then, this extracted LFs (data set based on both male and female) of gender independent characteristics are used to train the GI classifier based on triphone HMM, while the Viterbi algorithm is used for evaluating the test data set for male and female.

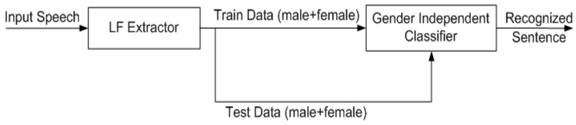


Fig. 4 The Proposed LF-based GI ASR System.

VI. Experimental Setup

The frame length and frame rate are set to 25 ms and 10 ms (frame shift between two consecutive frames), respectively, to obtain acoustic features (MFCCs) and local features (LFs) from an input speech. MFCCs and LFs comprised of 39 and 25 dimensional feature vectors, respectively.

For designing an accurate continuous word recognizer, word correct rate (WCR), word accuracy (WA) and sentence correct rate (SCR) for (D3+D4) data set are evaluated using an HMM-based classifier. The D1 (male) and D2 (female) data sets are used to design Bangla triphones HMMs with five states, three loops, and left-to-right models. Input features for the classifier are 39 dimensional MFCCs and 25 dimensional LFs. In the HMMs, the output probabilities are represented in the form of Gaussian mixtures, and diagonal matrices are used. The mixture components are set to 1, 2, 4 and 8.

For evaluating the performance of different methods including the proposed method, we have designed the following experiments:

Experiment-I [Exp-I]

(a) MFCC (Train: 3000 male, Test: 1000 male + 1000 female).
(b) LF(Train: 3000 male, Test: 1000 male + 1000 female).

Experiment-II [Exp-II]

(c) MFCC (3000 female, Test: 1000 male + 1000 female).
(d) LF(3000 female, Test: 1000 male + 1000 female).

Experiment-III [Exp-III]

(e) MFCC (Train: 3000 male + 3000 female, Test: 1000 male + 1000 female).
(f) LF(Train: 3000 male + 3000 female, Test: 1000 male + 1000 female) [Proposed].

VII. Experimental Result And Analysis

Figure 5 shows sentence correct rates for MFCC and LF-based ASR using the mixture component one, where total numbers of input sentences were 2000. From the figure it is shown that, LF-based ASR provides higher sentence correct rate over all the experiments evaluated by MFCC-based ASR.

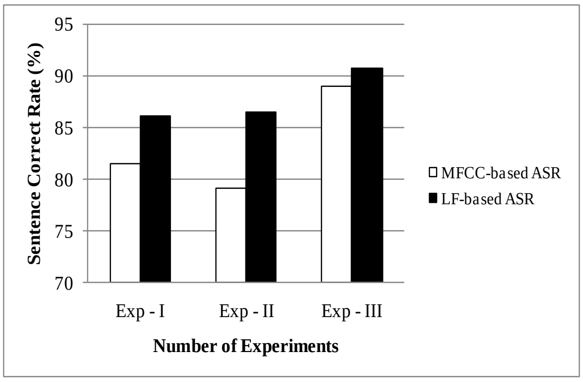


Fig.5 Sentence Correct Rates for MFCC and LF-based ASR.

It is noted that, MFCC-based method provides 81.45%, 79.05% and 88.90% SCRs for the experiments I, II and III, respectively, while corresponding experiments of LF-based method generates 86.10%, 86.45% and 90.65%, respectively On the other hand, experiment III, which is done by gender independent condition, provides significant improvement of SCR over the experiments I and II that are gender dependent. For an example, GI LF-based method (experiment III (f)) shows 90.65% SCR that is significant improvement in comparison with the values, 86.10% and 86.45% which are provided by experiments I(b) and II(d). The reason for the better results exhibited by the LF-based method is the incorporation of frequency and time domain information in the input features, where the MFCC-based method only includes time domain features. Moreover, the GI LF-based method (experiment III(f)) gives better result over experiments I(b) and II(d) because training of HMM-based classifier for GI LF-based method embeds both male and female voices.

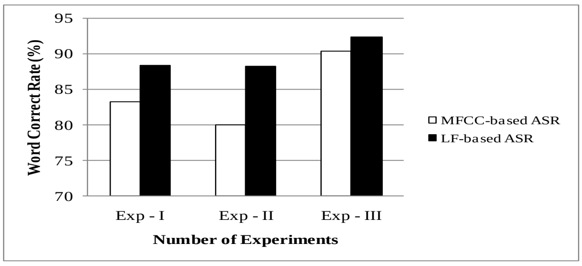


Fig. 6 Word Correct Rates for MFCC and LF-based ASR.

WCRs for the experiments I, II and III using the MFCC-based and LF-based methods are shown in Figure 6 using the mixture component one. From the experiments, it is exhibited that the LF-based methods provides higher word correct rates than the MFCC-based methods. Maximum improvement is shown in the Exp-II. On the other hand, the highest correctness is provided by the LF-based method for Exp-III, where gender-independent training was performed.

On the other hand, Figure 7 depicts the WAs for the experiments I, II and III using the MFCC-based and LF-based methods for the mixture component one. From the experiments, it is exhibited that the LF-based methods provides higher word accuracies than the MFCC-based methods. Exp-II exhibits the maximum improvement. Moreover, the highest level accuracy is generated by the LF-based method for Exp-III, where training was done by incorporating the male and female data sets.

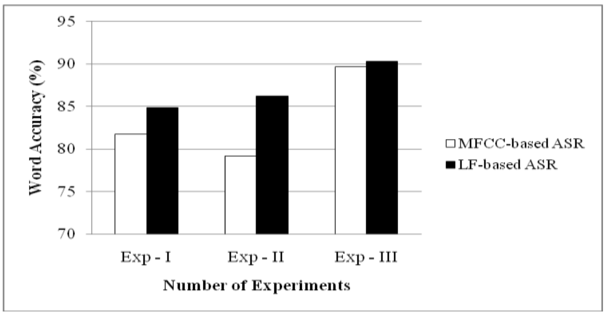


Fig. 7 Word Accuracy for MFCC and LF-based ASR.

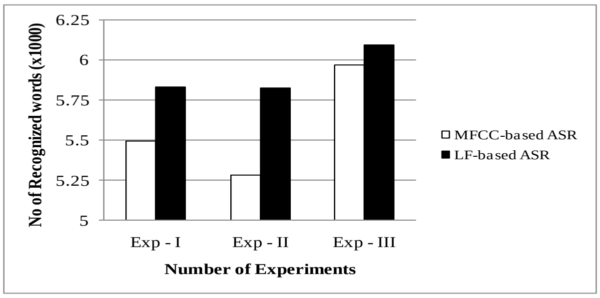


Fig. 8 No. of correctly recognized words for MFCC and LF-based ASR.

Again, the number of correctly recognized words out of 6600 input words is shown in Figure 8. From the figure, it is observed that the LF-based method in Exp-III recognizes the highest number of input words. Besides, the highest improvement by the LF-based method over the method based on MFCC is shown in Exp-II.

Tables 1 and 2 show the speech recognition Performance for the Exp-I where MFCC and LF-based methods for the mixture components 1, 2, 4 and 8 are investigated. Here, training and testing are done by using the D1 (male) and D3+D4 (male and female) speech corpora, respectively. For all the mixture components in the Table 1, LF-based method shows higher word correct rate, word accuracy and sentence correct rate in comparison with the method that incorporated MFCCs as input feature. It may be mentioned that the mixture component one provides the highest level performance among the entire mixture component investigated.

From the Table 2, it is exhibited that the methods incorporating LFs show the higher number of sentence recognition and the highest number at mixture component one compared to the counterpart.

Tables 3 and 4 generate the performance of similar pattern for the female dependent training in Exp-II. On the other hand, Tables 5 and 6 exhibit the gender independent performance where training and testing are done in the gender independent environment in Exp-III. It is claimed from the Tables 1, 3 and 5 that the LF-based method provides the higher recognition performance for all the mixture components.

Besides, the proposed LF-based method in Exp-III provides the higher performance among the three experiments and outputs the highest recognition performance for all the investigated mixture components. Among the experimented mixture components, the best result is achieved in component one.

On the other hand, from the Tables 2, 4 and 6 it is observed that the proposed method recognized the highest number of sentences.

Table I: Speech Recognition Performance for Exp-I using MFCC and LF-based methods using the mixture components 1, 2, 4 and 8. Training and testing are done by using the D1 and (D3+D4) speech corpora, respectively.

Mixture Components	Methods	Recognition Performance (%)		
		Word Correct Rate	Word Accuracy	Sentence Correct Rate
1	MFCC-based	83.20	81.71	81.45
	LF-based	88.32	84.85	86.10
2	MFCC-based	82.71	81.26	81.25
	LF-based	87.82	84.39	85.95
4	MFCC-based	78.02	77.26	77.20
	LF-based	86.79	83.24	84.65
8	MFCC-based	68.05	67.59	67.25
	LF-based	86.85	83.65	84.80

Table II: Word Recognition Performance for Exp-I using MFCC and LF-based methods using the mixture components 1, 2, 4 and 8. Training and testing are done by using the D1 and (D3+D4) speech corpora, respectively.

Mixture Components	Methods	Sentence recognition performance (out of 2000)		Word recognition performance (out of 6600)			
		Correctly recognized Sentence, H	Substitution, S	Correctly recognized Words, H	Deletion, D	Substitution, S	Insertion, I
1	MFCC-based	1629	371	5491	240	869	98
	LF-based	1722	278	5829	54	717	229
2	MFCC-based	1625	375	5459	264	877	96
	LF-based	1719	281	5796	57	747	226
4	MFCC-based	1544	456	5149	419	1032	50
	LF-based	1693	307	5728	67	805	234
8	MFCC-based	1345	655	4491	734	1375	30
	LF-based	1696	304	5732	71	797	211

Table III: Speech Recognition Performance for Exp-II using MFCC and LF-based methods using the mixture components 1, 2, 4 and 8. Training and testing are done by using the D2 and (D3+D4) speech corpora, respectively.

Mixture Components	Methods	Recognition Performance (%)		
		Word Correct Rate	Word Accuracy	Sentence Correct Rate
1	MFCC-based	79.94	79.14	79.05
	LF-based	88.20	86.23	86.45
2	MFCC-based	83.45	82.62	82.35
	LF-based	85.48	83.38	83.65
4	MFCC-based	80.33	79.65	79.20
	LF-based	84.05	82.09	82.35
8	MFCC-based	71.11	70.70	70.30
	LF-based	80.20	77.91	78.50

Table IV: Word Recognition Performance for Exp-II using MFCC and LF-based methods using the mixture components 1, 2, 4 and 8. Training and testing are done by using the D2 and (D3+D4) speech corpora, respectively.

Mixture Components	Methods	Sentence recognition performance (out of 2000)		Word recognition performance (out of 6600)			
		Correctly recognized Sentence, H	Substitution, S	Correctly recognized Words, H	Deletion, D	Substitution, S	Insertion, I
1	MFCC-based	1581	419	5276	322	1002	53
	LF-based	1729	271	5821	94	685	130
2	MFCC-based	1647	353	5508	239	853	55
	LF-based	1673	327	5642	145	813	139
4	MFCC-based	1584	416	5302	380	918	45
	LF-based	1647	353	5547	165	888	129
8	MFCC-based	1406	594	4693	661	1246	27
	LF-based	1570	430	5293	203	1104	151

Table V: Speech Recognition Performance for Exp-III using MFCC and LF-based methods using the mixture components 1, 2, 4 and 8. Training and testing are done by using the (D1+D2) and (D3+D4) speech corpora, respectively.

Mixture Components	Methods	Recognition Performance (%)		
		Word Correct Rate	Word Accuracy	Sentence Correct Rate
1	MFCC-based	90.36	89.67	88.90
	LF-based	92.27	90.30	90.65
2	MFCC-based	89.59	88.76	87.95
	LF-based	91.09	88.53	89.25
4	MFCC-based	91.23	90.53	89.85
	LF-based	91.50	89.18	89.50
8	MFCC-based	91.26	90.73	90.15
	LF-based	91.03	88.88	89.30

Table VI: Word Recognition Performance for Exp-III using MFCC and LF-based methods using the mixture components 1, 2, 4 and 8. Training and testing are done by using the (D1+D2) and (D3+D4) speech corpora, respectively.

Mixture Components	Methods	Sentence recognition performance (out of 2000)		Word recognition performance (out of 6600)			
		Correctly recognized Sentence, H	Substitution, S	Correctly recognized Words, H	Deletion, D	Substitution, S	Insertion, I
1	MFCC-based	1778	222	5964	123	513	46
	LF-based	1813	187	6090	40	470	130
2	MFCC-based	1759	241	5913	120	567	55
	LF-based	1785	215	6012	44	544	169
4	MFCC-based	1797	203	6021	100	479	46
	LF-based	1790	210	6039	37	524	153
8	MFCC-based	1803	197	6023	127	450	35
	LF-based	1786	214	6008	43	549	142

VIII. CONCLUSION

This paper has proposed a gender independent automatic speech recognition technique for Bangla language by inputting local features. The following information concludes the paper.

i) The methods based on local features provide a higher speech recognition performance than the method that incorporates the standard MFCCs for all the experimented mixture components.

ii) For the LF-based methods, the mixture component one generates the highest level performance.

iii) The proposed LF-based gender independent method has showed the significant improvement of word correct rate, word accuracy and sentence correct rate in comparison with the methods that are experimented in gender dependent environments.

