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

IJARAI

International Journal of
Advanced Research in Artificial Intelligence

Volume 2 Issue 7

www.ijarai.thesai.org

A Publication of
The Science and Information Organization

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN ARTIFICIAL INTELLIGENCE

THE SCIENCE AND INFORMATION ORGANIZATION

www.thesai.org | info@thesai.org












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 2 Issue 7 July 2013
ISSN: 2165-4069(Online)
ISSN: 2165-4050(Print)

Editorial Board

Reviewer Board Members

CONTENTS

Comparative study between the proposed shape independent clustering method and the conventional methods (K-means and the other)

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

Cahya Rahmad [2]
Electronic Engineering Department
The State Polytechnics of Malang,
East Java, Indonesia

***Abstract*—Cluster analysis aims at identifying groups of similar objects and, therefore helps to discover distribution of patterns and interesting correlations in the data sets. In this paper, we propose to provide a consistent partitioning of a dataset which allows identifying any shape of cluster patterns in case of numerical clustering, convex or non-convex. The method is based on layered structure representation that be obtained from measurement distance and angle of numerical data to the centroid data and based on the iterative clustering construction utilizing a nearest neighbor distance between clusters to merge. Encourage result show the effectiveness of the proposed technique.**

Keywords—clustering algorithms; mlccd; shape independence clustering;

I. INTRODUCTION

Clustering is one of the most useful tasks in data mining process for discovering groups and identifying interesting distributions and patterns in the underlying data and is also broadly recognized as a useful tool in many applications. It has been subject of wide research since it arises in many application domains in engineering, business and social sciences. Researchers of many disciplines have addressed the clustering problem. The objective of algorithms is to minimize the distance of the objects within a cluster from the representative point of this cluster[1]. The clustering means process to define a mapping, f: D → C from some data D={t1, t2,t3,…tn} to some clusters C={c1, c2,c3,…, cn} based on similarity between ti.

Clustering is a central task for which many algorithms have been proposed. The task of finding a good cluster is very critical issues in clustering[2]. Cluster analysis constructs good clusters when the members of a cluster have minimize distances (Intra-cluster distances are minimized or internal homogeneity) are also not like members of other clusters (Inter-cluster distances are maximized). Clustering algorithms can be hierarchical or partitional[3].

In hierarchical clustering, the output is a tree showing a sequence of clustering with each cluster being a partition of the data set. Hierarchical clustering proceeds successively by either merging smaller clusters into larger ones. The result of the algorithm is a tree of cluster, called dendogram, which shows how the cluster are related, by cutting the dendogram at a desired level, a clustering of data the data items into disjoint groups is obtained [1]. The most well known methods for clustering is K-means developed by Mac Queen in 1967. K-means is a partition clustering method that separates data into k mutually excessive groups. Partitional clustering attempts to directly decompose the data set into a set of disjoint clusters. By iterative such partitioning, K-means minimizes the sum of distance from each data to its clusters. This method ability to cluster huge data, and also outliers, quickly and efficiently[4]. However, K-means algorithm is very sensitive in initial starting points. Because of initial starting points generated randomly, K-means does not guarantee the unique clustering results. However, many of above clustering methods require some additional user specified parameters, such as shape of cluster, optimal number, similirity threshold etc. In this paper, a new simple algorithm of numerical clustering is proposed. The proposed method, in particular, for a shape independent clustering that can also be applied in the case of condensed clustering.

II. PROPOSED CLUSTERING ALGORITHM

In this paper, a new simple algorithm of numerical clustering is proposed. The proposed method, in particular, is for a shape independent clustering. The proposed method can also be applied in the case of condensed clustering. we implemented multi-layer centroid contour distance (mlccd)[5] with some modification to the clustering.

The algorithm is described as follow:

1) Begin with an assumption that every point "n" is it's own cluster c_i, where i =1,2,…n

2) Calculate the centroid location

3) Calculate the angle and distance for every point to the centroid

4) Make multi layer centroid contour distance based on step 2 and 3

5) Set i = 1 as initial counter

6) Increment i = i + 1

7) Measure distance between cluster in the location i with cluster in the location i-1 and cluster in the location i+1.

8) Merge two cluster become one cluster base on the a nearest neighbor distance between clusters (see equation 2.4) as shown in the step 7

9) Repeat from step 6 to step 8 while i< 360

10) Repeat step 5 to step 9 until the required criteria is met

Firstly in the step 1, let every point "n" is it's own cluster, if there are n data it mean there are n cluster. Secondly, in the step 2 obtain the centroid by using equation 2.1 in this case every point have contribution to find the centroid location. Step3 Calculate the distance and angle between centroid to the every point, the angle is obtained by using arctangent of dy/dx, dy is differences y position between position y of every point and position y in the centroid and dx is differences x position between position x of every point and position x in the centroid see equation 2.2. The distance between centroid to the every point of data can be obtained by using equation (2.3). The next step By using the angle of every point and its distance then make the multi layer centroid contour distance (mlccd). The next process set i as counter start from 1 to 360 then calculate distance by using Euclidian distance between every cluster that pointed by i and every cluster that pointed by i-1 and i+1. After the distance was calculated then merge two cluster become one cluster base on the nearest distance.

position of the centroid is:

$$X_c = \frac{X_1+X_2+X_3+....+X_n}{n} \quad , \quad Y_c = \frac{Y_1+Y_2+Y_3+....+Y_n}{n} \qquad (1)$$

Where:

X_c = position of the centroid in the x axis

Y_c = position of the centroid in the y axis

n = Total point or data (every point have x position and y position)

angle every point to the centroid is:

$$\text{Angle} = \text{arctangent}\left(\frac{dy}{dx}\right) \qquad (2)$$

Where:

$dx = x - x_c$

$dy = y - y_c$

(x_c, y_c) = Position centroid

(x,y) = Position point or data

The distance from the centroid to every point is :

$$\text{Dis} = \sqrt{dx^2 + dy^2} \qquad (3)$$

Where:

$dx = x - x_c$

$dy = y - y_c$

(x_c, y_c) = Position centroid

(x,y) = Position point or data

The a nearest neighbor distance between clusters is calculated by using Euclidian distance these equation is commonly used to calculate the distance in case of numerical data sets [6]. For two-dimensional dataset, it performs as:

$$d(A, B) = \sqrt{\sum_{i=1}^{n} |A_i - B_i|^2} \qquad (4)$$

III. Experiment Result

In order to analyze the accuracy of our proposed method we represent error percentage as performance measure in the experiment. It is calculated from number of misclassified patterns and the total number of patterns in the data sets[7] (see the equation 3.1). We compare our method with the conventional clustering methods (k-means and other) to the same dataset.

$$\text{Error} = \frac{number\ of\ misclassified\ patterns}{number\ of\ patterns}\ X\ 100\% \qquad (5)$$

We applied the proposed method to solve some various shape independent cases and also tried to apply the proposed method in condensed clustering case.The dataset consist Circular nested dataset contain 96 data and 3 cluster with cluster1 contain 8 data, cluster 2 contain 32 data and cluster 3 contain 56 data, inter related dataset contain 42 data and 2 cluster with cluster 1 contain 21 data and cluster 2 contain 21 data, S shape dataset contain 54 data and 3 cluster with cluster1 contain 6 data, cluster 2 contain 6 data and cluster 3 contain 42 data. u shape dataset contain 38 data and 2 cluster with cluster 1 contain 12 data and cluster 2 contain 26 data. The 2 cluster Random dataset contain 34 data and 2 cluster with cluster1 contain 15 data and cluster 2 contain 19. The 3 cluster condense dataset contain 47 data and 3 cluster with cluster1 contain 16 data, cluster 2 contain 14 data and cluster 3 contain 17. The last data is 4 cluster condense dataset contain 64 data and 4 cluster with cluster1 contain 14 data, cluster 2 contain 18 data cluster 3 contain 15 data and cluster 4 contain 17 data. In the Figure 1 up to Figure 7, A is unlabelled data, B is labelled data by using proposed method and C is labelled data by using K-mean method.

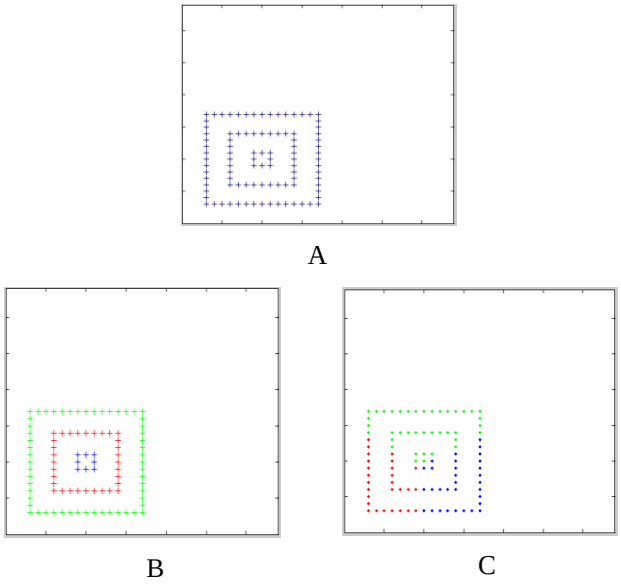


Fig.1. Circular nested dataset contain 96 data and 3 cluster with cluster1 contain 8 data, cluster 2 contain 32 data and cluster 3 contain 56 data

In Figure 1 by using proposed method there is no misclassified, by using k-mean there are misclassified in the cluster1 4 miss, cluster2 29 and cluster3 35miss, average error is 67.7%.