In future, the authors would like to incorporate neural network based systems in gender independent for evaluating the performance.

REFERENCES

[1] S. Matsuda, T. Jitsuhiro, K. Markov and S. Nakamura, "Speech Recognition system Robust to Noise and Speaking Styles," Proc. ICSLP'04, Vol.IV, pp.2817-2820, Oct. 2004.

[2] M. A. Hasnat, J. Mowla, and Mumit Khan, "Isolated and Continuous Bangla Speech Recognition: Implementation Performance and application perspective, " in Proc. International Symposium on Natural Language Processing (SNLP), Hanoi, Vietnam, December 2007.

[3] R. Karim, M. S. Rahman, and M. Z Iqbal, "Recognition of spoken letters in Bangla," in Proc. 5th International Conference on Computer and Information Technology (ICCIT02), Dhaka, Bangladesh, 2002.

[4] http://en.wikipedia.org/wiki/List_of_languages_by_total_speakers, Last accessed April 11, 2009.

[5] S. P. Kishore, A. W. Black, R. Kumar, and Rajeev Sangal, "Experiments with unit selection speech databases for Indian languages," Carnegie Mellon University.

[6] S. A. Hossain, M. L. Rahman, and F. Ahmed, "Bangla vowel characterization based on analysis by synthesis," Proc. WASET, vol. 20, pp. 327-330, April 2007.

[7] A. K. M. M. Houque, "Bengali segmented speech recognition system," Undergraduate thesis, BRAC University, Bangladesh, May 2006.

[8] K. Roy, D. Das, and M. G. Ali, "Development of the speech recognition system using artificial neural network," in *Proc. 5th International Conference on Computer and Information Technology (ICCIT02)*, Dhaka, Bangladesh, 2003.

[9] M. R. Hassan, B. Nath, and M. A. Bhuiyan, "Bengali phoneme recognition: a new approach," in *Proc. 6thInternationalConference on Computer and Information Technology (ICCIT03)*, Dhaka, Bangladesh, 2003.

[10] K. J. Rahman, M. A. Hossain, D. Das, T. Islam, and M. G. Ali, "Continuous bangla speech recognition system," in*Proc.6th International Conference on Computer and Information Technology (ICCIT03)*, Dhaka, Bangladesh, 2003.

[11] S. A. Hossain, M. L. Rahman, F. Ahmed, and M. Dewan, "Bangla speech synthesis, analysis, and recognition: an overview," in *Proc. NCCPB*, Dhaka, 2004.

[12] S. Young, et al, The HTK Book (for HTK Version. 3.3),Cambridge UniversityEngineeringDepartment,2005.http:///htk.eng.cam.ac.uk/prot-doc/ktkbook.pdf.

[13] http://en.wikipedia.org/wiki/Bengali_script, Last accessed April 11, 2009.

[14] C. Masica, *The Indo-Aryan Languages*, CambridgeUniversity Press, 1991.

[15] Ghulam Muhammad, Yousef A. Alotaibi, andMohammad Nurul Huda, "Automatic SpeechRecognition for Bangla Digits," ICCIT'09, Dhaka,Bangladesh, December 2009.

[16] Foyzul Hasan, Rokibul Alam Kotwal, Saiful Alam Khan and Mohammad Nurul Huda, **"**Gender Independent Bangla Automatic Speech Recognition," IEEE/IAPR Internationaol Conference on Informatics, Electronics and Vision (ICIEV) 2012, May 2012, Dhaka, Bangladesh .

[17] T. Nitta, "Feature extraction for speech recognition based on orthogonal acoustic-feature planes and LDA," Proc. ICASSP'99, pp.421-424, 1999.

[18] Daily Prothom Alo. Online: www.prothom-alo.com.

AUTHORS PROFILE

Bulbul Ahamed was born in Munshiganj, Bangladesh in 1982. He obtained his B. Sc. in Computer Science and Engineering and MBA (major in MIS &Marketing) from Northern University Bangladesh. Now he is pursuing his M.Sc. in Computer Science and Engineering at United International University, Bangladesh. He is now working as Senior Lecturer in Northern University Bangladesh. His research interests include Speech Recognition, Artificial Intelligence, Neural Network and Business. He has published his articles in different journals of Pakistan, Dubai and Bangladesh.

B.K.M. Mizanur Rahman was born in Jhenaidah, Bangladesh in 1972. He completed his B.Sc. in Electrical and Electronic Engineering Degree from BUET, Dhaka, Bangladesh. He is a student of Masters in Computer Science and Engineering at United International University, Dhaka, Bangladesh. He is now working as a Lecturer in the Department of Electrical and Electronic Engineering of the same university. His research interests include Speech Recognition, Digital Signal Processing and Renewable Energy.

Rasel Ahmed was born in Shariatpur, Bangladesh in 1983. He completed his Bachelor in Computer Science from National University, Bangladesh. Now he is pursuing his M.Sc. in Computer Science and Engineering at United International University, Bangladesh. He is now working as Lecturer in the Dhaka Residential Model College.

His research interests include Speech Recognition, Artificial Intelligence, and Information Technology.

Khaled Mahmud was born in 1984 at Pabna, Bangladesh. He was graduated from Bangladesh University of Engineering and Technology (BUET) in Computer Science and Engineering. He had his MBA (Marketing) from Institute of Business Administration, University of Dhaka. He was awarded gold medals both in his secondary and higher secondary school level for excellent academic performance. He is a Fulbright Scholar, now pursuing his MBA at Bentley University, Massachusetts, USA. He previously worked as Assistant Manager in Standard Chartered Bank. He has research interest business, technology, e-learning, e-governance, human resource management and social issues. He has his articles published in journals and conferences of USA, Canada, Australia, United Arab Emirates, Malaysia, Thailand, South Korea, India and Bangladesh.

Foyzul Hassan was born in Khulna, Bangladesh in 1985. He completed his B.Sc. in Computer Science and Engineering Degree from Military Institute of Science and Technology (MIST), Dhaka, Bangladesh in 2006. He has participated several national and ACM Regional Programming Contests. He is currently doing M. Sc. in CSE in United International University, Dhaka, Bangladesh and also has been working as a Senior Software Quality Assurance Engineer at Enosis Solutions, Dhaka, Bangladesh. His research interests include Speech Recognition, Robotics and Software Engineering.

Mohammad Nurul Huda was born in Lakshmipur, Bangladesh in 1973. He received his B. Sc. and M. Sc. in Computer Science and Engineering degrees from Bangladesh University of Engineering & Technology (BUET), Dhaka in 1997 and 2004, respectively. He also completed his Ph. D from the Department of Electronics and Information Engineering, Toyohashi University of Technology, Aichi, Japan. Now, he is working as an Associate Professor in United International University, Dhaka, Bangladesh. His research fields include Phonetics, Automatic Speech Recognition, Neural Networks, Artificial Intelligence and Algorithms. He is a member of International Speech Communication Association (ISCA).

A Mechanism of Generating Joint Plans for Self-interested Agents, and by the Agents

Wei HUANG
School of Computer Science and Technology
University of Science and Technology of China
Hefei, China

***Abstract*— Generating joint plans for multiple self-interested agents is one of the most challenging problems in AI, since complications arise when each agent brings into a multi-agent system its personal abilities and utilities. Some fully centralized approaches (which require agents to fully reveal their private information) have been proposed for the plan synthesis problem in the literature. However, in the real world, private information exists widely, and it is unacceptable for a self-interested agent to reveal its private information. In this paper, we define a class of multi-agent planning problems, in which self-interested agents' values are private information, and the agents are ready to cooperate with each other in order to cost efficiently achieve their individual goals. We further propose a semi-distributed mechanism to deal with this kind of problems. In this mechanism, the involved agents will bargain with each other to reach an agreement, and do not need to reveal their private information. We show that this agreement is a possible joint plan which is Pareto optimal and entails minimal concessions.**

Keywords- joint plan; self-interested agent; privat information; Pareto optimality; concession.

I. Introduction

In the context of multi-agent planning, it is common that agents have different capabilities and they usually have to cooperate each other if they want to achieve their goals [1, 2, 3]. In many real settings, agents are self-interested, i.e., they may consider personal goals and utilities. Cooperation means finding a common plan which can increase agents' personal net benefits. Fully centralized approaches for finding such common plans in subsets of multi-agent planning problems have been proposed in the literature (see [4, 5] for instances). In these approaches, it is assumed that each agent reveals its personal goals and utilities to the other agents or to an arbitrator.

However, individualism often leads to existence of private information [6]. That is to say, in many real world multi-agent environments, self-interested agents may take advantage of knowledge about other agents' key information. Consequently, the agents don't want to reveal private information each other, and fully centralized planning approaches are not appropriate. On another hand, in order to perform an action successfully or deal with inconsistence of agents' preferences over actions, agents often need to negotiate with each other. For example, a glass bottle will be broken if there are multiple robots that are going to clasp it at the same time. In cases like this, fully distributed planning approaches (see [2, 7] for examples) are also not appropriate.

So it is significant to deal with agents' capabilities (for cooperation and competition) and private information (for individualism) together in multi-agent planning. In this paper, we present an attempt to solve the problem. We define a rich class of planning problems in which **(1)** self-interested agents' values on joint plans are private information; **(2)** agents are ready to cooperate with each other in order to cost efficiently achieve their individual goals; **(3)** each agent can invite other agents to join a plan that is still beneficial to itself by providing certain amount of side payment [8]; and **(4)** each agent always persists in its personal goals even side payments are very attractive.

For this kind of problems, we provide a semi-distributed multi-agent planning mechanism (**MAPM**). In **MAPM**, each agent plays a role in the selection of the final joint plan. Agents will bargain with each other by giving joint plans (a joint plan includes a plan and a side payment function), and do not need to reveal their goals and utilities.

In traditional bargaining situations [9], the set of utility vectors (to the set of agents) derived form all possible proposals, is assumed to be compact and convex. In a planning domain, this assumption may be problematic, since the set of joint plans interesting to an agent is usually finite. In addition, each agent φ's utility function on joint plans, often takes integer values, since φ cannot value a joint plan more closely than to the nearest penny. So in the bargaining situation discussed in this paper, only applicable joint plans and side payment functions taking integer values are considered. We show that **MAPM** always terminates; and the agreement reached by the agents through **MAPM**, is a joint plan which is Pareto optimal and entails minimal concessions.

The paper is organized as follows. Firstly, we give some preliminaries for representing planning problems. Secondly, we introduce the bargaining situation considered in **MAPM**. Thirdly, we define **MAPM** and prove its properties. Finally we discuss related work and future research directions.

II. Planning Domains And Problems

The system we are interested to plan on is a multi-agent dynamic domain, modeling the potential evolutions of world.

Definition 1: A multi-agent planning domain is a tuple $\mathcal{D}=\langle \mathcal{S}, s_0, N, A, \rho \rangle$, where $\mathcal{S}$ is the finite set of domain states,

$s_0 \in \mathcal{S}$ is the initial state, N={1,2,…,n} denotes the set of agents, A is the finite set of domain actions, $\rho \subseteq \mathcal{S} \times N \times A \times \mathcal{S}$ is the domain transition relation.

At each time point, $\mathcal{D}$ is in one of its states; initially, s_0. Agent φ can perform action a on state s if $\langle s, \varphi, a, s' \rangle \in \rho$ for some s'. In this paper, we assume that all state transitions are deterministic, i.e., $(\forall \langle s, \varphi, a \rangle \in \mathcal{S} \times N \times A) |\{s' | \langle s, \varphi, a, s' \rangle \in \rho\}| \leq 1$ holds. All actions are assumed to be asynchronous. That is to say, at each time point, there is at most one agent that is going to perform an action. Therefore a plan can be formalized as a sequence of agent-action pairs.

Definition 2: A plan π over domain $\mathcal{D}=\langle \mathcal{S}, s_0, N, A, \rho \rangle$ is a finite sequence in the form $\langle \varphi_1, a_1 \rangle, \langle \varphi_2, a_2 \rangle, \ldots, \langle \varphi_m, a_m \rangle$ such that each $\langle \varphi_i, a_i \rangle \in N \times A$. ε denotes the empty sequence. π is applicable if there exist $s_1, s_2, \ldots, s_m \in \mathcal{S}$ such that $\langle s_{i-1}, \varphi_i, a_i, s_i \rangle \in \rho$ for $0<i\leq m$. m and $s_0;s_1;\ldots;s_m$ are called the length and the path of π, respectively.

In many realistic settings, agents are self-interested, have private goals and costs, and are motivated to act to increase their private net benefit. To capture such settings, for each agent φ, we associate a set of goal states, a cost function on actions, and the reward φ associates with the set of goal states. The set of goal states, cost function, and reward are assumed to be φ's private information (i.e., they are only observable to φ itself). Formally:

Definition 3: A planning problem is a tuple $\mathcal{P}=\langle \mathcal{D}, g, c, r, \delta \rangle$, where **(1)** $\mathcal{D}=\langle \mathcal{S}, s_0, N, A, \rho \rangle$ is a multi-agent planning domain; **(2)** g: $N \rightarrow 2^{\mathcal{S}} \backslash \varnothing$ is a goal function that assigns each agent a set of goal states; **(3)** c: $N \times A \rightarrow Z^+$ is a cost function that specifies the cost of action execution for each agent; **(4)** r: $N \rightarrow Z^+$ is a function capturing the reward each agent associates with its own goal states; and **(5)** $\delta \in Z^+$, and only plans bounded in length by δ are taken into consideration.

g(φ), c_φ (It is required that $c_\varphi(a)=c(\varphi,a)$ for all $a \in A$), and r(φ) are agent φ's private information that can not be revealed to other agents. So we freely interchange notations φ.g and g(φ),φ.c and c_φ, φ.r and r(φ).

Given a planning problem $\mathcal{P}$, let $\Omega(\mathcal{P})$ denote the set of all applicable plans considered in $\mathcal{P}$. Suppose φ is an agent and $\pi=\langle \varphi_1, a_1 \rangle, \langle \varphi_2, a_2 \rangle, \ldots, \langle \varphi_m, a_m \rangle \in \Omega(\mathcal{P})$ such that $s_0;s_1;\ldots;s_m$ is the path of π. The utility of π to φ is

$$u_\varphi(\pi) = r' - \sum_{i=1}^{m} c_i' \tag{1}$$

where $m \leq \delta$; r'=φ.r if $s_m \in$ φ.g, 0 otherwise; and $c'_i = \varphi.c(a_i)$ if $\varphi=\varphi_i$; 0 otherwise.

Example 1: A small planning domain $\mathcal{D}=\langle \mathcal{S}, s_0, N, A, \rho \rangle$ is depicted in Fig. 1, where $\mathcal{S}=\{s_0,s_1,s_2,s_3\}$, N={1,2,3}, A={a,b}, and $\langle s_0,2,a,s_1 \rangle, \langle s_0,3,a,s_1 \rangle, \ldots \in \rho$. It is easy to find that $\pi=\langle 3,b \rangle; \langle 2,a \rangle$ is an applicable plan over $\mathcal{D}$. Let $\mathcal{P}=\langle \mathcal{D}, g, c, r, 3 \rangle$ be a planning problem, where g, c, r are given in Table 1. So $\pi \in$

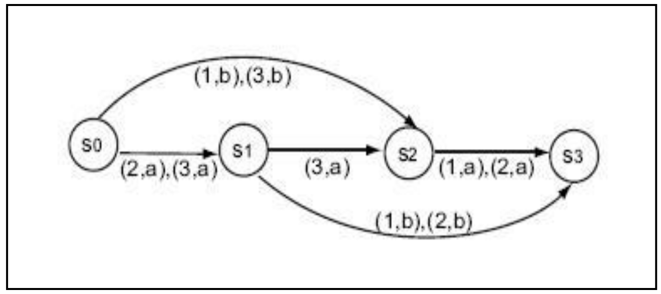


Figure 1. A small multi-agent planning domain $\mathcal{D}$.

TABLE I. Agents' Goals, Costs, and Rewards

φ∈N	g(φ)	c(φ,a)	c(φ,b)	r(φ)
1	$\{s_1,s_3\}$	4	6	12
2	$\{s_2,s_3\}$	2	2	3
3	$\{s_3\}$	5	3	6

$\Omega(\mathcal{P})$, $u_1(\pi)=12$, $u_2(\pi)=1$, and $u_3(\pi)=3$.

III. Bargaining Situation

Let $\mathcal{P}$ be a planning problem and N={1,2,…,n} be the set of agents in $\mathcal{P}$. Suppose $\varphi \in N$; $\pi, \pi' \in \Omega(\mathcal{P})$; and $u_\varphi(\pi) > u_\varphi(\pi')$. Then agent φ would prefer π to π'. If there is some $\varphi' \in N$ preferring π' to π, then φ can propose a side payment such the amount is not greater than $u_\varphi(\pi) - u_\varphi(\pi')$. If this proposal does not work, then φ must abandon π and consider π' instead. A joint plan can be seen as a structured contract, and consists of a plan and a side payment function.

Definition 4: A joint plan in $\mathcal{P}$ is a pair $p=\langle \pi,\zeta \rangle$, where $\pi \in \Omega(\mathcal{P})$ and $\zeta: N \rightarrow Z$ is a side payment function satisfies $\sum_{\varphi \in N} \zeta(\varphi)=0$. The utility of p to agent φ is

$$u_\varphi(p) = u_\varphi(\pi) + \zeta(\varphi) \tag{2}$$

In order to reach an agreement (i.e., a joint plan accepted by all the agents in N), the agents can bargain with each other by proposing joint plans. Once an agreement $p=\langle \pi,\zeta \rangle$ is reached, all the agents will cooperate to perform π, and the gross utility, i.e., $\sum_{\varphi \in N} u_\varphi(\pi)$ will be redistributed among N such that each agent φ's real income is $u_\varphi(p)$. If the agents fail to reach agreement, then an agent $\varphi \in N$ would be selected randomly to control the evolution of the domain. In this case, we assume that the other agents take the "null action", i.e., do not interfere with the evolution.

Suppose d_φ is the maximal value of utility φ can achieve without other agent's involvement, i.e.,

$$d_\varphi = \max_{\pi \in \Omega(P)} \{u_\varphi(\pi) \,|\, Agents(\pi) \subseteq \{\varphi\}\} \tag{3}$$

Where Agents(π) denotes the set of agents appear in π. Then $\lfloor d_\varphi/|N| \rfloor$ ($\lfloor\,\rfloor$ and $\lceil\,\rceil$ denote the ceil and the floor function on real numbers.) acts as φ's utility on the disagreement event, denoted by u_φ^d. In other words, φ is not willing to cooperate with other agents if the cooperation can not bring to φ a utility value which is strictly greater than u_φ^d (individual rationality). It is not difficult to find that, without access to other agents' private information, each φ∈N can compute (eg., by a backward breadth-first search) $\prod_\varphi=\{\pi\in\Omega(\mathcal{P})|\ u_\varphi(\pi)>u_\varphi^d\}$, i.e., the set of plans interesting for φ. The set $\Omega^\perp(\mathcal{P})$ of individual rational plans is defined as

$$\Omega^\perp(P)=\{\pi\in\Omega(P)\,|\,(\forall\varphi\in N)u_\varphi(\pi)>u_\varphi^d\} \quad (4)$$

It is easy to find that $\Omega^\perp(\mathcal{P})=\cap_{\varphi\in N}\prod_\varphi$. We use $u^\perp_\varphi$ and u^T_φ to denote the minimal utility and the maximal utility φ can gain in the situation where all agents are individual rational, respectively:

$$u_\varphi^\perp=\min_{\pi\in\Omega^\perp(P)}u_\varphi(\pi);u_\varphi^T=\max_{\pi\in\Omega^\perp(P)}u_\varphi(\pi) \quad (5)$$

Indeed $u^\perp_\varphi$ is φ's bottom line for bargaining, and u^T_φ is the ideal outcome of φ. $\Lambda(\mathcal{P})$ denotes the set of all possible joint plans. Joint plan $p=\langle\pi,\zeta\rangle\in\Lambda(\mathcal{P})$ if and only if **(1)** $\pi\in\Omega^\perp(\mathcal{P})$, and **(2)** $u_\varphi(p)\geq u^\perp_\varphi$ for each φ∈N.

$(u^T_1,\ldots,u^T_n)$ is called the ideal point. For a joint plan p,

$$\Delta(p)=\sum_{\varphi\in N}(u_\varphi^T-u_\varphi(p))^2 \quad (6)$$

describes the distance between the ideal point and the utility vector derived from p. In other words, Δ(p) describes the concessions made by the agents to achieve p. This leads to the notion of solution which characterizes the Pareto optimal joint plans which entail minimal concessions.

Definition 5: Joint plan p is a solution to $\mathcal{P}$ if it satisfies: **(1)** $p\in\Lambda(\mathcal{P})$, **(2)** there is no joint plan $p'\in\Lambda(\mathcal{P})$ such that $u_\varphi(p')>u_\varphi(p)$ for all φ∈N, and **(3)** $\Delta(p)=\min_{p'\in\Lambda(\mathcal{P})}\Delta(p')$.

This definition states that all self-interested agents should be individual rational at first (i.e., each φ∈N will not commit to a joint plan p if $u_\varphi(p)<u^\perp_\varphi$. Second, a solution should be a Pareto optimal joint plan, in which no agent can increase its utility without decreasing other agents' utility. Finally, among the possible joint plans, a solution should entail minimal concessions.

Example 2: See Example 1. We can find that $u_1^d=u_2^d=u_3^d=$ =0. So $\pi\in\Omega^\perp(\mathcal{P})$. Let $p=\langle\pi,\zeta\rangle$ be a joint plan such that ζ(1)=-2 and ζ(2)= ζ(3)=1. Then $u_1(p)=10$, $u_2(p)=2$, and $u_2(p)=4$. In fact, $u^T_1=12$, $u^T_2=3$, $u^T_3=6$, $\Delta(p)=2^2+1^2+2^2=9$, and p is a solution to $\mathcal{P}$.

Each φ∈N can compute $\prod_\varphi$. And $\Omega^\perp(\mathcal{P})=\cap_{\varphi\in N}\prod_\varphi=\{\pi_i\,|1\leq i\leq 8\}$, where π_1=⟨1,b⟩; ⟨1,a⟩, π_2=⟨1,b⟩; ⟨2,a⟩, π_3=⟨3,b⟩; ⟨1,a⟩, π_4=⟨3,b⟩; ⟨2,a⟩, π_5=⟨3,a⟩; ⟨1,b⟩, π_6=⟨3,a⟩; ⟨2,b⟩, π_7=⟨2,a⟩; ⟨1,b⟩, π_8=⟨2,a⟩; ⟨3,a⟩; ⟨1,a⟩. Agents' values are shown in Table 2, where u^i_φ denotes $u_\varphi(\pi_i)$.

TABLE II. Agents' Values on Each Plan

φ	u^1_φ	u^2_φ	u^3_φ	u^4_φ	u^5_φ	u^6_φ	u^7_φ	u^8_φ
1	2	6	8	12	6	12	6	8
2	3	1	3	1	3	1	1	1
3	6	6	3	3	1	1	6	1

IV. MECHANISM FOR GENERATING JOINT PLANS

In this section, we present **MAPM** a semi-distributed mechanism of generating joint plans for multiple self-interested agents, in which all the involved agents will bargain on the possible joint plans, and each agent's utility function and goal keep secret to other agents. **MAPM** is defined as follows.

Step1 Each φ∈N calculates $\prod_\varphi=\{\pi|u_\varphi(\pi)>u_\varphi^d\}$, and sends $\prod_\varphi$ to an arbitrator φ*∉N.

Step2 φ* calculates $\Omega^\perp(\mathcal{P})=\cap_{\varphi\in N}\prod_\varphi$. If $\Omega^\perp(\mathcal{P})=\varnothing$, then the result is **failure**, and the process stops. Otherwise, φ* sets N'← ∅, $\prod\leftarrow\Omega^\perp(\mathcal{P})$, **d**[π][φ]←∞, and **c**[φ]←0 for all φ∈N and $\pi\in\Omega^\perp(\mathcal{P})$, and sends $\Omega^\perp(\mathcal{P})$ to each φ∈N.

Step3 Each φ∈N puts the plans of $\Omega^\perp(\mathcal{P})$ in a sequence $\mathbf{sq}_\varphi=\pi_1;\pi_2;\ldots$ such that $u_\varphi(\pi_i)\geq u_\varphi(\pi_{i+1})$ for $|\Omega^\perp(\mathcal{P})|>i\geq1$, and sets $\mathbf{con}_\varphi\leftarrow0$. Let t←0.

Step4 φ* sends ``Next proposal!'' to each φ∈N-N'.

Step5 Let t←t+1. For each φ∈N-N':

- If $\mathbf{con}_\varphi=0$ then: φ sets $p_\varphi^t\leftarrow$Head($\mathbf{sq}_\varphi$); $\mathbf{sq}_\varphi\leftarrow$Tail($\mathbf{sq}_\varphi$); if $\mathbf{sq}_\varphi\neq\varepsilon$ then φ sets $con_\varphi\leftarrow u_\varphi(p_\varphi^t)$-$u_\varphi$(Head($\mathbf{sq}_\varphi$)).[1]
- Otherwise φ sets $p_\varphi^t\leftarrow$hold, and $\mathbf{con}_\varphi\leftarrow\mathbf{con}_\varphi$-1.

Step6 Each φ∈N-N' sends p_φ^t to φ*. If p_φ^t=hold then φ* sets **c**[φ]←**c**[φ]+1. Otherwise φ* sets $\mathbf{d}[p_\varphi^t][\varphi]\leftarrow\mathbf{c}[\varphi]$. Let $ps_\varphi\leftarrow\{\pi|\mathbf{d}[\pi][\varphi]\neq\infty\}$. φ* sets N' ←$\{\varphi|ps_\varphi=\Omega^\perp(\mathcal{P})\}$, $\omega\leftarrow\cap_{\varphi\in N}ps_\varphi$. If ω=∅, then goto step 4.