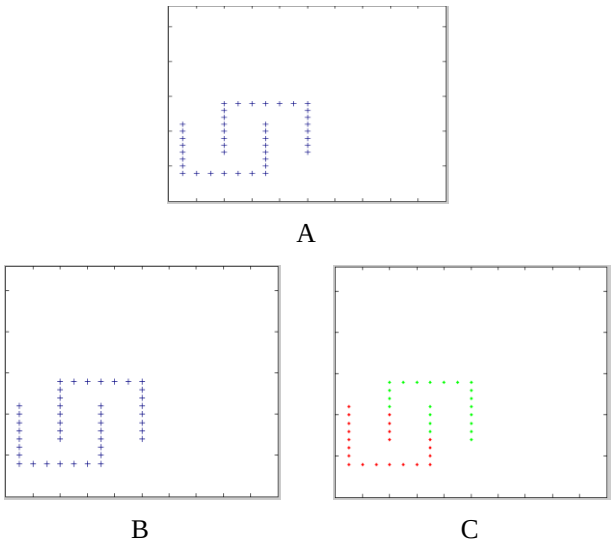


Fig.2. inter related dataset contain 42 data and 2 cluster with cluster 1 contain 21 data and cluster 2 contain 21 data

In figure 2 by using proposed method there is no misclassified, by using k-mean there are misclassified in the cluster1 4 miss and cluster2 4 miss, average error is 19.04%.

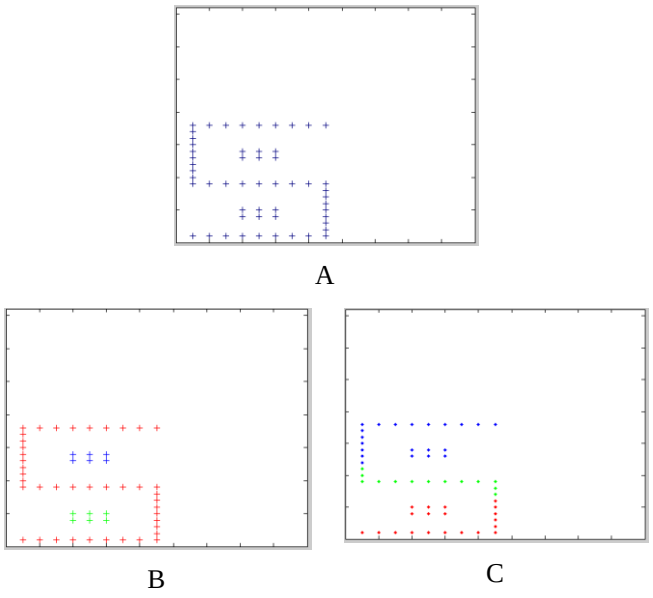


Fig.3. S shape dataset contain 54 data and 3 cluster with cluster1 contain 6 data, cluster 2 contain 6 data and cluster 3 contain 42 data

In Figure 3 by using proposed method there is no misclassified, by using k-mean there are misclassified in the cluster1 0 miss, cluster2 0 and cluster3 29miss, average error is 23.01%.

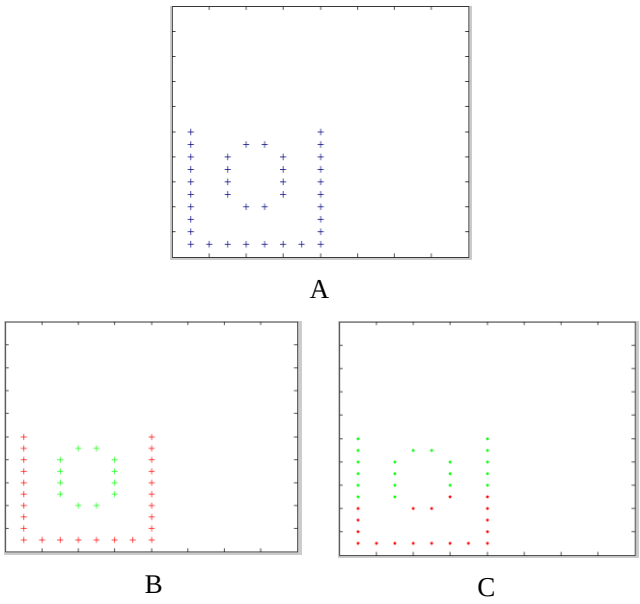


Fig.4. u shape dataset contain 38 data and 2 cluster with cluster 1 contain 12 data and cluster 2 contain 26 data.

In Figure 4 by using proposed method there is no misclassified, by using k-mean there are misclassified in the cluster1 3 miss and cluster2 11 miss, average error is 33.65%.

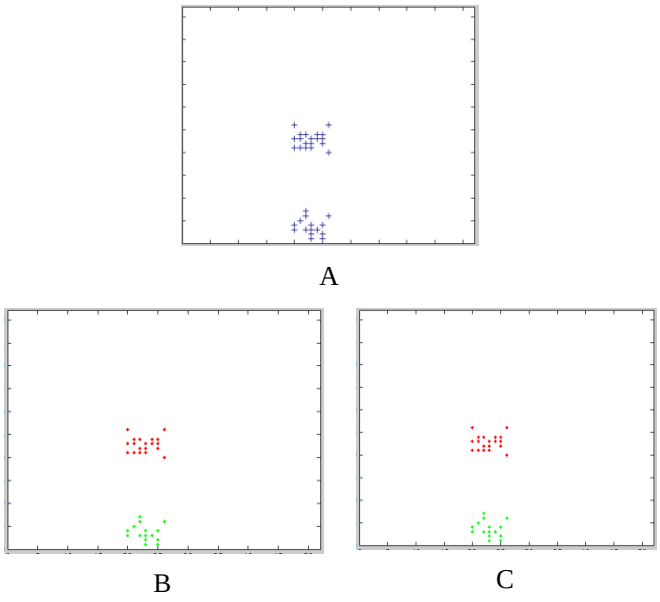


Fig.5. The 2 cluster Random dataset contain 34 data and 2 cluster with cluster1 contain 15 data and cluster 2 contain 19.

In Figure 5 by using proposed method and k-mean there is no misclassified average error is 0%.

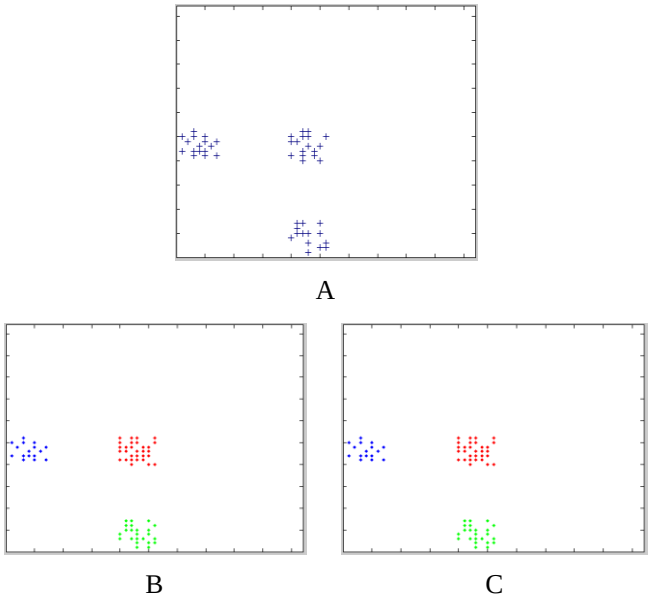


Fig.6. The 3 cluster condense dataset contain 47 data and 3 cluster with cluster1 contain 16 data, cluster 2 contain 14 data and cluster 3 contain 17

In Figure 6 and Figure 7 by using proposed method and k-mean there is no misclassified average error is 0%.

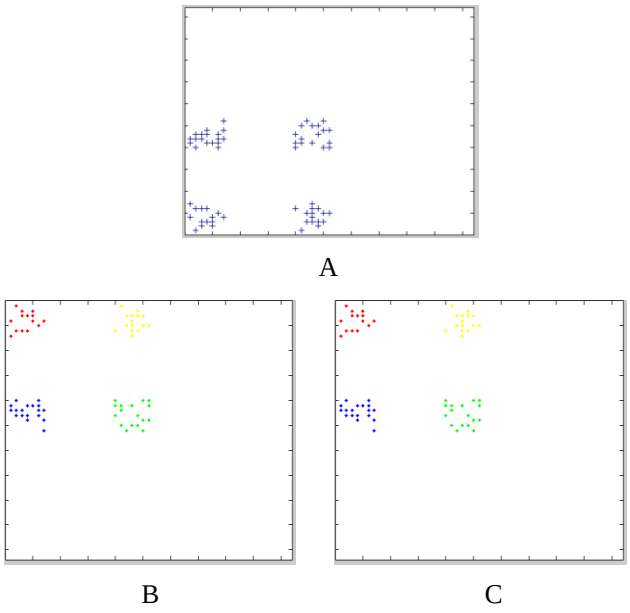


Fig.7. The 4 cluster condense dataset contain 64 data and 4 cluster with cluster1 contain 14 data, cluster 2 contain 18 data cluster 3 contain 15 data and cluster 4 contain 17 data.

We also implement some other clustering method that are Hierarchical algorithms (Single Linkage, Centroid Linkage, Complete Linkage and Average Linkage) to the same dataset the average result shown in the Table2.

TABLE I. AVERAGE ERROR IN PERCENT OF CLUSTERING BY USING LAYERED STRUCTURE REPRESENTATION AND CLUSTERING BY USING K-MEAN

Dataset	Propose method Error in %	k-mean Error in %
Circular nested	0	67.7
inter related	0	19.04
S shape	0	23.01
u shape	0	33.65
2 cluster condense	0	0
3 cluster condense	0	0
4 cluster condense	0	0
Average	0	20.48

TABLE II. AVERAGE ERROR IN PERCENT OF CLUSTERING BY USING HIERARCHICAL CLUSTERING

Clustering Method	Single linkage	Centroid linkage	Complete linkage	average linkage
Average	19.32	57.82	57.94	58.21

In the Table 1 and Table2 are shown that the proposed method compare with other method allows to identifying any shape of cluster as well as condensed dataset.

IV. CONCLUSION

In this work a new clustering methodology has been introduced based on the layered structure representation that be obtained from measurement distance and angle of numerical data to the centroid data and based on the iterative clustering construction utilizing a nearest neighbor distance between clusters to merge. From the experimental results with some various clustering cases, the proposed method can solve the clustering problem and create well-separated clusters. It is found that the proposed provide a consistent partitioning of a dataset which allows identifying any shape of cluster patterns in case of numerical clustering as well as condensed clustering.

ACKNOWLEDGEMENT

The authors would like to thank all laboratory members for their valuable discussions through this research.