Step7 φ* selects a plan π from ω such that $\sum_{\varphi\in N}\mathbf{d}[\pi][\varphi]\leq\sum_{\varphi\in N}\mathbf{d}[\pi'][\varphi]$ for all π'∈ω, sets π*←π, $\theta\leftarrow\sum_{\varphi\in N}\mathbf{d}[\pi^*][\varphi]$, and $\prod\leftarrow\{\pi|\ \sum_{\varphi\in N}\min(\mathbf{d}[\pi][\varphi],\mathbf{c}[\varphi])<\theta\}$. If ∏≠∅ then goto step 4. φ* sets M←N.

Step8 Let $N''\leftarrow\{\varphi\in M\cap N'|\mathbf{c}[\varphi]<\lceil\theta/|M|\rceil\}$. φ* sets M←M-N'', $\theta\leftarrow\theta-\sum_{\varphi\in N''}\mathbf{c}[\varphi]$, M'←Stop(M, θ, **c**)(see Fig. 2), and ζ[φ]←**d**[π*][φ]-**c**[φ] for each φ∈N''. If M⊆M'∪N' then goto step 12.

[1] Given a nonempty sequence **sq**, Head returns the first item of **sq**, and Tail returns a sequence **sq'** such that **sq**=Head (**sq**)•**sq'**. For example, let **sq**=$q_1;q_2;q_3$, then Head(**sq**)=q_1, and Tail (**sq**)=$q_2;q_3$.

Step9 φ^* sends “Next proposal!” to each $\varphi \in M-(M' \cup N')$.

Step10 Let $t \leftarrow t+1$. For each $\varphi \in M-(M' \cup N')$:

- If $\mathbf{con}_\varphi=0$ then: φ sets $p_\varphi^t \leftarrow \mathrm{Head}(\mathbf{sq}_\varphi)$; $\mathbf{sq}_\varphi \leftarrow \mathrm{Tail}(\mathbf{sq}_\varphi)$; if $\mathbf{sq}_\varphi \neq \varepsilon$ then φ sets $\mathrm{con}_\varphi \leftarrow u_\varphi(p_\varphi^t)-u_\varphi(\mathrm{Head}(\mathbf{sq}_\varphi))$.
- Otherwise φ sets $p_\varphi^t \leftarrow \mathrm{hold}$, and $\mathbf{con}_\varphi \leftarrow \mathbf{con}_\varphi-1$.

Step11 Each $\varphi \in M-(M' \cup N')$ sends p_φ^t to φ^*. If $p_\varphi^t=\mathrm{hold}$ then φ^* sets $\mathbf{c}[\varphi] \leftarrow \mathbf{c}[\varphi]+1$. Otherwise φ^* sets $\mathbf{d}[p_\varphi^t][\varphi] \leftarrow \mathbf{c}[\varphi]$. Let $\mathrm{ps}_\varphi \leftarrow \{\pi | \mathbf{d}[\pi][\varphi] \neq \infty\}$. φ^* sets $N' \leftarrow \{\varphi | \mathrm{ps}_\varphi=\Omega^\perp(\mathcal{P})\}$. Goto step 8.

Step12 φ^* sets $\theta \leftarrow \theta-\sum_{\varphi \in M-M'}\mathbf{c}[\varphi]$, $\zeta[\varphi] \leftarrow \mathbf{d}[\pi^*][\varphi]-\mathbf{c}[\varphi]$ for each $\varphi \in M-M'$, and $\zeta[\varphi] \leftarrow \mathbf{d}[\pi^*][\varphi]-\lfloor \theta/|M'| \rfloor$ for each $\varphi \in M'$. Let $k \leftarrow \theta-|M'|\lfloor \theta/|M'| \rfloor$. φ^* selects k agents (denoted as $\varphi'_1,\ldots,\varphi'_k$) from M' randomly, and sets $\zeta[\varphi'_i]:=\zeta[\varphi'_i]-1$ for $1 \le i \le k$. φ^* announces the result of the procedure is $\langle \pi^*,\zeta \rangle$, and the process stops.

Observe this mechanism. We can find that, for all $\varphi \in N$: **(1)** φ does not communicate with other agents in N directly; **(2)** φ gets no information about other agents' proposals from φ^* during the course of bargaining; and **(3)** φ^* only knows $u_\varphi(\pi)-u_\varphi(\pi')$ if π and π' have been proposed by φ. So in the course of bargaining, no agent $\varphi \in N$ can extract information that would allow it to infer something about other agents' private information. In addition, for all $\varphi \in N$ and $\pi,\pi' \in \Omega^\perp(\mathcal{P})$, the arbitrator φ^* does not know: **(1)** $u_\varphi(\pi)$ (and of course, also φ.g, φ.c, and φ.r), and **(2)** $u_\varphi(\pi)-u_\varphi(\pi')$ if π or π' has not been proposed by φ.

Consider the planning problem $\mathcal{P}$ depicted in Example 1. We apply **MAPM** to $\mathcal{P}$. Firstly, each $\varphi \in N$ calculates $\prod_\varphi$, and φ^* finds that $\Omega^\perp(\mathcal{P})=\{\pi_i | 8 \ge i \ge 1\}$, where each π_i is given in Example 2 (agents' values on these plans are shown in TABLE II). And then, agents in N begin to bargain. The relevant data generated by **MAPM** is illustrated in TABLE III, where $\prod'=\{\pi_1, \pi_2, \pi_4, \pi_6, \pi_7\}$, **h** and Ω denote hold and $\Omega^\perp(\mathcal{P})$, respectively. Lastly, the procedure stops at time t=9, and φ^* announces the result is $p=\langle \pi_4, \zeta \rangle$, where $\langle \zeta(1), \zeta(2), \zeta(3) \rangle = \langle -1,0,1 \rangle$ ($\langle -2,1,1 \rangle$, or $\langle -2,0,2 \rangle$). So $\langle u_1(p), u_2(p), u_3(p) \rangle = \langle 11,1,4 \rangle$ ($\langle 10,2,4 \rangle$, or $\langle 10,1,5 \rangle$).

In fact, agent $\varphi \in N$ proposes a side payment (i.e., makes a concession) such the amount is 1 based on p_φ^t, if it sends $p_\varphi^{t+1}=\mathrm{hold}$ to φ^*. Suppose $\pi, \pi' \in \Omega^\perp(\mathcal{P})$ such that $u_\varphi(\pi) > u_\varphi(\pi')$, and φ sends π and π' to φ^* at time t and t', respectively. Let $C=|\{t'>t''>t| \, p_\varphi^{t''}=\mathrm{hold}\}|$. In **MAPM**, it is required that t<t' and $C=u_\varphi(\pi)-u_\varphi(\pi')$. Please note that we do not aim at dominant-strategy truthful mechanisms in which private information does not influence the final result. In this paper, it is assumed that private information has value to agents, and to keep it secrete is each agent's responsibility.

Consequently, we aim at mechanisms in which any agent is not sure that lying is better than truth telling if she does not know other agents' private information. Now let us see what would happen if φ deviates from **MAPM**:

- If t>t' then it is possible that π will be discarded and φ will suffer loss. For example, see TABLE II, TABLE III and suppose that agent 1 proposes π_3 instead of π_4 at t=1. Then φ^* will announce that $p'=\langle \pi_3,\zeta' \rangle$ is the result such that $\zeta'(1)=-1$. It is easy to find that $u_1(p')=7<u_1(p)$.
- If $C<u_\varphi(\pi)-u_\varphi(\pi')$ then it is possible that π' will be chosen and the concession made by φ will be underestimated (see $\mathbf{d}[\pi^*][\varphi]$ at step 8 and step 12). For example, see TABLE II, TABLE III and suppose that agent 3 proposes π_3 instead of hold at t=4. Then φ^* will announce that $p'=\langle \pi_3,\zeta' \rangle$ is the result such that $\zeta'(3)=-1$ or -2. It is easy to find that $u_3(p')=2$ or $1< u_3(p)$.
- If $C>u_\varphi(\pi)-u_\varphi(\pi')$ then it is possible that π will be chosen and the benefit gained by φ will be overestimated (see $\mathbf{c}[\varphi]$ and $\lfloor \theta/|N| \rfloor$ at step 8 and step 12). For example, see TABLE II, TABLE III and suppose that agent 3 insists on π_1 (i.e., $p_3^1=\pi_1$ and $p_3^t=\mathrm{hold}$ for all t>1). Then φ^* will announce that $p'=\langle \pi_1,\zeta' \rangle$ is the result such that $\zeta'(3)=-4$. It is easy to find that $u_3(p')=2< u_3(p)$.

Consequently, the agents in N will follow **MAPM**, even though they are self-interested. We now show some properties of **MAPM**. The first key result states that MAPM always terminates.

```
subroutine Stop(M, θ, c)
  M'←M;
  while M'≠∅
    mc←min_{φ∈M'} c[φ];
    if mc∗|M'|≥θ then break;
    M'←{φ∈M'| c[φ]>mc};
  return M';
```

Figure 2. Stop subroutine.

TABLE III. Data Generated by **MAPM** for the Example

t	p_1^t	p_2^t	p_3^t	ω	π^*	$\prod$	M	M’	θ
1	π_4	π_1	π_1	$\varnothing$		Ω			
2	π_6	π_3	π_2	$\varnothing$		Ω			
3	h	π_5	π_7	$\varnothing$		Ω			
4	h	h	h	$\varnothing$		Ω			
5	h	h	h	$\varnothing$		Ω			
6	h	π_2	h	$\varnothing$		Ω			
7	π_3	π_4	π_3	$\{\pi_3\}$	π_3	$\prod'$			
8	π_8	π_6	π_4	$\{\pi_3, \pi_4\}$	π_4	$\{\pi_1\}$			
9	h	π_7	h	$\{\pi_3, \pi_4\}$	π_4	$\varnothing$	N	N	5

Theorem 1: **MAPM** is guaranteed to terminate at some time $T \le |\Omega^{\perp}(\mathcal{P})| + \max_{\varphi \in N}(u^T_{\varphi} - u^{\perp}_{\varphi})$.

Proof. Observe **MAPM**. It is easy to find that computation of every step of **MAPM** always terminates. Pick any $\varphi \in N$, $\pi \in \Omega^{\perp}(\mathcal{P})$, and $t \ge 1$. Then:

1. if **MAPM** does not stop at time t, then there exists $\varphi' \in N$ sending a joint plan to φ^* at t;
2. if $\{\pi' \in \Omega^{\perp}(\mathcal{P}) | \varphi$ has sent π' before $t\} = \Omega^{\perp}(\mathcal{P})$ then φ will not send any joint plan to φ^* at any $t' > t$;
3. if $t \ge |\{\pi' \in \Omega^{\perp}(\mathcal{P}) | u_{\varphi}(\pi') \ge u_{\varphi}(\pi)\}| + u^T_{\varphi} - u_{\varphi}(\pi)$ then there exists $1 \le t' \le t$ such that φ sends π at t'.

According to items 1-3, we can find that **MAPM** stops at some $T \le |\Omega^{\perp}(\mathcal{P})| + \max_{\varphi \in N}(u^T_{\varphi} - u^{\perp}_{\varphi})$. □

The second property states that if there is a solution for $\mathcal{P}$, then **MAPM** will not fail.

Proposition 1: **failure** is the result of **MAPM** if and only if there is no solution to $\mathcal{P}$.

Proof. Observe **MAPM** and we can find that **failure** is the result of **MAPM** if and only if $\Omega^{\perp}(\mathcal{P}) = \varnothing$. Now we show that $\Omega^{\perp}(\mathcal{P}) = \varnothing$ if and only if there is no solution to $\mathcal{P}$.

Obviously, if $\Omega^{\perp}(\mathcal{P}) = \varnothing$ then there is no solution to $\mathcal{P}$.

Suppose $\Omega^{\perp}(\mathcal{P}) \ne \varnothing$. Then there exists $\pi \in \Omega^{\perp}(\mathcal{P})$ such that $(\forall \pi' \in \Omega^{\perp}(\mathcal{P})) \sum_{\varphi \in N} u_{\varphi}(\pi') \le \sum_{\varphi \in N} u_{\varphi}(\pi)$. Let $P = \{\langle \pi', \zeta \rangle \in \Lambda(\mathcal{P}) | \pi' = \pi\}$. It is easy to find that $P \ne \varnothing$. So there exists $p \in P$ such that $(\forall p' \in P) \Delta(p) \le \Delta(p')$. Pick any $p' \in \Lambda(\mathcal{P})$(suppose $p' = \langle \pi', \zeta' \rangle$). We have $\sum_{\varphi \in N} u_{\varphi}(p') = \sum_{\varphi \in N} u_{\varphi}(\pi') \le \sum_{\varphi \in N} u_{\varphi}(\pi) = \sum_{\varphi \in N} u_{\varphi}(p)$, i.e., there is no $p' \in \Lambda(\mathcal{P})$ such that $u_{\varphi}(p') > u_{\varphi}(p)$ for all $\varphi \in N$. Obviously, $(\forall p \in \Lambda(\mathcal{P})) \sum_{\varphi \in N} u_{\varphi}(p) \le \sum_{\varphi \in N} u^T_{\varphi}$. So there exists $p'' \in P$ such that for each $\varphi \in N$:

$$\begin{cases} u_{\varphi}(p'') = u_{\varphi}(p') \quad \text{if} \quad u_{\varphi}(p') \ge u_{\varphi}^{T} \\ u_{\varphi}^{T} \ge u_{\varphi}(p'') \ge u_{\varphi}(p') \quad \text{if} \quad u_{\varphi}(p') < u_{\varphi}^{T} \end{cases} \tag{7}$$

Then $\Delta(p') \ge \Delta(p'') \ge \Delta(p)$. Consequently, $\Delta(p) = \min_{p' \in \Lambda(\mathcal{P})} \Delta(p')$ and p is a solution to $\mathcal{P}$. That is to say, if there is no solution to $\mathcal{P}$, then $\Omega^{\perp}(\mathcal{P}) = \varnothing$. □

In the following discussion, we suppose that, **MAPM** goes out of the loop consisting of step 4, 5, and 6 at time T'', goes out of the loop consisting of step 4, 5, 6, and 7 at time T', and terminates at time T. ω_t, π^*_t, $\prod_t$, M'_t, and θ_t denote the values of ω, π^*, $\prod$, M', and θ at time $1 \le t \le T$.

Let us now consider the quality of the result. As shown by the following proposition, the plan given by **MAPM** maximizes the gross utility.

Proposition 2: If $p = \langle \pi, \zeta \rangle$ is the result of **MAPM**, then $\sum_{\varphi \in N} u_{\varphi}(\pi') \le \sum_{\varphi \in N} u_{\varphi}(\pi)$ for each $\pi' \in \Omega^{\perp}(\mathcal{P})$.

Proof. It is easy to find that, if $p = \langle \pi, \zeta \rangle$ is the result then $\pi = \pi^*_{T'}$ (see step 7 of **MAPM**). Now we show that $(\forall \pi' \in \Omega^{\perp}(\mathcal{P})) \sum_{\varphi \in N} u_{\varphi}(\pi') \le \sum_{\varphi \in N} u_{\varphi}(\pi^*_{T'})$.

According to the steps from 4 to 7 in **MAPM**, we have (for any $T'' \le t \le T'$ and $T'' \le t' < T'$):

1. $\sum_{\varphi \in N} u_{\varphi}(\pi^*_t) = \max_{\pi \in \omega t} \sum_{\varphi \in N} u_{\varphi}(\pi)$;
2. $\pi^*_t \in \omega_t$, $\omega_{t'} \subseteq \omega_{t'+1}$;
3. $\prod_{t-1} \supseteq \prod_t$, $\prod_{t-1} - \prod_t = \{\pi \in \prod_{t-1} | \sum_{\varphi \in N} \min(\mathbf{d}[\pi][\varphi], \mathbf{c}[\varphi]) \ge \sum_{\varphi \in N} \mathbf{d}[\pi^*_t][\varphi]\}$;
4. $\omega_{T''} \subseteq \prod_{T''-1} = \Omega^{\perp}(\mathcal{P})$, and $\prod_{T'} = \varnothing$.

Pick any $T'' \le t \le T'$. According to item 3, we have

$$(\forall \pi \in \Pi_{t-1} - \Pi_t) \sum_{\varphi \in N} u_{\varphi}(\pi) \le \sum_{\varphi \in N} u_{\varphi}(\pi^*_t) \tag{8}$$

According to items 1 and 2, we have

$$\sum_{\varphi \in N} u_{\varphi}(\pi^*_t) \le \sum_{\varphi \in N} u_{\varphi}(\pi^*_{T'}) \tag{9}$$

According to formula (8), (9), and item 4, we have $(\forall \pi \in \Omega^{\perp}(\mathcal{P})) \sum_{\varphi \in N} u_{\varphi}(\pi) \le \sum_{\varphi \in N} u_{\varphi}(\pi^*_{T'})$. □

Let us now give our final result which characterizes the quality of the result of **MAPM**. The following theorem shows that the resulting joint plan is a solution to $\mathcal{P}$.

Theorem 2: If $p \ne$ **failure** is the result of **MAPM**, then p is a solution to $\mathcal{P}$.

Proof. Observe **MAPM**. We can find that $\mathbf{d}[\pi^*_{T'}][\varphi] = u^T_{\varphi} - u_{\varphi}(\pi^*_T)$ at any $t \ge T'$ for all $\varphi \in N$. So (see step 8, 9, 10, 11, and 12) the result of **MAPM** should be $p = \langle \pi^*_{T'}, \zeta \rangle$, such that:

- $u_{\varphi}(p) = u^{\perp}_{\varphi}$ for all $\varphi \in N - M'_T$,
- $\theta_T = \sum_{\varphi \in N} (u^T_{\varphi} - u_{\varphi}(\pi^*_{T'})) - \sum_{\varphi \in N-M'T} \mathbf{c}[\varphi]$,
- $M'' \subseteq M'_T$ and $|M''| = \theta_T - |M'_T| \lfloor \theta_T / |M'_T| \rfloor$,
- $u_{\varphi}(p) = u^T_{\varphi} - \lfloor \theta_T / |M'_T| \rfloor - 1$ for all $\varphi \in M''$, and
- $u_{\varphi}(p) = u^T_{\varphi} - \lfloor \theta_T / |M'_T| \rfloor$ for all $\varphi \in M'_T - M''$.

Now we show that p is a solution to $\mathcal{P}$.

It is easy to find that $\pi^*_{T'} \in \Omega^{\perp}(\mathcal{P})$. According to the Stop subroutine (see Fig. 2), $u^T_{\varphi} - u^{\perp}_{\varphi} \ge \lceil \theta_T / |M'_T| \rceil$ for all $\varphi \in M'_T$. So $u_{\varphi}(p) \ge u^{\perp}_{\varphi}$ for all $\varphi \in N$. Consequently, $p \in \Lambda(\mathcal{P})$. Pick any $p' = \langle \pi', \zeta' \rangle \in \Lambda(\mathcal{P})$. According to Proposition 2, we have:

$$\sum_{\varphi \in N} u_{\varphi}(p') = \sum_{\varphi \in N} u_{\varphi}(\pi') \le \sum_{\varphi \in N} u_{\varphi}(\pi^*_{T'}) = \sum_{\varphi \in N} u_{\varphi}(p) \tag{10}$$

So there is no $p' \in \Lambda(\mathcal{P})$ such that $u_{\varphi}(p') > u_{\varphi}(p)$ for all $\varphi \in N$.

Let $P = \{\langle \pi^*_{T'}, \zeta'' \rangle \in \Lambda(\mathcal{P}) | (\forall \varphi \in N) \zeta''(\varphi) \le u^T_{\varphi} - u_{\varphi}(\pi^*_{T'})\}$ and $P' = \{p'' \in P | (\forall \varphi \in N - M'_T) u_{\varphi}(p'') = u^{\perp}_{\varphi}\}$. It is easy to find that:

1. there exists $p'' \in P$ such that $\Delta(p'') \le \Delta(p')$, and
2. $\Delta(p) = \min_{p'' \in P'} \Delta(p'')$.

According to steps 8, 10, 11, and 12, $(\forall \varphi \in M'_T)(\forall \varphi' \in N - M'_T) 0 \le u^T_{\varphi'} - u^{\perp}_{\varphi'} < \theta_T / |M'_T| \le u^T_{\varphi} - u^{\perp}_{\varphi}$ and $\theta_T + \sum_{\varphi \in N-M'T}(u^T_{\varphi}-$

$u^{\perp}_{\varphi}$)=$\sum_{\varphi\in N}$ (u^{T}_{φ}-u_{φ}(π^*_{T} ·)). So there exists p^∈P' such that Δ(p^)≤Δ(p"). According to items 1 and 2, Δ(p)≤Δ(p^)≤Δ(p")≤ Δ(p'), i.e., Δ(p)=$\min_{p'\in\Lambda(\mathcal{P})}$Δ(p'). □

V. Related Work, And Future Work

Unlike the previous work that consider planning for self-interested agents in fully adversarial settings (e.g., [10, 11]), [4, 5] propose the notion of planning games in which self-interested agents are ready to cooperate with each other to increase their personal net benefits. Fully centralized planning approaches have been proposed for subsets of planning games (see [4, 5] for examples). All these approaches require agents to fully disclose their private information. However, this requirement is not acceptable in this paper. Following the tradition of non-cooperative game theory, [7] presents a distributed multi-agent planning and plan improvement method, which is guaranteed to converge to stable solutions for congestion planning problems [12, 13]. But multi-agent planning problems discussed in this paper are different from congestion planning problems, in which each agent can plan individually to reach its goal from its initial state and no other agent can contribute to achieving its goal. On another hand, probably because of the loose interaction between agents in congestion planning problems, private information is not a focus of the work. [14] investigates algorithms for solving a restricted class of “safe” coalition-planning games, in which all the joint plans constructed by combining local solutions to the smaller planning games over disjoint subsets of agents are valid. It is easy to find that the multi-agent planning problems discussed in this paper are not “safe”.

Different from the mechanisms [15, 16, 17] of single agent's plan execution, a mechanism of multiple self-interested agents' joint plan execution must deal with values' realization. As future work, we plan to investigate agents' power [8] and values' realization in joint plans' execution. Planning domains can be nondeterministic. So we will redefine the concept of solutions and multi-agent planning algorithms in strong [18] or probabilistic style. On another hand, we will design a more general bargaining mechanism for generating joint plans, which can deal with changing goals [19], incomplete information [20], and concurrent actions.

Acknowledgment

The work is sponsored by the Fundamental Research Funds for the Central Universities of China (Issue Number WK0110000026), and the National Natural Science Foundation of China under grant No.61105039.

References

[1] Ronen I. Brafman, Carmel Domshlak: From one to many: Planning for loosely coupled multi-agent systems. In: ICAPS (2008)

[2] Raz Nissim, Ronen I. Brafman, Carmel Domshlak: A general, fully distributed multi--agent planning algorithm. In: AAMAS, pp. 1323-1330. (2010)

[3] Alexandros Belesiotis, Michael Rovatsos, Iyad Rahwan: Agreeing on plans through iterated disputes. In: AAMAS, pp. 765-772. (2010)

[4] Ronen I. Brafman, Carmel Domshlak, Yagil Engel, Moshe Tennenholtz: Planning games. In: IJCAI, pp. 73-78. (2009)

[5] Ronen I. Brafman, Carmel Domshlak, Yagil Engel, Moshe Tennenholtz: Transferable utility planning games. In: AAAI, pp. 709-714. (2010)

[6] Ning Sun, Zaifu Yang: A double-track adjustment process for discrete markets with substitutes and complements. Econometrica. 77(3), 933-952 (2009)

[7] Anders Jonsson, Michael Rovatsos: Scaling up multiagent planning: A best-response approach. In: ICAPS, pp. 114-121. (2011)

[8] Sviatoslav Brainov, Tuomas Sandholm: Power, dependence and stability in multiagent Plans. In: AAAI, pp. 11-16. (1999)

[9] John Nash: The bargaining problem. Econometrica. 18(2), 155-162 (1950)

[10] M. H. Bowling, R. M. Jensen, M. M. Veloso: A formalization of equilibria for multiagent planning. In: IJCAI, pp. 1460-1462. (2003)

[11] R. Ben Larbi, S. Konieczny, P. Marquis: Extending classical planning to the multi-agent case: A Game-theoretic Approach. In: ECSQARU, pp. 731-742. (2007)

[12] Rosenthal, R. W: A class of games possessing pure-strategy Nash equilibria. International Journal of Game Theory. 2, 65-67 (1973)

[13] Monderer, D., Shapley, L. S.: Potential games. Games and Economic Behaviour. 14(1), 124-143 (1996)

[14] Matt Crosby, Michael Rovatsos: Heuristic multiagent planning with self-interested Agents. In: AAMAS, pp. 1213-1214. (2011)

[15] Piergiorgio Bertoli, Alessandro Cimatti, Marco Pistore, Paolo Traverso: A framework for planning with extended goals under partial observability. In: ICAPS, pp. 215--225. (2003)

[16] Sebastian Sardina, Giuseppe De Giacomo, Yves Lesperance, Hector J. Levesque: On the limits of planning over belief states under strict uncertainty. In: KR, pp. 463-471. (2006)

[17] Wei Huang, Zhonghua Wen, Yunfei Jiang, Hong Peng: Structured plans and observation reduction for plans with contexts. In: IJCAI, pp. 1721-1727. (2009)

[18] Alessandro Cimatti, Marco Pistore, Marco Roveri, Paolo Traverso: Weak, strong, and strong cyclic planning via symbolic model checking. Artificial Intelligence. 147, 35-84 (2003)

[19] Tran Cao Son, Chiaki Sakama: Negotiation using logic programming with consistency restoring rules. In: IJCAI, pp. 930-935. (2009)

[20] Piergiorgio Bertoli, Alessandro Cimatti, Marco Roveri, Paolo Traverso: Strong planning under partial observability. Artificial Intelligence. 170, 337-384 (2006)

Optimisation of Resource Scheduling in VCIM Systems Using Genetic Algorithm

Son Duy Dao
School of Advanced Manufacturing and Mechanical Engineering,
University of South Australia,
Australia

Kazem Abhary
School of Advanced Manufacturing and Mechanical Engineering,
University of South Australia,
Australia

Romeo Marian
School of Advanced Manufacturing and Mechanical Engineering,
University of South Australia,
Australia

***Abstract*— The concept of Virtual Computer-Integrated Manufacturing (VCIM) has been proposed for one and a half decade with purpose of overcoming the limitation of traditional Computer-Integrated Manufacturing (CIM) as it only works within an enterprise. VCIM system is a promising solution for enterprises to survive in the globally competitive market because it can exploit effectively locally as well as globally distributed resources. In this paper, a Genetic Algorithm (GA) based approach for optimising resource scheduling in the VCIM system is proposed. Firstly, based on the latest concept of VCIM system, a class of resource scheduling problems in the system is modelled by using agent-based approach. Secondly, GA with new strategies of handling constraint, chromosome encoding, crossover and mutation is developed to search for optimal solution for the problem. Finally, a case study is given to demonstrate the robustness of the proposed approach.**

Keywords- optimisation; genetic algorithm; resource scheduling; virtual computer-integrated manufacturing system.