REFERENCES

[1] M. Halkidi, "On Clustering Validation Techniques," *Journal of Intelligent Information Systems*, pp. 107–145, 2001.
[2] V. Estivill-Castro, "Why so many clustering algorithms: a position paper," *ACM SIGKDD Explorations Newsletter*, vol. 4, no. 1, pp. 65–75, 2002
[3] A. Jain, M. Murty, and P. Flynn, "Data clustering: a review," *ACM computing surveys (CSUR)*, vol. 31, no. 3, pp. 264–323, 1999.
[4] H. Ralambondrainy, "A conceptual version of the K-means algorithm," *Pattern Recognition Lett.*, pp. 1147–1157, 1995.
[5] K. Arai and C. Rahmad, "Content Based Image Retrieval by using Multi Layer Centroid Contour Distance," vol. 2, no. 3, pp. 16–20, 2013.
[6] P. A. Vijaya, M.N. Murty, and D. K. Subramanian, "An Efficient Hierarchical Clustering Algorithm for Large Data Sets," *Pattern Recognition Letters 25*, pp. 505–513, 2004.
[7] K. Arai and A. R. Barakbah, "Cluster construction method based on global optimum cluster determination with the newly defined moving variance," vol. 36, no. 1, pp. 9–15, 2007.

AUTHIORS 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 April 1974 to December 1978 and also was with National Space Development Agency of Japan from January, 1979 to March, 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 moved to Saga University as a Professor in Department of Information Science on April 1990. He was a 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 a councilor of Saga University for 2002 and 2003. He also was an executive councilor for the Remote Sensing Society of Japan for 2003 to 2005. He is an Adjunct Professor of University of Arizona, USA since 1998. He also is Vice Chairman of the Commission A of ICSU/COSPAR since 2008. He wrote 30 books and published 307 journal papers.

Cahya Rahmad, He received BS from Brawijaya University Indonesia in 1998 and MS degrees from Informatics engineering at Tenth of November Institute of Technology Surabaya Indonesia in 2005. He is a lecturer in The State Polytechnic of Malang Since 2005 also a doctoral student at Saga University Japan Since 2010. His interest researches are image processing, data mining and patterns recognition.

Effect of Driver Scope Awareness in the Lane Changing Maneuvers Using Cellular Automaton Model

Kohei Arai
Graduate School of Science and Engineering
Saga University
Saga, Japan

Steven Ray Sentinuwo
Department of Electrical Engineering
Sam Ratulangi University
Manado, Indonesia

***Abstract*—This paper investigated the effect of drivers' visibility and their perception (e.g., to estimate the speed and arrival time of another vehicle) on the lane changing maneuver. The term of scope awareness was used to describe the visibility required by the driver to make a perception about road condition and the speed of vehicle that exist in that road. A computer simulation model was conducted to show this driver awareness behavior. This studying attempt to precisely catching the lane changing behavior and illustrate the scope awareness parameter that reflects driver behavior. This paper proposes a simple cellular automata model for studying driver visibility effects of lane changing maneuver and driver perception of estimated speed. Different values of scope awareness were examined to capture its effect on the traffic flow. Simulation results show the ability of this model to capture the important features of lane changing maneuver and revealed the appearance of the short-thin solid line jam and the wide solid line jam in the traffic flow as the consequences of lane changing maneuver.**

Keywords—scope awareness; lane changing meneuver; speed estimation; spontaneous braking.

I. INTRODUCTION

The simulation model that can express the real traffic condition becomes the most important aspect in the field of traffic analysis and modeling. Study of traffic flow tries to capture and analyze the movement of individual vehicles between two points and the interactions between them. Traffic systems are characterized by a number of entities and features that make them hard to capture, analyze, control, and modify. The real traffic systems are formed by a combination of human interaction, that is interaction between driver entities, and human-environment interaction, such as driver interaction with the vehicle, with traffic information, and with the physical road condition. Studies about traffic and transportation have shown that driver behavior is one of the main contributors to some traffic event or phenomena. Our recent simulation study about the traffic flow showed that the traffic congestion can be influenced not only by the road capacity condition, but also by the driver behavior [1]. The other studies also found the strong relationship between the driver' speed behavior and accidents [2] [3] [4] [5]. Safe driving is a very important element for all the people on the road at any given time. Study of traffic accidents shows that human factors are a sole or a primary contributory factor in road traffic accidents [7]. There are two separate components that affect human factors in driving, driving skills and driving style [8]. Driving style has a direct relationship to the individual driving behavior. The U.S. Department of Transportation recently reported that driver behavior leading to lane-change crashes and near-crashes [6]. In some countries, the reckless driving behaviors such as sudden-stop by public-buses, tailgating, or vehicles which changing lane too quickly also could give an impact to the traffic flow. The lane changing maneuver is one of the phenomena in the highway. A Lane changing is defined as a driving maneuver that moves a vehicle laterally from one lane into another where both lanes have the same direction of travel. Lane changing maneuvers are occasionally performed in order to avoid hazards, obstacles, vehicle collision, or pass through the slow vehicle ahead. Changing lanes requires high attention and visual demand compared to normal highway of freeway driving due to the need to continually monitor areas around the subject vehicle [9]. However, in the real traffic situations there are some reckless drivers that changing lanes at the moment they signal or who make last minutes decision on the road. Frequent lane changing in roadway could affect traffic flow and even lead to accidents. The lane changing behaviors can be vary depend on the characteristic of the driver [10]. Some crashes accidents typically referred to as Look-But-Fail-To-See errors because drivers involved in these accidents frequently report that they failed to notice the conflicting vehicle in spite of looking in the appropriate direction, commonly occur when drivers change lanes [25]. This mean the driver typically use their perception in order to estimate the speed and the arrival time of the other vehicles before making a maneuver, e.g., lane changing maneuver. A psychology study has shown the accuracy level of this perception may contribute to both failures to detect the collision and to judge the crash risk (e.g., time-to-contact). From a certain distance, a short fixation may be enough to identify an approaching vehicle. Duration of gaze interpreted as the amount of time devoted to processing a stimulus, longer and shorter gazes reflect difficult and simple processing, respectively. Inaccuracy of the gazes duration are likely to reflect a failure to process these stimuli [26].

This paper was interested to investigate the effect of drivers' visibility and their perception (e.g., to estimate the speed and arrival time of another vehicle) on the lane changing maneuver. One purpose of this study was to examine how different driver visibility and scope awareness might affect traffic flow. We consider that the driver decision to make a lane

changing is influenced by the condition of both its current and target lane. The estimation about the gap with ahead and backward vehicle in target lane, includes their speed, will affect the human perception to make a safety lane changing. This paper introduces one of the driver behavior parameter; that is scope awareness parameter. The term of scope awareness was used to describe the visibility required by the driver to make a perception of road condition and the speed of vehicle that exist in that road. Since there are various types of driving skill and style of the drivers that exist in the roadway then the value of scope awareness probabilities could be vary. This studying attempt to precisely catching the lane changing behavior and illustrate the scope awareness parameter that reflects driver behavior. A computer simulation model was conducted to show this scope awareness behavior. In this simulation model, the scope awareness parameter reflected as the length of the road at the adjacent lane that is considered as safely area by the subject driver before making a lane changing.

The Cellular Automata model of Nagel and Schreckenberg (NaSch) [12] was improved to better capture the effect of scope awareness that reflect drivers' behavior when making a lane changing. This NaSch model has been modified to describe more realistic movement of individual vehicle when make a lane changing maneuver. Moreover, the recent study of spontaneous braking behavior [1] has been enhanced through the investigation of its relationship with the driver's scope awareness behavior.

This paper is organized as follows. Some studies relating with traffic modeling principles and this study is quick reviewed in Section 2. Section 3 presents a short description of the theoretical aspect of traffic CA model. Section 4 explains about the proposed model. Section 5 contains simulation process and the results in the form of fundamental diagrams and space-time diagrams. Finally, in section 6, we present a summary and conclusion of this work.

II. Related Research Works

Due to the rapid development of computer technology then research about traffic simulation and modeling has increasingly grown. Computer simulation in traffic model has developed from a research tool of experts to a widely used technology for practitioners and researchers in the research, planning, demonstration, and development of traffic systems.

The increasing of computational speed and power make the scope of research of traffic simulation have been growing. Since the early 1950's, the research simulation have evolved from local road analysis into more complex systems where several type of parameters are integrated in one system. The research about traffic modeling can be divided into two categories: microscopic model and macroscopic model. Microscopic model described traffic behavior as resulting from discrete interaction between vehicles as entities. This microscopic model range from simple analytical models such as car-following model, to more detailed analytical models, such as FRESIM and NETSIM simulation software. While the macroscopic models concern to describe the aggregate traffic behavior phenomena by considering the fundamental relationships between vehicles speed, flow, and density. Macroscopic models include Input-Output, Simple Continuum, and Higher Order Continuum [27].

Most microscopic models (e.g., the car-following model) use the assumption the all the vehicles have a uniform driving behavior. These microscopic models use deterministic approach and, therefore difficult to capture inherent stochastic nature of real traffic. On the other hand, a major limitation of macroscopic models is their aggregate nature. The macroscopic models concern the traffic flow as continuous system, then these models cannot capture the discrete dynamic aspects that arise from vehicles interaction [27].

The interaction between vehicles has strong relationship with the driver behavior. Some research studies have shown that the driver behavior play an important role for the traffic events. One cause of those traffic events is due to the observations and reactions of drivers are governed by human perception and not by technology based sensor and monitoring systems. The emotional aspect of the driver contributes to the many situations in traffic such as car crashes and congestion [28]. Another study also shown that the driver behavior is a fundamental factor and a key source of complexity in predicting traffic network states unfolding over time [29].

Nagel et.al. [13], discusses two lane traffic and lane changing rules based on a cellular automata model. Furthermore, Arai 1 enhances the original NaSch model by introduced the spontaneous braking parameter as a driver behavior that periodically affect the traffic flow and lane changing decision. However, these models have not considered about drivers' visibility and speed estimation of the vehicles within the monitoring area which may have important influence on human' hazard perception and lane changing decision.