I. Introduction

Computer-Integrated Manufacturing (CIM) is a manufacturing strategy of using computers and communication networks to fully integrate all of production processes into a highly interconnected manufacturing system. A network of interconnected CIM systems which are globally distributed is defined as Virtual Computer-Integrated Manufacturing (VCIM) system [1,2]. The concept of VCIM system has been proposed for one and a half decade with purpose of overcoming the limitation of traditional CIM system as it only works within an enterprise. VCIM system is a promising solution for enterprises to survive in the globally competitive market because it can exploit effectively globally as well as locally distributed resources.

The VCIM system is still being developed and the latest concept of VCIM system can be explained as follows. Main purpose of VCIM is to effectively share distributed resources and management objectives.

Although, different enterprises which might be located at different areas have their own manufacturing resources, they intend to temporarily work together in an integrated manner to fulfill specific product orders. When the orders are fulfilled, this temporary system disappears. The virtual status of those temporary manufacturing systems is a special characteristic of VCIM system, compared with traditional manufacturing system. That is why the word "virtual" is used in the VCIM concept [2].

One of the key issues that have to be resolved to make the VCIM happen is how to organise manufacturing resources to produce a specific product. When receiving an order from customer, VCIM system generates automatically all possible manufacturing resource schedules in which the resources could be locally or globally distributed and find the best schedule in real time to form a temporary manufacturing system to fulfill the customer's order. Different orders or even the same orders but at different time might require different temporary manufacturing systems. This is a class of highly constrained optimisation problems which requires a robust approach to deal with the dynamic environment in the system and then find an optimal solution, especially for large-scale practical problems.

Literature review shows that although resource scheduling is very important to VCIM system, the number of research dealing with the problem is very limited. There has been only one approach proposed and developed by the same group of researchers [2-7]. Furthermore, this approach is not capable to find a global optimum.

In this paper, Genetic Algorithm (GA) based approach for optimising resource scheduling in VCIM system is proposed. First, the latest concept of resource scheduling in VCIM system proposed by [2] is modelled as a multi-agent system and then GA is developed to optimise the model. The proposed GA is capable to deal with dynamic environment of the system by automatically changing the size of chromosome according to specific product orders of customers. It is also capable to search for optimal solution for the resource scheduling very quickly. To do so, some strategies of crossover, mutation and handling constraints to make the chromosomes in the GA feasible are developed and presented herein.

The rest of this paper is organised as follows. Section 2 discusses the literature review on the latest VCIM concept and approach used to optimise resource scheduling in the system. Research methodology based on GA, which is proposed to optimise the resource scheduling, is introduced in Section 3. Section 4 is about case study used to demonstrate the robustness of the GA based approach. Some conclusions and future work are given in Section 5.

II. Literature Review

The concept of VCIM system has been proposed by Lin [8] since 1997. This concept is relatively new and still being developed. As the nature of a distributed system, the agent-based approach is the most appropriate approach to model the VCIM system. The latest agent-based VCIM architecture can be described as follows. There are three types of agents: facilitator agent, customer agents, and resource agents. Facilitator agents work as coordinators to manage the information flow across the VCIM system. Customer agents are to provide interfaces for customers to participate in the system. Resource agents are manufacturing functionalities in the VCIM system such as assembly workshop, material provider, delivery facility, design department, etc. All of the agents are connected to each other via Internet. The system works as follows. When receiving a product-request message, customer agent passes it to facilitator agent. The request is then decomposed into interrelated subtasks by the facilitator agent, according to the registered functionalities of resource agents and the built-in knowledge base of the facilitator agent. Functionalities of resource agents could be product design, material supply, manufacturing, assembly, test, package and warehousing. Next, the facilitator agent selects resource agents to perform the subtasks based on multiple criteria such as suitable functionality, time frame, availability, cost, quality, friendship, credit, transportation service and delivery reliability. The result of this selection is resource schedule from which facilitator agent forms a temporary manufacturing system to fulfill the customer's order [2-5,9-11]. It is clear that different customer order or even the same order requested at different time in VCIM system requires different temporary manufacturing system.

It can be clearly seen that resource scheduling problem in the VCIM system associates with hard precedence constraint, supply chain management, resource allocation and task sequencing. Precedence constraints are generally divided into two categories: hard precedence constraints and soft precedence constraints. A hard precedence constraint is a constraint that makes the solution illegal or infeasible if violated, while a soft precedence constraint is a constraint that imposes a penalty if violated rather than rendering the sequence and schedule infeasible. Assembly sequence and selection of resource agent to do a given subtask are examples of hard precedence constraints. In addition, one of the special characteristic of this problem is that the size of its solution is variable which depends on the nature of the requested product. This is because different products could be decomposed into different number of subtasks and sometimes one resource agent can do several different subtasks. Therefore, the number of resource agents required to do those subtasks could be different. To optimise production scheduling problem associated with soft precedence constraint and variability of solution size, GAs with variable length chromosomes have been proposed by some researchers [12-15]. However, supply chain management and hard precedence constraints have not been integrated into those schedules yet.

Main advantage of VCIM system is the effective sharing globally and locally distributed resources. Therefore, resource scheduling plays an important role in the system. However, the research on this problem is still very limited. Literature review shows that there has been only one approach proposed to optimise the resource scheduling in VCIM system. To optimise the resource scheduling for a given product request, some researches [2-7] proposed and developed the Backward Network Algorithm for Shortest-path. In general, this approach includes the following steps in the resource allocation process which is done after the customer order is decomposed into interrelated subtasks:

Step 1: Facilitator agent selects a subtask by using a backward sequence (the final subtask is selected first and vice versa).

Step 2: Facilitator agent selects resource agents to perform the selected subtask in Step 1 by matching the subtask and resource agent's functionality.

Step 3: Facilitator agent requests all transportation proposals to transport the parts from the selected resource agent in Step 2 to the previously selected resource agent due to the backward sequence. It is noted that for the final subtask, transportation proposals are to transport the requested product to the customer.

Step 4: Facilitator agent short-lists transportation proposals, achieved in Step 3. In this selection process, a transportation proposal having the lowest cost is selected. If there are more than one transportation proposals having the lowest cost, the one with the latest start time is selected.

Step 5: Facilitator agent combines the selected subtask in Step 1, the selected resource agent in Step 2 and the selected transportation proposal in Step 4 to form an intermediate schedule. There could be many intermediate schedules, if there are many in the short-listed transportation proposals in Step 4.

Step 6: Facilitator agent short-lists intermediate schedules that are achieved in Step 5. Intermediate schedules having a lower cost or a higher cost but with a later start time are selected. The first 6 Steps result in the intermediate schedules to complete the selected subtask in Step 1 by the selected resource agent in Step 2.

Step 7: Facilitator agent repeats Steps 2 to 6 until all resource agents have provided proposals to perform the subtask selected in Step 1. After the first 7 Steps, a set of intermediate schedules to complete one selected subtask in Step 1 is achieved.

Step 8: Facilitator agent repeats Steps 1 to 7 until all subtasks have been completed. When the schedules for a new subtask are achieved, new intermediate schedules are formed by including the new subtask's schedule information such as: the selected subtask, resource agent and transportation proposals. As a result, full production schedules are achieved when all subtasks are completed.

Step 9: Facilitator agent selects the best full resource schedule based on total cost and time required.

It can be clearly seen that the existing approach cannot be able to find a globally optimised solution for the resource scheduling problem in the VCIM system because of the way that the full resource schedules are formed. Given that the intermediate schedules to complete every single subtask achieved in Step 7 are optimised, there is no way to guarantee that the full resource schedules formed by adding above optimal intermediate schedules without any negotiation between them as presented in Step 8 are globally optimised.

To overcome this limitation, this paper proposes GA based approach for optimising resource scheduling in the VCIM system.

III. RESEARCH METHODOLOGY

A. Problem Modelling

Based on the ideas from the previous works [2-5,7,9-11], a class of resource scheduling problem in the VCIM system considered in this research, at this stage, is modelled as follows.

Consider:

- There is a VCIM system capable to produce N different products.
- The products can be made by assembling number of different parts or groups of parts.
- The VCIM system has M different manufacturing agents with certain functionalities, which are locally or globally distributed.
- Each manufacturing agent is capable to produce a limited number of different parts of a product and it is also able to assemble particular groups of parts.
- The system also has K final assembly agents in which the product's parts or groups of parts can be assembled together to form final products and then they are tested and packed before being transported to the customers. It is noted that locations and operation costs of those final assembly agents are known.
- Costs of each part or a group of parts produced in different manufacturing agents are different but known in advance.
- Distances and transportation costs between the manufacturing agents and the final assembly agents or the final assembly agents and the customers are all known.

Determine:

Which manufacturing agents and final assembly agent to be selected to form a temporary production system to fulfill the customer's orders.

So that:

Cost of the product is minimised while all of the given constraints are simultaneously satisfied.

Conditions:

- All of the agents have no limitation in capability of performing their designed functions.
- Time constraints in this scheduling problem are ignored.
- All of the products can be decomposed into a number of parts or groups of parts which are all known in advance.
- Final assembly agents are all able to form any final product in the VCIM system but with different costs.

B. Optimisation Methodology

To find an optimal solution to the problem described above, a GA with new features in chromosome encoding, mutation and crossover operations is proposed herein. There are five main components of the GA as follows:

- *Chromosome Encoding*

To encode solutions for the problem, chromosomes in this GA have two parts: main part and additional part to represent part-manufacturing agent allocation and final assembly agent selection, respectively.

Main part of a chromosome is defined as a matrix with the size of C x R where C is the number of decomposed parts or groups of parts needed to assemble a requested product and R is the number of manufacturing agents in the VCIM system. In Table 1, there is one example of main part of chromosome where C and R are 20 and 15, respectively.

Clearly, for a given VCIM system, R is constant and C is variable. That is because C depends on nature of different products requested. All elements in the C x R matrix are binary, 0 or 1. The number "1" in the matrix means that the manufacturing agent in its row is selected to complete the part or group of parts in its column. The number "0" or blank cell in that matrix means that the corresponding manufacturing agent is not selected to complete the corresponding part or group of parts. This part of chromosome represents the allocation of manufacturing agents to complete the parts or group of parts.

TABLE 1. A MAIN PART OF A CHROMOSOME

Manufacturing agent	Parts or groups of parts needed to assemble a product																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1						1														
2	1																			
3												1			1				1	1
4		1									1									
5																		1		
6																				
7			1											1						
8										1										
9							1													
10				1													1			
11													1							
12					1															
13																				
14								1								1				
15									1											

The other part of the chromosome is a binary string with the length of K representing the selection of final assembly agent to form final product as customer's requirements. As all final assembly agents are assumed to be capable to form any final products in the system, the best one in regard to the main part should be selected to form additional part of chromosome which looks like Table 2. As a result, the additional part is related to the main part and there is no need to select at random a final assembly agent to form the additional part. It means that the additional part of chromosome is generated according to its main part.

TABLE 2. AN ADDITIONAL PART OF A CHROMOSOME

	Final assembly agents in the VCIM system																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Agent selected	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0

Because of the assumption that all final assembly agents are capable to form any final product in the system, the feasibility of the chromosome depends only on its main part. To generate a main part of feasible chromosomes, the following steps are proposed.

Step 1: Check the size of the main part of chromosome based on information from database about the requested product.

Step 2: Generate a zero matrix with the same size as obtained in Step 1.

Step 3: Select at random one part or group of parts of the requested product which corresponds to one column in the zero matrix, denoted as x.

Step 4: Check database to find out which manufacturing agents are able to complete the selected part or group of parts in Step 3 based on their functionalities.

Step 5: Randomly select one manufacturing agent among those in Step 4, denoted as y.

Step 6: Insert number “1” in a cell located at column x and row y in the zero matrix.

Step 7: Repeat Step 3-6 until all parts or group of parts of a given product are selected.

As a result, the main part of feasible chromosome is as shown in Table 1.

- *Genetic Operators*

Due to nature of the chromosome, a modified crossover and mutation operations are required. This paper proposes that crossover and mutation operations are only applied to main part of chromosome. That is because the additional part is automatically determined after the main part is generated, based on overall cost of schedule encoded in the corresponding chromosome. In addition, after crossover and mutation operations applied as shown in Tables 3 and 4, offspring chromosomes are checked and repaired to make them feasible. To check the feasibility of the offspring, gene-by-gene feasible checking approach is proposed. It is noted that a gene herein is defined as a column in the main part of chromosome, the highlighted columns in Table 4, for examples. By matching the manufacturing agent’s functionality to the part or group of parts in the gene, the feasibility of a gene can be determined. The principle in the proposed gene-by-gene feasible checking approach is that if a gene is feasible, it is accepted, otherwise, a new gene is generated by following steps:

Step 1: Delete all information in the gene. This gene is now a zero matrix with size of 1 x R.

Step 2: Check the database to find out which manufacturing agents are able to complete the part or group of parts in the selected gene.

Step 3: Randomly select one manufacturing agent among those in Step 2, denoted as z.

Step 4: Insert number “1” in a cell located at column 1 and row z in the zero matrix in Step 1 to form a new feasible gene.

- *Evaluation*

The chromosomes in the GA are evaluated through one objective function which is the total cost of the requested product. Total cost for a product is calculated as follows.

$$TC = CP + CA + CT1 + CT2$$

Where: TC is total cost of the product; CP is total cost of parts or groups of parts made in the selected manufacturing agents; CA is final assembly cost in the selected final assembly agent; CT1 is total cost of transportations between the selected manufacturing agents and final assembly agent; CT2 is cost of transportation between the selected final assembly agent and the customer’s product delivery destination.

TABLE 3. CROSSOVER OPERATION

Manufacturing agent	Parts or groups of parts needed to assemble a product																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1						1														
2	1																			
3												1			1				1	1
4		1									1									
5																		1		
6																				
7			1											1						
8										1										
9							1													
10				1													1			
11													1							
12					1															
13																				
14								1								1				
15									1											

Manufacturing agent	Parts or groups of parts needed to assemble a product																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1		1				1								1						
2																				
3	1								1				1							
4																				
5											1									
6					1														1	
7																	1			
8										1										
9			1																	
10				1														1		
11																				
12							1													
13																				1
14								1							1	1				
15												1								

TABLE 4. MUTATION OPERATION

Manufacturing agent	Parts or groups of parts needed to assemble a product																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1		1				1								1						
2																	1			
3	1								1				1							
4																				
5											1									
6					1														1	
7																				
8										1										
9			1																	
10				1														1		
11																				
12							1													
13																				1
14								1							1	1				
15												1								

Manufacturing agent	Parts or groups of parts needed to assemble a product																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1		1				1								1						
2				1																
3	1								1				1							
4																				
5											1									
6					1														1	
7																				
8										1										
9			1																	
10																	1	1		
11																				
12							1													
13																				1
14								1							1	1				
15												1								

- *Selection*

In this paper, *roulette wheel* approach [16] is used to select the population for the next generations. This approach belongs to the fitness-proportional selection and selects a new population based on the probability distribution associated with fitness value, calculated through the objective function above.

The proposed GA has classical structure and it is implemented in Matlab to search the optimal solution for the problem. A case study is given to illustrate the proposed approach.

IV. A Case Study

A. Modelling of the Scheduling Problem

To illustrate the proposed approach and demonstrate its capability, a case study is given and described as follows.

There is a VCIM system with 15 manufacturing agents, 5 final assembly agents which is capable to produce 10 different products. Distances between the manufacturing agents and the final assembly agents are given in Table 5. There are 14 different product delivery destinations required by customers. Distances between the final assembly agents and the product delivery destinations are given in Table 6. The rates of different transportations are given in Table 7. In addition, the products can be made by assembling a number of different parts as shown in Table 8. Moreover, final assembly costs for different products are detailed in Table 9. Capabilities of the manufacturing agents in completing parts or groups of parts for different products are detailed in Table 10. It is noted that because of lengthiness, a part of Table 10 is given herein. Lastly, costs of parts or groups of part for different products made in different manufacturing agents are given in Table 11. Again, due to lengthiness, a part of Table 11 is given herein.

Problem here is that which manufacturing agents and final assembly agent should be selected to form a temporary production system to fulfill the customer's order so that total cost of the requested product is minimised.

Conditions:

- All of the agents have no limitation in capability of performing their designed functions.
- Time constraints in this scheduling problem are ignored.
- Final assembly agents are all able to form any final product in the VCIM system.

B. Results and Discussion

The proposed GA has been successfully implemented in Matlab and has been comprehensively tested for convergence and consistence of the solutions. It is evident that GA cannot guarantee to find best solution after only one run. However, it is very good at finding good/best solution(s) reasonably quickly. Therefore, it is easy to evaluate the quality of solutions by comparing different runs. Accordingly, the best solution among the ones obtained from different runs should be selected. If the number of runs is large enough, this validates the global optimality of the solution [17, 18].

Without losing generality, it is assumed that product 4 is requested by a customer and the required delivery destination is 5. In this case study, the GA was run for 200 times with different parameters and experiment results are shown in Tables 12 and 13. The probability of obtaining the minimum value of the objective function, the cost of the requested product, is visualised and estimated as the three-parameter lognormal distribution by Minitab software as shown in Fig. 1.

TABLE 5. DISTANCES BETWEEN AGENTS (In Km)

		Manufacturing agent														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Final assembly agent	1	1960	1191	236	172	1462	1927	1249	76	525	215	1808	62	1221	367	337
	2	879	525	594	526	978	1095	1359	1771	672	1309	1783	1489	1236	481	1958
	3	223	1207	639	1603	1158	1043	792	1828	1360	989	669	1001	1720	1774	1426
	4	517	1423	849	59	476	464	736	1593	274	1559	1398	961	1612	58	1002
	5	818	444	1017	1859	919	979	1977	198	1443	1431	397	1810	1154	981	943

TABLE 6. DISTANCES BETWEEN AGENTS & CUSTOMERS (In Km)

		Product delivery destinations required by customers													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
Final assembly agent	1	7	11	67	46	18	41	30	12	96	43	31	68	4	38
	2	69	83	53	44	40	54	44	38	93	56	71	19	57	47
	3	5	83	98	84	84	43	3	21	6	95	68	14	89	99
	4	8	73	66	9	81	67	99	50	75	43	55	101	68	17
	5	53	16	81	14	7	64	18	35	28	99	71	18	20	87

TABLE 7. TRANSPORTATION RATE (In $)

		Manufacturing agent														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Final assembly agent	1	0.8	0.2	0.2	0.2	0.7	0.8	0.7	0.8	0.5	0.5	0.3	0.5	0.8	1.0	0.9
	2	0.9	0.4	1.0	0.5	0.1	0.8	0.1	0.7	0.4	0.5	0.7	1.0	0.3	0.6	0.3
	3	0.2	0.6	1.0	0.9	0.9	0.5	0.3	0.4	0.8	0.7	0.7	0.4	0.6	0.2	0.8
	4	0.9	1.0	0.5	0.8	0.9	0.7	0.1	1.0	0.8	0.7	0.2	0.6	0.7	0.2	0.3
	5	0.7	1.0	0.8	1.0	0.7	0.3	0.2	0.1	0.3	0.8	0.2	0.3	0.9	0.3	0.9

		Product delivery destinations required by customers													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
Final assembly agent	1	0.9	0.5	0.9	0.5	0.3	0.6	0.9	0.5	0.5	1.0	0.2	0.7	0.7	0.3
	2	0.3	0.4	1.0	0.6	0.2	0.2	0.8	1.0	0.7	0.1	0.3	0.2	0.9	0.1
	3	0.3	0.9	0.5	0.3	0.4	0.3	0.5	0.6	0.7	0.9	0.4	0.7	0.9	0.8
	4	0.3	0.5	0.2	0.6	0.4	0.8	0.6	0.6	0.5	0.9	0.7	0.5	0.4	0.6
	5	0.3	0.3	0.3	0.7	0.5	0.1	0.3	0.3	0.4	0.8	0.2	0.8	0.7	0.5

TABLE 8. THE DECOMPOSED PARTS OR GROUP OF PARTS

	Product									
	1	2	3	4	5	6	7	8	9	10
Number of parts or group of parts	30	24	27	36	31	23	33	26	32	28

TABLE 9. FINAL ASSEMBLY COST FOR PRODUCTS (In $)

		Product									
		1	2	3	4	5	6	7	8	9	10
Final assembly agent	1	355	474	297	380	274	500	319	401	370	490
	2	609	383	444	236	531	458	469	317	149	646
	3	219	387	316	125	635	466	342	319	238	613
	4	101	157	104	563	423	616	173	513	346	151
	5	599	648	400	127	528	394	342	387	451	624

The result shown in Tables 12 and 13 indicates that the GA with population size of 100, crossover rate of 50%, mutation rate of 60% gives the better solutions – the lowest average cost of $4887.8. The lowest cost of the product achieved, corresponding to the globally optimal solution, is $4803.1. The convergence of the GA for that solution is shown in Fig. 2 and detail of the solution is shown in Tables 14 and 15.

The experiments show that the proposed GA is consistently convergent towards an optimum within a very small amount of computational time, just less than 2 minutes. It is evident that the proposed approach can easily accommodate much larger and more complex resource scheduling problems in the VCIM system.

To the best of our knowledge, there has not been any quantitative research on resource scheduling in the VCIM system yet. Therefore, it is impossible to compare the quality of the solution obtained here in this case study to the others.

TABLE 10. FUNCTIONALITIES OF MANUFACTURING AGENTS

Product	Part or group of parts	Manufacturing agents														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	1	1	1	0	0	0	0	0	0	0	1	1	1	0	1
	2	0	0	0	1	0	1	1	0	0	0	1	0	0	1	0
	3	1	0	0	1	1	0	1	0	1	1	0	1	1	1	0
	4	1	1	0	1	1	0	0	1	0	0	0	1	1	0	0
	5	0	1	0	0	1	0	0	1	0	0	1	1	1	1	1
	6	0	0	1	1	0	0	0	0	1	1	0	1	0	0	0
	7	0	0	1	0	0	1	1	1	1	0	0	0	1	0	1
	8	0	1	1	0	1	1	1	0	1	1	1	1	0	1	1
	9	0	1	0	1	0	1	0	0	0	0	0	1	1	1	1
	10	1	0	1	1	0	1	0	1	0	1	0	1	0	0	0
	11	1	0	1	1	0	0	1	1	1	0	1	1	0	1	0
	12	0	0	1	0	0	0	1	0	0	0	1	1	0	0	1
	13	1	1	1	1	1	1	1	1	0	1	0	1	1	0	0
	14	0	1	0	1	0	1	1	1	1	1	0	0	1	0	1
	15	0	1	0	0	1	1	1	0	1	0	1	1	1	0	1
	16	0	0	1	0	0	0	1	1	0	0	0	0	1	1	0
	17	0	1	1	1	1	1	1	0	1	0	1	1	0	0	1
	18	1	0	1	0	1	1	1	1	0	0	1	0	0	0	0
	19	0	0	1	1	0	0	0	0	1	0	1	0	1	0	0
	20	1	1	1	0	1	1	0	0	1	1	1	0	0	0	1
	21	0	0	0	0	1	0	0	1	0	0	0	0	1	1	1
	22	1	0	0	1	1	0	1	1	1	0	0	1	1	1	1
	23	1	1	0	1	1	1	0	0	1	0	0	0	1	0	1
	24	1	0	0	0	1	1	1	1	0	0	0	1	1	1	1
	25	1	1	0	1	0	1	0	0	1	1	1	0	0	0	0
	26	0	1	1	0	1	0	1	0	0	0	1	0	1	0	0
	27	0	0	1	1	1	0	0	0	0	1	1	0	0	0	1
	28	0	0	0	1	1	0	0	0	0	0	0	0	1	1	1
	29	0	1	1	1	0	0	1	0	1	1	1	1	1	1	0
	30	0	1	1	0	1	1	0	1	1	0	1	1	1	1	0
2	1	1	1	0	1	1	1	0	1	0	1	1	1	1	1	0
	2	1	1	0	0	1	1	1	1	0	1	1	1	1	0	0
	3	0	0	1	0	0	1	1	1	1	0	0	0	1	0	1

TABLE 11. COSTS OF PARTS MADE IN AGENTS (In $)