III. Traffic Cellular Automata model

One of the famous microscopic models for the simulation of road traffic flow is Cellular Automata (CA) model. CA is a discrete model studied in computability theory, mathematics, physics, complexity science, theoretical biology and microstructure modeling. In comparison with another microscopic model, the CA model proposes an efficient and fast performance when used in computer simulation [11]. CA is a dynamic model developed to model and simulates complex dynamical system. The set of CA rules may illustrate complex evolution patterns, such as time and space evolution in a system. Those evolutions can be shown just by use simple rules of CA. The CA model consists of two components, a cellular space and a set of state. The state of a cell is completely determined by its nearest neighborhood cells. All neighborhood cells have the same size in the lattice. Each cell can either be empty, or is occupied by exactly one node. There is a set of local transition rule that is applied to each cell from one discrete time step to another (i.e., iteration of the system). This parallel updating from local simple interaction leads to the emergence of global complex behavior.

Furthermore, the utilization of CA successfully explains the phenomenon of transportation. These so-called traffic cellular automata (TCA) are dynamical systems that are discrete in nature and powerful to capture all previously mentioned basic

phenomena that occur in traffic flows[11]. The one dimensional cellular automata model for single lane freeway traffic introduced by Nagel and Schreckenberg (NaSch) [12] is simple and elegant that captures the transition from laminar flow to start-stop waves with increasing vehicle density. This model shows how traffic congestion can be thought of as an emergent or collective phenomenon due to interactions between cars on the road, when the density of cars is high and so cars are close to each on average. The simplicity of the NaSch model has prompted the use of it for studying many traffic situations. The NaSch model also known as stochastic traffic cellular automaton (STCA) because it included a stochastic term in one of its rules. Like in deterministic traffic CA models (e.g., CA-184 or DFI-TCA), this NaSch model contains a rule that reflect vehicle increasing speed and braking to avoid collision. However, the stochasticity term also introduced in the system by its additional rule. In one of its rules, at each time-step t, a random number $\xi(t) \in [0,1]$ is generated from a uniform distribution. This random number is then compared with a stochastic noise parameter $p \in [0,1]$. For it is based on this probability p then a vehicle will slow down to (v(i) - 1) cells/time-step. According to NaSch model, the randomization rule captures natural speed fluctuations due to human behavior or varying external conditions [14].

In real traffic, most highways consist two or more lanes. Regarding this road condition, there are a few analytical models for multi-lane traffic. Nagatani was one of the first researchers that introduced a CA model for two lane traffic [24]. His model used deterministic approach and the maximum velocity $v_{max} = 1$. Then, building on Nagatani's model, Ricket et al. [15] considered a model with $v_{max} \geq 1$. Rickert investigated a simple model for two-lane traffic. Their model introduced the lane changing behavior for two lanes traffic. They proposed a symmetric rule set where the vehicle changes lanes if the following criteria are fulfilled:

$$gap(i) < l \quad (1)$$

$$gap_0(i) > l_0 \quad (2)$$

$$gap_{0,back}(i) > l_{0,back} \quad (3)$$

The variable gap(i), $gap_0(i)$, and $gap_{0,back}(i)$ denote the number of empty cells between the vehicle and its predecessor on its current lane, and forward gap on the desired lane, and backward gap on the desired lane, respectively. Rickert also used the parameters which decide how far the vehicle look ahead in current lane for l, ahead on the desired lane for l_0, and how far the vehicle look back on the desired lane for $l_{0,back}$.

The advance analysis about lane-changing behavior has been done, which includes symmetric and asymmetric rules of lane-changing [16-21]. Symmetric rule can be considered as rules that threat both lanes equally, while asymmetric rule can be applied in special characters highway, like German highways simulation [22], where lane changes are dominated by right lane rather than left lane. While the NaSch model could reproduce some of basic phenomenon observed in real traffic situations such as the start-stop waves in congested traffic, but it has been observed that the base NaSch model lacks the ability to produce other more realistic traffic patterns [23].

In this paper we consider the parameter of scope awareness of the driver that occur in the real traffic condition. Here, in the real traffic situations, this parameter of scope awareness has a strong relationship with human perception in order to make a lane changing maneuver decisions.

IV. Model Definition

This proposed model uses two-lane highway with unidirectional traffic character in periodic boundaries condition. Two-lane model is necessary in order to accommodate the lane changing behavior in the real traffic condition. A one-dimensional chain of L cells of length 7.5 m represents each lane. This value is considered as the length of vehicle plus the distance between vehicles in a stopped position. A one-lane loop consists of 10^3 cells. There are just two possibility states of each cell. Each cell can only be empty or containing by just one vehicle. The speed of each vehicle is integer value between $v = 0,1, \ldots, v_{max}$. In this model, all vehicles are considered as homogeneous then have the same maximum speed $v_{max} = 5$. The speed value number corresponds to the number of cell that the vehicle proceeds at one time step. The state of a road cell at the next time step, form t to $t + 1$ is dependent on the states of the direct frontal neighborhood cell of the vehicle and the core cell itself of the vehicle.

Rickert et al. [15], among others, have discussed about criteria of safety by introduced the parameters which decide how far the vehicle looks ahead on current lane, looks ahead on desired lane, and looks back on desired lane. Those criteria have to be fulfilled before a vehicle makes a lane changing. However, in real traffic condition, these criteria of safety rules by Rickert are not sufficient to describe driver's behaviors in highway traffic. This paper introduces a new additional parameter to accommodate the driver behavior when making a lane changing. In addition to considering the gap of cell that consists of vehicle, we also consider about the speed parameter of the other vehicle that situated in the desired lane. In this paper, we discuss in more detail the parameter of scope awareness S_a that reflects the various characters of driver. This scope awareness parameter takes into account the dynamic characteristic of the driver while decide to make a lane changing. Here, the smaller S_a value reflect the degree of driver aggressiveness and awareness. Figure 1 describes the scope awareness definition from the perspective of vehicle 1.

The updating rule for lane changing maneuver is done according to a set of rules. The set of rules of the lane changing maneuver is analogous as the liquid movement. Compare to the lane changing model of Ricket et al. [15], there are two basic differences rules in our model. The first one, as the result of traffic conditions ahead of subject driver (eq. (4)), the subject vehicle would consider changing its lane not only due to the comparison value between number of gap and condition which decide how far the vehicle look ahead in current lane (eq. (1)), but also depending on the current speed of the subject vehicle that can be vary based on traffic situation. Another difference is the scope awareness value (S_a). The subject vehicle would consider the velocity of every vehicle that situated along its

scope awareness area then decide whether possible or not to change the lane (eq. (8)).

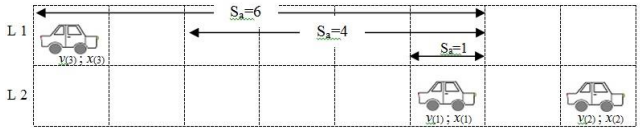


Fig. 1 Schematic definition diagram of scope awareness S_a from the perspective of vehicle (1) in its current speed and position $v_{(1);}\ x_{(1)}$.

At the beginning of each iteration, the subject driver checks whether a lane changing is desirable or not. The subject driver looks ahead to check if the existing gap in the current lane can accommodate his current speed. If not, then due to the randomness number of percentage ratio, the subject driver decides whether he will maintain or decelerate the vehicle speed due to the existing gap number or change his lane. When the subject driver chooses to change lanes, then he looks sideways at the other lane to check whether the cell next to the subject vehicle is empty and the forward gap on the other lane is equal or longer than his current lane. If one cell is unoccupied or free-cell then its state is 0. Moreover, the subject driver also looks back at the other lane to check road condition. In the real traffic situation, a subject driver also has to look back on the other lane in order to estimate the velocity of the following vehicle to avoid a collision. Eq. (8) accommodates the driver behavior that estimate the velocity of vehicle at the moment before making a lane changing.

As mentioned before, this paper uses the parameter of scope awareness S_a which decide how far the coverage area on the desired lane that is considered as the scope of awareness by the driver. If there is another vehicle within the area of scope awareness then the subject driver estimates the speed of the vehicle in order to avoid collision during the lane changing maneuver. The subject driver will make a lane changing maneuver if the speed of the vehicle that located within the area of scope awareness is less than the existing gap. The lane changing rules can be summarized as follows:

$$gap_{same} < v_{current} \tag{4}$$

$$cell_{next} = 0 \tag{5}$$

$$rand() < p_{change} \tag{6}$$

$$gap_{target} > gap_{same} \tag{7}$$

$$v_{Vehicle,back} \leq gap_{back}\ ;\ X(vehicle_{back}) \epsilon\ S_a \tag{8}$$

The lane changing rules are applied to vehicle that change from right lane to left lane and conversely. Vehicle is only move sideways and it does not advance. Once all the lane changing maneuvers are made then the updating rules from a single lane model are applied independently to each lane. Fig. 2 shows the schematic diagram of lane changing operation. In this fig. 2, the subject vehicle $v_{(1)}$; $x_{(1)}$ is assumed that have current speed $v_{(1)}^{\ t} = 3$ cell per time step and the parameter of scope awareness $S_a = 4$ cells.

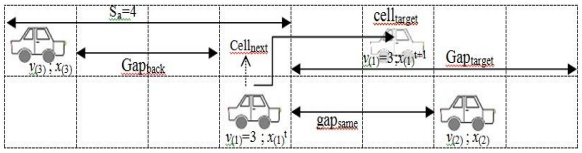


Fig. 2 Schematic diagram of a lane changing operation.

In order to avoid the introduction of any unrealistic artifacts in the simulation then this proposed model uses eq. (7) to express the more realistic lane changing decision. According to eq. (7), the driver must consider that the forward gap in the desired lane is more than the gap in the current lane. This consideration is important because this proposed model uses the different desired velocities into the vehicles.

Once the lane changing maneuvers are made to all possibility vehicles then the updating rules from a single lane model are applied independently to each lane. Together with a set of lane changing rules, the road state can be obtained by applying the following rules to all by parallel updated:

$$Acceleration:\ v(i) \rightarrow min(v(i)+1,\ v_{max}) \tag{8}$$

$$Deceleration:\ v(i) \rightarrow min(v(i),\ gap_{same}(i)) \tag{9}$$

$$Driving:\ x(i) \rightarrow x(i)+v(i) \tag{10}$$

V. Simulation Results

The simulation starts with an initial configuration of N vehicles, with fixed distributions of positions on both lanes. This simulation uses the same initial velocity for all vehicle $v_{min} = 0$ and the maximum vehicle speed has been set to $v_{max} = 5$ cell/time-step. The velocity corresponds to the number of cells that a vehicle advances in one iteration. Many simulations performed with different density ρ. The density ρ can be defined as number of vehicles N along the highway over number of cells on the highway L.

This traffic model uses close (periodic) boundary conditions. This means that during one simulation, the total number of vehicles on the highway cannot change. Vehicles go from left to right. If a vehicle arrives on the right boundary then it moves to the left boundary. Since this model assumes symmetry character of the both lanes then the traffic flow characteristics on both lanes are identical.

A. Traffic Flow

In order to examine the effect of scope awareness on the traffic flow then the model was simulated over 1000 iterations on 10^3 cells for all possibility density level. The flow indicates the number of number of moving vehicles per unit of time. Along with the study of this proposed model, this paper also conducted a comparison study for the case of traffic without using scope awareness parameter. Results from the simulation are summarized in fig. 3. After 1000 time step, when the system reaches a stationary velocity state then the flow was computed. The whole process then repeated over 50 times for both each density level and each scope awareness value to make statistics and the flow-density diagram was obtained.

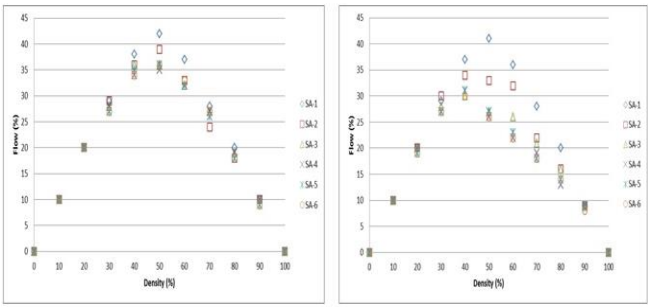


Fig. 3 The average flow-density diagram of the proposed model (left) is compared to a two-lane traffic system without using scope awareness parameter (right).

A number of interesting observations can be made:

- The proposed model reproduces a recognizable diagram of flow towards density relationship. Flow is linearly increasing together with the increases in density level. A maximum flow level is achieved at density level ρ=0.5 for each value of S_a. After reaches the critical point of flow at ρ=0.5, the flow at each level of S_a becomes linearly decreasing in density. In other words, the laminar flow turn into back travelling start-stop waves after density level ρ=0.5. Another thing that also interest is in the scope awareness value Sa=3, Sa=4, and Sa=5, this simulation produced almost the same flow level at all density levels. Scope awareness value Sa=1 reached the highest number of flow. This may happen because in the Sa=1, the driver can be described as the most aggressive driver, who makes a lane changing maneuver with only consider the empty area beside him. This behavior also confirms the result on fig. 4 that compared the number of lane changing for each value of scope awareness.
- Compared to the model without scope awareness consideration (Fig. 3-right diagram), the usage of S_a parameter produced a better flow of vehicles, especially above density ρ=0.4. This Sa parameter maintained the traffic to keep flowing by carefully calculate the appropriate time to make a lane changing decision, thus the lane changing maneuver does not disturb the traffic in the target lane.
- Since the parameter of scope awareness has a strong relationship with the lane changing decision then fig. 4 shown the ratio of lane changing number over density. The results of each S_a value are not surprising. If the distance of scope awareness becomes longer then the lane changing number becomes lower. However, the surprise thing is the behavior of the lane changing variance. For each changes of S_a value into the higher one, the critical point of maximum lane changing decreases almost half than before. The critical point of maximum lane changing for each S_a value is same at ρ=0.2, except for S_a=5 at ρ=0.3. This is due to the fact that the chances of the vehicles to change the lane become fewer caused by density increases.

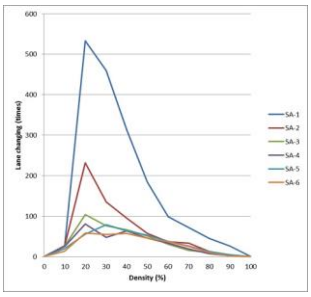


Fig. 4 The ratio of lane changing number over density.

B. *Space-Time diagram*

In order to explore more clearly the effect of scope awareness on the traffic flow then the space-time diagram was reproduced. The space-time diagram represents the location of the vehicles at the certain time. This paper conducted the space-time diagram for density ρ=0.25, ρ=0.5, and ρ=0.75. These three values of density assumed as the light traffic, moderate traffic, and heavy traffic in the real traffic condition, respectively. Fig. 5, fig. 6, and fig. 7 show the result for each density at the all values of scope awareness. To make the comparison fairly then this simulation used initial fixed distributions of positions of the vehicles. The horizontal axis represents space and the vertical axis represents the time. Vehicles go from left to right (space axis) and from top to bottom (time axis).

In the light traffic condition ρ=0.25 (fig.5), it can be seen that the increases of scope awareness distance affect the vehicles flow. Free flow phase showed in Sa=1 diagram (fig.5-a), which are drawn as light area and have a more shallow negative inclination. However, when the Sa value was increased then some solid area starting appear. This solid area with steep positive inclination reflects the traffic jam. One can observed from fig. 5, there are many regions that show the high frequency of short vehicle life lines appearing and disappearing that indicate the great number of lane changing at this traffic density ρ=0.25. Once the scope awareness increases then this frequency of short vehicle life lines become smaller than before (fig. 5-f).

In the moderate traffic ρ=0.5 (fig. 6), the phenomena that showed in the fig. 5 also appear in this density. The diagram of Sa=1 (fig. 6-a) shows the appearance of slight traffic jam distributed in the whole simulation area. In this density level, a solid moving jam appears since Sa=2 (fig. 6-b). Along with a wide solid line, there are also some short-thin solid lines appear in the diagram of Sa=2. While in the diagram of Sa=6 (fig. 6-f), some of these short-thin solid lines disappeared. Refer to the fig. 4., at the certain density value, once the S_a value was increased then the number of lane changing decreased. One can be observed that the short-thin solid line caused by the lane changing maneuver of another vehicle from adjacent lane, which resulted the subject vehicle has to make a spontaneous braking in order to avoid collision. As the result of this spontaneous braking causing another following vehicles has to adjust or decrease their speed with the vehicle ahead. This phenomenon produces a short traffic jam. On the other hand, the wide solid line appeared as a result of deceleration into the

minimum speed of the vehicle as the consequence of the reduced opportunities for lane changing maneuver.

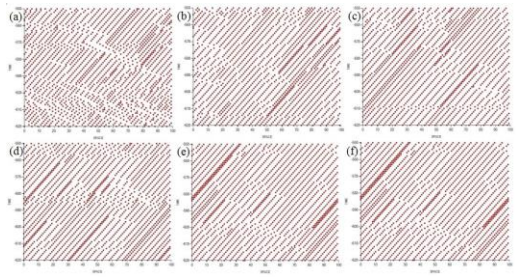


Fig. 5 Space-time diagram for light traffic condition (density ρ = 25%). Lane changing probability 100%. (a) for Scope awareness Sa=1 ; (b) for Scope awareness Sa=2 ; (c) for Scope awareness Sa=3 ; (d) for Scope awareness Sa=4 ; (e) for Scope awareness Sa=5 ; (f) for Scope awareness Sa=6.

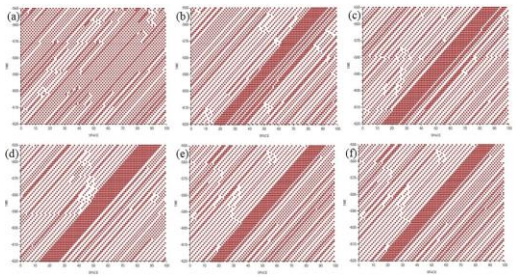


Fig. 6 Space-time diagram for moderate traffic condition (density ρ = 25%). Lane changing probability 100%. (a) for Scope awareness SA=1 ; (b) for Scope awareness SA=2 ; (c) for Scope awareness SA=3 ; (d) for Scope awareness SA=4 ; (e) for Scope awareness SA=5 ; (f) for Scope awareness SA=6.

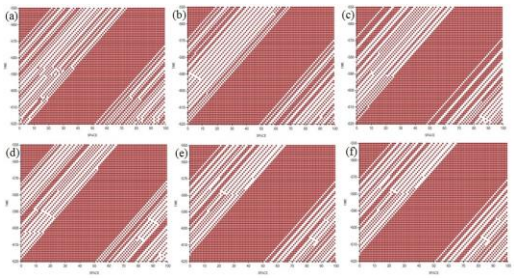


Fig. 7 Space-time diagram for heavy traffic condition (density ρ = 25%). Lane changing probability 100%. (a) for Scope awareness SA=1 ; (b) for Scope awareness SA=2 ; (c) for Scope awareness SA=3 ; (d) for Scope awareness SA=4 ; (e) for Scope awareness SA=5 ; (f) for Scope awareness SA=6.

However, in the heavy traffic condition (Figure 7), this model showed that the differences value of scope awareness did not affect the traffic condition. In this traffic condition, the opportunity to make a lane changing is very small. This result imply that in the heavy traffic condition, the driver characters that related to lane changing style have no influence to the traffic condition.