Product	Part or group of parts	Manufacturing agents														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	243	227	298	-	-	-	-	-	-	-	66	134	72	-	183
	2	-	-	-	111	-	270	266	-	-	-	125	-	-	224	-
	3	246	-	-	161	92	-	171	-	212	276	-	235	142	240	-
	4	168	202	-	222	265	-	-	154	-	-	-	76	221	-	-
	5	-	163	-	-	297	-	-	101	-	-	240	82	199	214	84
	6	-	-	151	234	-	-	-	-	190	278	-	187	-	-	-
	7	-	-	162	-	-	297	207	71	264	-	-	-	142	-	86
	8	-	243	141	-	197	182	58	-	137	236	70	273	-	97	92
	9	-	138	-	226	-	170	-	-	-	-	-	250	72	117	99
	10	135	-	207	161	-	250	-	92	-	190	-	234	-	-	-
	11	202	-	243	55	-	-	62	205	94	-	183	63	-	172	-
	12	-	-	283	-	-	-	172	-	-	-	77	68	-	-	104
	13	235	258	293	156	256	275	98	63	-	125	-	72	188	-	-
	14	-	114	-	118	-	194	81	283	275	84	-	-	107	-	273
	15	-	203	-	-	130	261	101	-	80	-	123	286	210	-	226
	16	-	-	224	-	-	-	87	234	-	-	-	-	171	218	-
	17	-	185	73	157	72	196	97	-	185	-	53	83	-	-	96
	18	97	-	181	-	78	112	61	265	-	-	62	-	-	-	-
	19	-	-	183	148	-	-	-	-	300	-	217	-	75	-	-
	20	73	130	265	-	220	71	-	-	122	160	201	-	-	-	278
	21	-	-	-	-	174	-	-	265	-	-	-	-	109	129	227
	22	221	-	-	252	97	-	224	246	166	-	-	132	183	243	189
	23	187	211	-	239	174	232	-	-	241	-	-	-	73	-	128
	24	156	-	-	-	87	273	184	94	-	-	-	237	151	81	92
	25	211	210	-	104	-	296	-	-	75	127	122	-	-	-	-
	26	-	186	137	-	263	-	81	-	-	-	223	-	78	-	-
	27	-	-	87	287	190	-	-	-	-	75	189	-	-	-	93
	28	-	-	-	132	282	-	-	-	-	-	-	-	123	156	114
	29	-	230	116	218	-	-	268	-	180	133	65	293	201	214	-
	30	-	181	61	-	196	54	-	124	134	-	245	267	291	231	-
2	1	221	157	-	194	139	61	-	242	-	96	274	153	233	294	-
	2	151	81	-	-	299	231	297	134	-	75	84	173	84	-	-
	3	-	-	88	-	-	137	67	216	232	-	-	-	259	-	230

TABLE 12. THE COST ACHIEVED BY THE GA WITH DIFFERENT MUTATION RATE (POPULATION: 100, CROSSOVER: 50%)

		Mutation Rate (%)									
		10	20	30	40	50	60	70	80	90	100
Run	1	5528.0	5256.8	4931.3	5827.5	5042.8	5042.8	5256.8	4931.3	5074.1	5796.0
	2	5962.2	4803.1	4803.1	4803.1	5256.8	4803.1	4803.1	5171.0	4803.1	4803.1
	3	5385.0	6158.4	4803.1	5042.8	5042.8	4803.1	4807.1	5080.8	6273.3	5256.8
	4	6222.7	5256.8	5256.8	5171.0	5246.0	5042.8	5329.5	5967.2	5705.8	4803.1
	5	6656.3	4931.3	4931.3	5689.7	4803.1	5042.8	4803.1	4803.1	5689.7	6031.9
	6	5528.0	4931.3	5042.8	5042.8	4803.1	4803.1	4803.1	5062.8	4803.1	5171.0
	7	5689.7	6576.1	5042.8	4803.1	4931.3	4803.1	4803.1	4803.1	5256.8	5840.5
	8	5042.8	5528.0	4803.1	5042.8	4803.1	4931.3	5042.8	4803.1	4803.1	4803.1
	9	5903.7	5256.8	4803.1	5256.8	4803.1	4803.1	5102.6	5817.9	4931.3	4803.1
	10	4803.1	4803.1	4803.1	5528.0	5042.8	4803.1	4803.1	4931.3	5042.8	4803.1
Average		$5,672.2	$5,350.2	$4,922.1	$5,220.8	$4,977.5	$4,887.8	$4,955.4	$5,137.2	$5,238.3	$5,211.2

TABLE 13. THE COST ACHIEVED BY THE GA WITH DIFFERENT CROSSOVER RATE (POPULATION: 100, MUTATION: 50%)

		Crossover Rate (%)									
		10	20	30	40	50	60	70	80	90	100
Run	1	4931.3	6158.4	5042.8	5042.8	4803.1	5042.8	6160.5	5042.8	5042.8	6376.9
	2	4803.1	5042.8	4803.1	4803.1	5042.8	4931.3	4803.1	4803.1	4803.1	6232.9
	3	5903.7	5164.8	4931.3	5029.8	4931.3	4803.1	5905.2	4931.3	5256.8	5171.0
	4	4931.3	5062.8	5042.8	6158.4	4803.1	6286.6	5042.8	5042.8	5102.6	4931.3
	5	5256.8	6286.6	5256.8	5042.8	5903.7	5385.0	4931.3	5042.8	4803.1	5683.7
	6	5029.8	4803.1	4803.1	4803.1	5256.8	4803.1	5377.8	5627.5	5042.8	4803.1
	7	6158.4	5042.8	5256.8	5171.0	5074.1	5042.8	4931.3	5102.6	4803.1	5042.8
	8	4803.1	5042.8	5796.0	5042.8	5256.8	5042.8	4803.1	5066.8	4803.1	5385.0
	9	6465.4	5171.0	4803.1	4803.1	5171.0	4803.1	4803.1	5609.5	4803.1	4803.1
	10	4803.1	4803.1	4803.1	5256.8	6031.9	4803.1	4803.1	4803.1	4931.3	4931.3
Average		$5,308.6	$5,257.8	$5,053.9	$5,115.4	$5,227.5	$5,094.4	$5,156.1	$5,107.2	$4,939.2	$5,336.1

V. CONCLUSIONS AND FUTURE WORK

In this paper, a class of resource scheduling problems in the VCIM system has been modelled. Due to the nature of constraints, the GA with new strategies for chromosome encoding, crossover and mutation operations have been developed to optimise the model.

Robustness of the proposed GA has been verified in the case study by extensive tests with various input parameters. The evolution of the output is consistently convergent towards the optimum within a very small amount of processing time. It is evident that the proposed approach can easily accommodate much larger and more complex resource scheduling problems in the VCIM system.

Future work will focus on:

- Generalisation of the VCIM model by adding more realistic and complex constraints.
- Incorporation of stochastic events into the model.
- Improvement of the proposed GA for the extended resource scheduling problems.
- Verification of the solutions obtained from the GA by simulation software such as Automod.

TABLE 14. FIRST PART OF THE OPTIMAL SOLUTION

	Final assembly agents in the VCIM system																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Agent selected	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

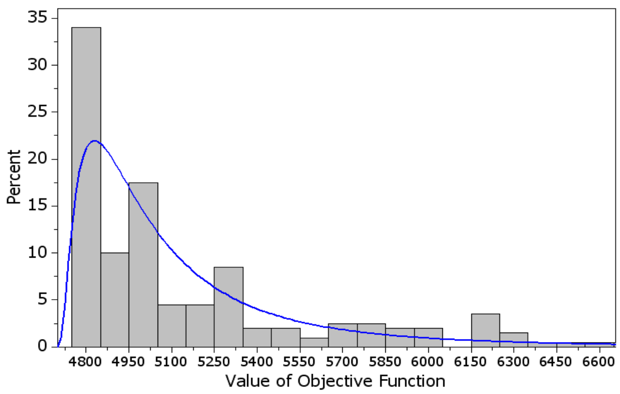


Figure 1. Probability of obtaining the globally optimal solution

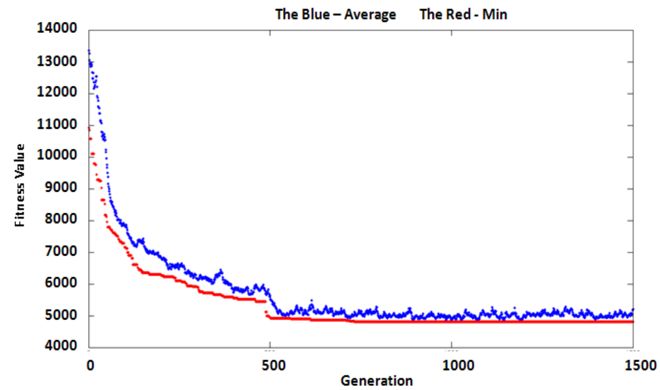


Figure 2. Evolution of fitness value – output of the GA

TABLE 15. SECOND PART OF THE OPTIMAL SOLUTION

Manufacturing agent	Parts or groups of parts needed to assemble a product																																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
4	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	1	0	1	1	0	0	1	1	0	1	0	0	0	1	0	0	1	0	0	0	0	0	1	0	1	0	0	0	1	0	0
9	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1

REFERENCES

[1] S.V. Nagalingam and G.C.I. Lin, "Latest developments in CIM", Robotics and Computer-Integrated Manufacturing, 1999. 15(6): p.423.

[2] N. Zhou, K. Xing, S.V. Nagalingam, and G.C.I. Lin, "Development of an Agent Based VCIM Resource Scheduling Process for Small and Medium Enterprises", Proceeding of the International MultiConference of Engineers and Computer Scientists 2010, Hong Kong, 17-19 March, 2010, pp. 39-44, Vol. 1.

[3] D. Wang, S.V. Nagalingam, and G.C.I. Lin, "Development of an agent-based Virtual CIM architecture for small and medium manufacturers", Robotics and Computer Integrated Manufacturing, Vol. 23, Issue 1, February 2007, pp.1-16.

[4] S.V. Nagalingam, G.C.I Lin, and D. Wang, Resource Scheduling for a Virtual CIM System. Process Planning and Scheduling for Distributed Manufacturing, Springer-Verlag; May 2007, pp 269-292.

[5] D. Wang, S.V. Nagalingam, and G.C.I Lin, "Production and logistic planning towards agent-based Virtual CIM system in an SME network", Proceeding of the Eighth International Conference on Manufacturing & Management, 8-10 December 2004, Gold Coast, Australia.

[6] N. Zhou, K. Xing, and S.V. Nagalingam, "An Agent-based Cross-enterprise Resource Planning for Small and Medium Enterprises", IAENG International Journal of Computer Science, August 2010, Vol. 37, Issue 3.

[7] N. Zhou, S.V. Naglingam, K. Xing, and G.C.I. Lin, Inside Virtual CIM: Multi-Agent Based Resource Integration for Small to Medium Sized Manufacturing Enterprises. Intelligent Control and Computer Engineering/Lecture Notes in Electrical Engineering, 2011, Vol. 70, Springer-Verlag, New York, USA.

[8] G.C.I. Lin, "The latest research trends in CIM" Keynote paper, Proceedings of the 4th International Conference on Computer Integrated Manufacturing, Singapore, 1997, pp. 26-33.

[9] D. Wang, S.V. Nagalingam, and G.C.I. Lin, "Implementation approaches for a multi-agent Virtual CIM System". Proceeding of the 9th International Conference on Manufacturing Excellence, 2003. Melbourne, Australia.

[10] D. Wang, S.V. Nagalingam, and G.C.I. Lin, "Development of a parallel processing multi-agent architecture for a virtual CIM system", International Journal of Production Research, Volume Number 17, September 2004 , pp. 3765-3785.

[11] D. Wang, S.V. Nagalingam, and G.C.I Lin, "A Novel Multi-Agent Architecture for Virtual CIM System",. International Journal of Agile Manufacturing System, 2005. 8(8): p. 69-82.

[12] S.D. Dao and R. Marian, "Optimisation of Precedence-Constrained Production Sequencing and Scheduling Using Genetic Algorithms", Proceeding of the IAENG International Conference on Artificial Intelligence and Applications, Hong Kong, 16-18 March 2011.

[13] S.D. Dao and R. Marian, "Modeling and Optimization of Precedence-

Constrained Production Sequencing and Scheduling Using Multi-Objective Genetic Algorithm", Proceeding of the International Conference of Computational Intelligence and Intelligent Systems, London, U.K, 6-8 July 2011.

[14] S.D. Dao and R. Marian, "Modeling and Optimisation of Precedence-Constrained Production Sequencing and Scheduling for Multiple Production Lines Using Genetic Algorithm", Journal of Computer Technology and Application, David publishing Company, USA, Vol. 2, No. 6, 2011.

[15] S.D. Dao and R. Marian, Genetic Algorithms for Integrated Optimisation of Precedence-Constrained Production Sequencing and Scheduling. Electrical Engineering and Intelligent System/Lecture Notes in Electrical Engineering, 2012, Vol. 130, Springer-Verlag, New York, USA.

[16] M. Gen and R. Cheng, Genetic algorithms and engineering design. Wiley, New York, 1997.

[17] R. Marian, L.H.S. Luong, and A. Raknoi, "Optimisation of distribution networks using Genetic Algorithms Part 1 - Problem modeling and automatic generation of solutions", International journal of manufacturing technology and management, Special issue on: Latest concepts and approaches in logistics engineering and management, Vol. 15, No 1, 2008, pp 64 - 83.

[18] R. Marian, L.H.S. Luong, and A. Raknoi, "Optimisation of Distribution Networks Using Genetic Algorithms, Part 2 - The Genetic Algorithm and Genetic Operators", International Journal of Manufacturing Technology and Management, Special Issue on: Latest concepts and approaches in logistics engineering and management, Vol. 15, No 1, 2008, pp 84-101.