VI. Summary and Conclusion

This paper has presented a simple model of the traffic cellular automata to describe a driver behavior in a two lane highway model. The term of scope awareness introduced to reflect the visibility required by the driver to make a perception of a road condition and the speed of vehicle that exist within the certain area of the road before making a lane changing maneuver. The relation between flow-density and space-time has been investigated in order to examine the effect of scope awareness parameter in the traffic flow. Some conclusions can be observed from this study:

- This model describes the realistic traffic situation, in particular capture the situation when driver make a lane changing maneuver. Compared to the conventional approach, the usage of scope awareness model approach produce a better flow of vehicles.
- The various value of the scope awareness may represent the characteristic and the experience level of the drivers. The increases of the scope awareness value means the driver become more aware to estimate the road condition in order to make a lane changing maneuver.
- This proposed model has revealed the phenomena of the short-thin solid line jam and the wide solid line jam in the traffic flow. This study found that the short-thin solid line caused by the lane changing maneuver of another vehicle from adjacent lane which resulted the subject vehicle has to make a spontaneous braking in order to avoid collision. As the result of this spontaneous braking causing another following vehicles has to adjust or decrease their speed with the vehicle ahead. This phenomenon then produces a short queue of vehicles. On the other hand, a wide solid line appeared as a result of deceleration into the minimum speed of the vehicle as the consequence of the reduced opportunities to make a lane changing maneuver.
- This simulation results showed that lane changing maneuvers with taking into account another vehicle speed could reduce the level of traffic congestion. However, in the heavy traffic (high dense) situation, the opportunity to make a lane changing is small, so that the congestion will always exist.

By taking into consideration the scope awareness parameter, the traffic cellular automata model proposed here can reflect certain characteristics of lane changing maneuver in the real traffic situation.

This simulation result can serve as a reference for transportation planning, evaluation, and control. Moreover, this result will pave the way for accurate simulation of a more complex traffic system. Based on the result of this paper, the effect of road shape towards the vehicle deceleration will be studied hereafter.

References

[1] K. Arai and S. Sentinuwo, “Spontaneous-braking and lane-changing effect on traffic congestion using cellular automata model applied to the two lane traffic”, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 3 (8), 2012.

[2] Finch, D. J., Kompfner, P., Lockwood, C. R. & Maycock, G. (1994) Speed, speed limits and crashes. Project Record S211G/RB/Project Report PR 58. Transport Research Laboratory TRL, Crowthorne, Berkshire.

[3] Nilsson, G. (2004). Traffic safety dimensions and the power model to describe the effect of speed on safety. Lund Bulletin 221. Lund Institute of Technology, Lund.

[4] Salusjärvi, M., 1981. The speed limit experiments on public roads in Finland. Technical Research Centre of Finland. Publication 7/1981. Espoo, Finland.

[5] G. M. Fitch, S. E. Lee, S. Klauer, J. Hankey, J. Sudweeks, and T. Dingus. Analysis of Lane-Change Crashes and Near-Crashes. National Technical Information Service, Springfield, VA 22161, 2009.

[6] Lewin, I. (1982). Driver training a perceptual-motor skill approach. Ergonomics,25, 917–924.

[7] Elander, J., West, R., & French, D. (1993). Behavioral correlates of individual differences in road traffic crash risk: An examination of methods and findings. Psychological Bulletin, 113, 279–294.

[8] Shinar, D., Psychology on the road: The human factor in traffic safety. Wiley New York, (1978).

[9] Sun, D. J., & Elefteriadou, L. (2011). Lane-changing behavior on urban streets: a focus group-based study. Applied ergonomics, 42(5), 682–91. doi:10.1016/j.apergo.2010.11.001.

[10] X. G. Li, B. Jia, Z. Y. Gao, and R. Jiang, "A realistic two-lane cellular automata traffic model considering aggressive lane- changing behavior of fast vehicle," PhysicaA, vol. 367, pp. 479– 486, 2006.

[11] K. Nagel and M. Schreckenberg, "A cellular automaton model for freeway traffic," Journal of Physics I France, vol. 2, no. 12, pp.2221-2229, 1992

[12] K. Nagel, Wolf, Wagner, and Simon, "Two-lane traffic rules for cellular automata: A systematic approach," Physical Review E, vol.58, no.2, 1998.

[13] S. Maerivoet and B. D. Moor, "Transportation Planning and Traffic Flow Models," 05-155, Katholieke Universiteit Leuven, Department of Electrical Engineering ESAT-SCD (SISTA), July 2005.

[14] M. Rickert, K. Nagel, M. Schreckenberg, and A. Latour, "Two Lane Traffic Simulations using Cellular Automata," vol. 4367, no. 95, 1995.

[15] W. Knospe, L. Santen, A. Schadschneider, and M. Schrekenberg, "Disorder effects in cellular automata for two lane traffic," Physica A, vol. 265, no. 3-4, pp. 614–633, 1998.

[16] A. Awazu, "Dynamics of two equivalent lanes traffic flow model: selforganization of the slow lane and fast lane," Journal of Physical Society of Japan, vol. 64, no. 4, pp. 1071– 1074, 1998.

[17] E. G. Campri and G. Levi, "A cellular automata model for highway traffic," The European Physica Journal B, vol. 17, no. 1, pp. 159–166, 2000.

[18] L. Wang, B. H. Wang, and B. Hu, "Cellular automaton traffic flow model between the Fukui-Ishibashi and Nagel- Schreckenberg models," Physical Review E, vol. 63, no. 5, Article ID 056117, 5 pages, 2001.

[19] B. Jia, R. Jiang, Q. S. Wu, and M. B. Hu, "Honk effect in the two-lane cellular automaton model for traffic flow," Physica A, vol. 348, pp. 544–552, 2005.

[20] D. Chowdhury, L. Santen, and A. Schadschneider, "Statistical physics of vehicular traffic and some related systems," Physics Report, vol. 329, no. 4-6, pp. 199–329, 2000.

[21] W. Knospe, L. Santen, A. Schadschneider, and M. Schreckenberg, "A realistic two-lane traffic model for highway traffic," Journal of Physics A, vol. 35, no. 15, pp. 3369–3388, 2002.

[22] W. Knospe, L. Santen, A. Schadschneider, and M. Schreckenberg, "Empirical test for cellular automaton models of traffic flow," Phys. Rev. E, vol. 70, 2004.

[23] Nagatani, T., "Self Organization and Phase Transition in the Traffic Flow Model of a Two-Lane Roadway," Journal of Physics A, Vol. 26, pp. 781-787, 1993.

[24] Shahar, A., Van Loon, E., Clarke, D., & Crundall, D. Attending overtaking cars and motorcycles through the mirrors before changing lanes. Accident; analysis and prevention, 44(1), 104–10, 2012.

[25] Rayner, K., Warren, T., Juhasz, B.J., Liversedge, S.P. The effect of plausibility on eye movements in reading. Journal of Experimental Psychology: Learning, Memory and Cognition 30, 1290–1301, 2004.

[26] Benjaafar, S., & Dooley, K. (1997). Cellular automata for traffic flow modeling. Minneapolis, MN, University of. Retrieved from http://ntl.bts.gov/lib/21000/21100/21189/PB99103996.pdf.

[27] Laagland, J. (2005). How To Model Aggressive Behavior In Traffic simulation. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.66.2976&rep=rep1&type=pdf

[28] Paz, A., & Peeta, S. (2009). Information-based network control strategies consistent with estimated driver behavior. Transportation Research Part B: Methodological, 43(1), 73–96. doi:10.1016/j.trb.2008.06.007.

AUTHORS PROFILE

Kohei Arai 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 April 1974 to December 1978 and also was with National Space Development Agency of Japan from January, 1979 to March, 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 moved to Saga University as a Professor in Department of Information Science on April 1990. He was a counselor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was a councilor of Saga University for 2002 and 2003. He also was an executive councilor for the Remote Sensing Society of Japan for 2003 to 2005. He is an Adjunct Professor of University of Arizona, USA since 1998. He also is Vice Chairman of the Commission A of ICSU/COSPAR since 2008. He wrote 30 books and published 322 journal papers.

Steven Ray Sentinuwo, received the B.Eng. degree in Electrical Engineering from Sam Ratulangi University, and the M.Eng. degree in Information Engineering, from University of Indonesia, in 2004 and 2006, respectively. He is currently a PhD Student at Information Science in Saga University, Japan. His research interest includes robot path planning, traffic simulation and modeling, and information system management.

Validity of Spontaneous Braking and Lane Changing with Scope of Awareness by Using Measured Traffic Flow

Kohei Arai
Graduate School of Science and Engineering
Saga University
Saga, Japan

Steven Ray Sentinuwo
Department of Electrical Engineering
Sam Ratulangi University
Manado, Indonesia

***Abstract*—This paper presents the validation method and its evaluation of the spontaneous braking and lane changing with scope awareness parameter. By using the real traffic flow data, the traffic cellular automaton model that accommodate these two driver behaviors, e.g., spontaneous braking and driver scope awareness has been compared and evaluated. The real traffic flow data have been observed via video-recording captured from real traffic situation. The validation results shown that by accommodate spontaneous braking and scope awareness parameters, the model can produced traffic flow's accuracy value 83.9% compared to the real traffic flow data.**

Keywords—traffic model validation; spontaneous braking; scope awareness; traffic cellular automaton.

I. INTRODUCTION

Validation is one of the important processes in the field of simulation and modeling. The validation process is concerned with determining whether the conceptual simulation model is an accurate representation of the system under study. However, the validation process cannot be defined to result a perfect model, since the perfect one would be the real system itself [1]. Naturally, any model is the simplification of the real world. On the other hand, simulation uses a model to develop conclusion providing insight on the behavior of the real world elements being studied. In the field of computer simulation, this term enhanced as the uses of computer programming to capture the real world situation. The origin of computer simulation and modeling is in the desire to forecast future behavior due to current phenomena. In the discipline of traffic engineering and transportation planning, computer simulation and modeling is needed because it can study models of traffic and its phenomena for analytical or numerical treatment can be used for experimental studies to describe detail evolution of the system over time, and produce the picture of current reality, as well as future estimation.

On the other words, the increasing trend of traffic congestion in most cities becomes the important issue in transportation system. Since travel demand increases at a rate often greater than the addition of road capacity, the situation will continue to deteriorate unless better traffic management strategies are implemented. To coup this problem, traffic simulation models are becoming as one of the important tool for traffic control. These simulation models is needed to asses, generate scenarios, optimize control, and estimate the future behavior of the system at the operational level. Through simulation the overall picture of traffic system can be pictured as well as the ability to assess current problems and the possible solutions immediately. Simulation and model can be a good tool to show some characteristics of complex traffic system, e.g., stable and unstable states, deterministic, chaotic or even stochastic behavior with phase transitions, fractal dimension and self-organized criticality. However, since the advance of technologies and application of transportation system in urban network and road way were not envisioned when many simulation models were developed, the existing models may not be directly applicable to such of this road system[2].

This paper is a continuation of the research work that have been done before[3]. The results from the previous research work have been evaluated by using real traffic data. In previous works, a simple traffic cellular automaton model has been developed in order to capture the real traffic flow with more naturally. The parameter of spontaneous braking and driver scope awareness was introduced. The spontaneous-braking probability rule captures the natural of braking behavior due to driver characteristic. The simulation results showed that traffic congestion can be effected by spontaneous braking behavior in the urban roadway with density $\rho \leq 0.75$. In the density $\rho > 0.75$, the effect of spontaneous braking in traffic congestion cannot be clearly distinguished. It is because in the high density level $\rho > 0.75$ the congestion already occurred before spontaneous braking parameter applied.

Moreover, in addition to spontaneous braking parameter, the evaluation study of lane changing maneuvers has been done. The driver scope awareness parameter was introduced to reflect the visibility required by the driver to make a perception of a road condition and the speed of vehicle that exist within the certain area of the road before making a lane changing maneuver. Various value of scope awareness has been evaluated regarding their effect to the traffic flow. By taking into consideration the scope awareness parameter, the proposed traffic cellular automata model can reflect certain characteristics of lane changing maneuver in the real traffic situation.

This paper evaluates the simulation model that accommodates the driver behavior rules of spontaneous braking probability and lane changing scope awareness, by

compare their simulation results to the real traffic data. The real traffic data has been observed via recorded traffic video.

This paper is organized as follow. The brief description about spontaneous braking and driver scope awareness in the lane changing maneuvers is quick reviewed in Section 2. Section 3 presents the validation and data gathering method that was used in this research. The comparison results and analysis are described in Section 4. Finally, in section 5, we present a summary and conclusion of this work

II. Spontaneous Braking and Scope of Awareness

Spontaneous braking and scope awareness parameter introduced to reflect the characteristic of driver in real traffic situation. In reality, vehicle would make a braking as the response to avoid collision with another vehicle or avoid some obstacle like potholes, snow, or pedestrian that crosses the road unexpectedly. In many cases, the reckless driving behaviors such as sudden-stop by public-buses, motorcycle which changing lane too quickly, or tailgating make the probability of braking getting increase. Arai et.al.[3], using traffic cellular automaton (TCA) model to describe these characters through a traffic simulation and models. In that simulation, the effect of spontaneous braking and lane changing maneuvers on the traffic flow have been investigated. A set of movement rules include the parameter of spontaneous braking behavior can be described as follow:

- R(1) Acceleration:

$$v_{(i)} \rightarrow min(v_{(i)} + 1, v_{max}) \quad (1)$$

- R(2) Deceleration

$$v_{(i)} \rightarrow min(v_{(i)}, gap_{(i)}) \quad (2)$$

- R(3) Stochastic randomization

$$p : v_{(i)} \rightarrow v_{(i)} - 1 \quad (3)$$

- R(4) Spontaneous braking probability

$$p_{br} : v_{(i)} \rightarrow v_{(i)} - b \quad (4)$$

- R(5) Driving

$$x_{(i)} \rightarrow x_{(i)} + v_{(i)} \quad (5)$$

Furthermore, in our previous research works, the effect of drivers' visibility and their perception (e.g., to estimate the speed and arrival time of another vehicle) on the lane changing maneuver have been investigated. The term of scope awareness was used to describe the visibility required by the driver to make a perception about road-lane condition and the speed of vehicle that exist in such road-lane. In that simulation study, several different values of scope awareness were examined to capture its effect on the traffic flow. A simple traffic cellular automaton model introduced to accommodate various braking characters of driver. Simulation results showed the ability of this model to capture the important features of lane changing maneuver and revealed the appearance of the short-thin solid line jam and the wide solid line jam in the traffic flow as the consequences of lane changing maneuver. The driver scope awareness rules can be summarized as follow:

$$gap_{same} < v_{current} \quad (6)$$

$$cell_{next} = 0 \quad (7)$$

$$rand() < p_{change} \quad (8)$$

$$gap_{target} > gap_{same} \quad (9)$$

$$v_{vehicle,back} \leq gap_{back} \; ; \; X(vehicle_{back}) \in Sa \quad (10)$$

III. Validation Method

Validation is used to determine the real world system being studied is accurately represented by the simulation model. Referring to ISO standard, the following steps in validation are listed[4]:

- Component testing: checking of software subcomponent (the model);
- Functional validation: checking of model capabilities and inherent assumptions;
- Qualitative verification: comparison of predicted traffic behavior with informed expectations.
- Quantitative verification: comparison of model predictions with reliable experimental data.

The first two of these items are usually based on simple test cases and do not require empirical data. The third is based on comparison with observation, and the last on comparison with quantitative and experiment data.

Often, the test has to be done in several times to obtain the best result of validity. By a thorough analysis of the simulation's output data then the best result would be taken. If the model's output data closely represents the expected values for the system's real-world data, then validity is more likely.

When a model has been developed for an existing system, a validity test becomes a statistical comparison. Data collected from the situation of real system can be used as theoretical comparator[5]. However, when the system does not yet exist, validity becomes harder to prove. In many cases, validity cannot be definitely proven until some point in the future when the system being modeled has been deployed and running.

A. Actual Data Gathering

Empirical data is used in validation of simulation results. The data used in this validation is based on video analysis. The usual approach towards data recording is observation and counting. The analysis is done manually, i.e., there was no automatic device that extracted the information from video. The evaluation of the data presented below is based on the following assumptions and methods. Since we interest to evaluate the effect of spontaneous braking behavior in the

traffic flow then their number through video recording have been counted.

A field observation using micro scenario where the actual traffic condition were captured by a 30 minutes video recorded. The parameter that was counted is traffic flow and the number of spontaneous braking. The actual data were taken from two-lane urban roadway with a length approximately 500m. The video camera was placed on the pedestrian bridge. Figureure 1 shows the observation location. Among the location, there are two traffic signal at the end of lane, then to distinguish between the normal braking and the spontaneous braking of vehicle, we use an assumption. For the vehicle that stop due to traffic signal would be categorized as normal braking, other than that would be categorized as spontaneous braking.

Through the video recording, the number of traffic flow and spontaneous braking has been counted for each of the recorded video data. The observation data were taken from 16 different real urban road traffic video data. These data were taken in the morning and afternoon as being assumed as peak traffic time. Table 1 shows the average density values of the observation results. In traffic data analysis, there are three related types of data: speed, flow, and density. Speed v (km/hr) is defined as the distance covered per unit time.

Since the speed of every vehicle is almost impossible to track on the roadway, therefore, in practice, average speed is based on the sampling of vehicles over a period of time or area and is calculated and used in formulae. Some sensing systems can directly measure it. Flow q is the rate in which vehicles arrive at a particular point on a roadway and described in terms of vehicles per hour (cars/hr). Traffic sensing systems usually record the traffic volume, which is the actual number of vehicles to arrive during a sampling period (e.g., 30 seconds). Thus, volume can be converted to a flow rate by multiplying the recorded volume by the number of sampling periods in an hour. Density k is defined as the number of vehicles per unit area of the roadway. The density value is described in terms of vehicle per unit area (cars/km). By measure flow and speed, the density is calculated by dividing the flow rate by the speed.

Since the real traffic data was recorded in 30 minutes video then based on this traffic flow counter, then we estimated the traffic flow for 1 hour. The average speed of vehicles was obtained by field experiment, e.g., driving a car along the observation area then calculated the average speed among such area. Once the flow q and speed values v were obtained then the density value k in the observation area was calculated by using the equation:

$$k = \frac{q\ (cars/hr)}{v\ (km/hr)} \qquad (6)$$

In the traffic cellular automaton model, most of the typical models use a consideration that one cell of the simulation model equal to the 7.5m length of the real system. This value is considered as the length of vehicle plus the distance between vehicles in a stopped position. Referring to this assumption then for 1 km road length there must be 133 vehicles that equal to maximum density among the road lane. Table 1 shows the summarized data. Data from the sequences video that have same density value have been calculated and retrieved their average value.

Fig.1. The location of observation.

TABLE I. THE AVERAGE VALUES OF THE OBSERVATION RESULTS

AVG. SPEED (km/hr)	DENSITY (cars/km)	DENSITY (%/km)	SP. Braking (%)	Real Data FLOW (cars/hr)
50	18.28	7	4	914
50	20.18	8	4	1009
40	24.3	9	4	972
40	25.8	10	5	1032
40	28.75	11	3	1150

IV. COMPARISON AND ANALYSIS

In this observation, we evaluated the number of traffic parameters provided by video recording, i.e., spontaneous braking events, flow number, cars density value, and average speed of the cars. There were several types of vehicle exist in this observation, e.g., motorbike, bus, passenger car, and bicycle. However, in the data analyzing, we just considered for the car types, include truck and bus. After those traffic parameters had been calculated then through the simulation model those real traffic data has been compared. We compared the number of traffic flow resulting from real traffic data and the proposed model. This evaluation used spontaneous braking number and density level from real traffic data. By using those parameters as input value, we obtained the traffic flow results of simulation model.

The comparison result of traffic flow is shown by Figure. 2. Referring to the video observation, in the simulation model we used the probability of lane changing 0.1. In this evaluation, we also compared the traffic flow result by using Nagel-Schreckenberg model[6]. The comparison result shows the model that accommodates spontaneous braking and driver scope awareness produced the better result of traffic flow rather than the original Nagel-Schreckenberg model[6]. The accuracy values between real data flow and our model are presented in Table II, as well as Nagel-Schreckenberg model. It can be seen from the comparison results (Figure. 2), there is a discrepancy between real data and simulation results. The simulation model produced .25 higher cars flow than real traffic flow. Therefore, in this work, we also tried to compare the real traffic data flow to various values of spontaneous braking probability and scope

awareness. Figure 3 presents the scatter graph of comparison between real data flow and several spontaneous braking probability values. Meanwhile, Figure. 4 presents the comparison result between real traffic flow data and several value of driver scope awareness parameter, i.e., scope awareness 3 cells and scope awareness 6 cells, respectively.

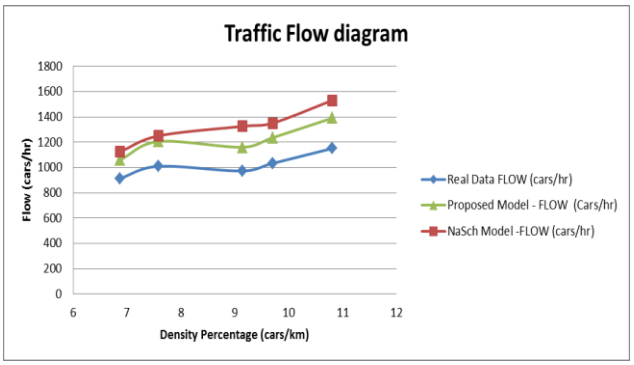


Fig.2. Comparison results of real data vs proposed model vs NaSch model.

TABLE II. TRAFFIC FLOW ACURRATION VALUES

Data Source	Accuration (%)
Real Data	100
NaSch Model	75.9
Proposed Model	83.9

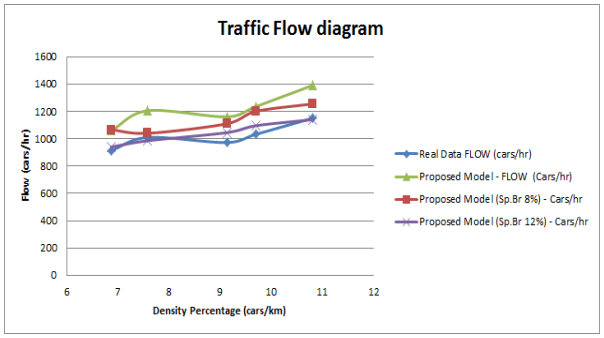


Fig.3. Comparison of real data to several spontaneous braking values

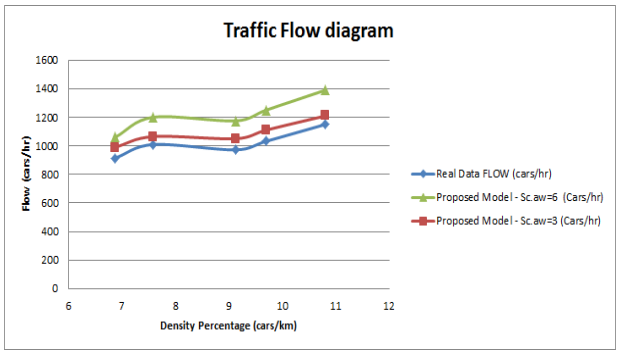


Fig.4. Comparison of real data to several Scope Awareness values

A. *Traffic Flow and Average Speed Estimation*

In addition, this paper also evaluated and made estimation for traffic flow and average speed on every density values. Referring to the Figure 3, it can be seen that by using spontaneous braking probability value 0.12, the simulation can produces a similar flow behavior to the real traffic flow data.

Thus, in order to predict the future traffic behavior, the simulation model used spontaneous braking probability 0.12. Figure 5 and Figure 6 present the traffic flow and speed estimation for all density values, respectively. The horizontal axis represents the percentage of car's density levels on the roadway-length while vertical axis in Figure 6 represents the percentage of the cars that can move at one time step in simulation. In Figure 7, the vertical axis describes the maximum speed that can be reached by the car at the specific density level.

V. CONCLUSION

A validation of spontaneous braking and scope awareness model using measured real traffic flow has been introduced. This validation shows that the traffic cellular automaton model that accommodate the probability of spontaneous braking and scope awareness have given more accurate description about traffic flow situation. Moreover, this paper also presents the estimation behavior of the traffic flow and average speed on every density values.

REFERENCES

[1] J. P. C. Kneijnen, "Theory and Methodology Verification and validation of simulation models," *European Journal of Operational Research*, vol. 82, pp. 145–162, 1995.

[2] S. Boxill and L. Yu, "An Evaluation of Traffic Simulation Models for Supporting ITS," Houston, 2000

[3] K. Arai and S. Sentinuwo, "Spontaneous-braking and lane-changing effect on traffic congestion using cellular automata model applied to the two-lane traffic," *International Journal of Advanced Computer Science and Applications*, vol. 3, no. 8, pp. 39–47, 2012.

[4] H. Klüpfel, "A cellular automaton model for crowd movement and egress simulation," Universitat Duisburg–Essen, 2003.

[5] R. McHaney, *Understanding computer simulation*. RogerMcHaney & Ventus Publishing Aps, 2009.

[6] K. Nagel and M. Schreckenberg, "A cellular automaton model for freeway traffic," *Journal of Physics I France*, vol. 2, no. 12, pp. 2221–2229, 1992.

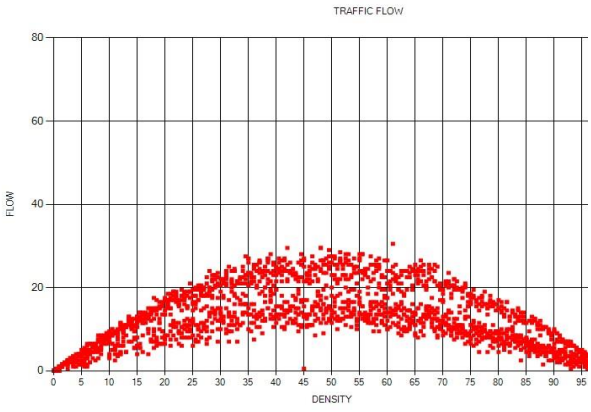


Fig.5. Traffic flow estimation for all density values

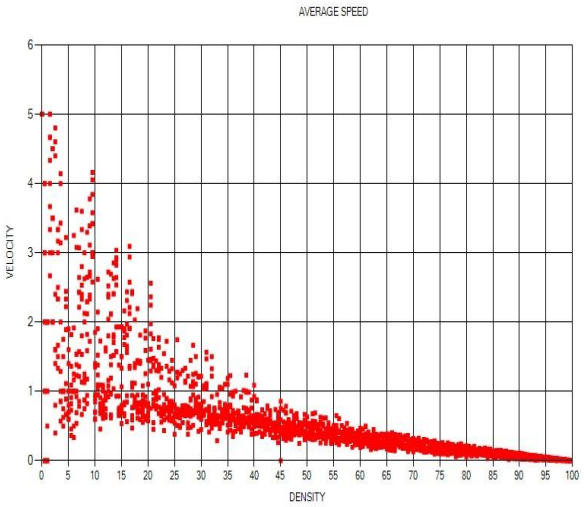


Fig.6. Average speed estimation for all density values

AUTHORS PROFILE

Kohei Arai 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 April 1974 to December 1978 and also was with National Space Development Agency of Japan from January, 1979 to March, 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 moved to Saga University as a Professor in Department of Information Science on April 1990. He was a counselor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was a councilor of Saga University for 2002 and 2003. He also was an executive councilor for the Remote Sensing Society of Japan for 2003 to 2005. He is an Adjunct Professor of University of Arizona, USA since 1998. He also is Vice Chairman of the Commission A of ICSU/COSPAR since 2008. He wrote 30 books and published 322 journal papers

Steven Ray Sentinuwo, received the B.Eng. degree in Electrical Engineering from Sam Ratulangi University, and the M.Eng. degree in Information Engineering, from University of Indonesia, in 2004 and 2006, respectively. He is currently a PhD Student at Information Science in Saga University, Japan. His research interest includes robot path planning, traffic simulation and modeling, and information system management.

Recovering Method of Missing Data Based on Proposed Modified Kalman Filter When Time Series of Mean Data is Known

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

***Abstract*—Recovering method of missing data based on the proposed modified Kalman filter for the case that the time series of mean data is know is proposed. There are some cases of which although a portion of data is missing, mean value of the time series of data is known. For instance, although coarse resolution of imagery data are acquired every day, fine resolution of imagery data are missing sometimes. In other words, coarse resolution of imaging sensor has wide swath width while fine resolution of imaging sensor has narrow swath, in general. Therefore, coarse resolution of sensor data can be acquired every day while fine resolution of sensor data can be acquired not so frequently. It would be nice to become able to create frequently acquired fine resolution of sensor data (every day) using the previously acquired fine resolution of sensor data together with the coarse resolution of sensor data. The proposed method allows creation of fine resolution sensor data with the aforementioned method based on a modified Kalman filter. As an example of the proposed method, prediction of missing ASTER/VNIR data based on Kalman filter using simultaneously acquired MODIS data as a mean value of time series data in revision of filter status is attempted together with a comparative study of prediction errors for both conventional Kalman filter and the proposed modified Kalman filter which utilizes mean value of time series data derived from the other sources. Experimental data shows that 4 to 111% of prediction error reduction can be achieved by the proposed modified Kalman filter in comparison to the conventional Kalman filter. It is found that the reduction rate depends on the mean value accuracy of time series data derived from the other data sources. The experimental results with remote sensing satellite imagery data show a validity of the proposed method**

Keywords*—Kalman filter; nremote sensing satellite image; time series analysis

I. Introduction

There are some cases of which although a portion of data is missing, mean value of the time series of data in concern is known. For instance, coarse resolution of imaging sensor has wide swath width while fine resolution of imaging sensor has narrow swath, in general. Therefore, coarse resolution of sensor data can be acquired every day while fine resolution of sensor data can be acquired not so frequently. It would be nice to become able to create frequently acquired fine resolution of sensor data (every day) using the previously acquired fine resolution of sensor data together with the coarse resolution of sensor data. The proposed method allows creation of fine resolution sensor data with the aforementioned method based on a modified Kalman filter.

Kalman filter is widely used for prediction of missing data [1],[2]. The parameters required for Kalman filter can be basically estimated with the Recursive Least Squares: RLS method. Applicability of the RLS method is discussed [3] and the method for overcoming the problem on the limitation of applicability of the RLS method is also discussed [4[-[6].

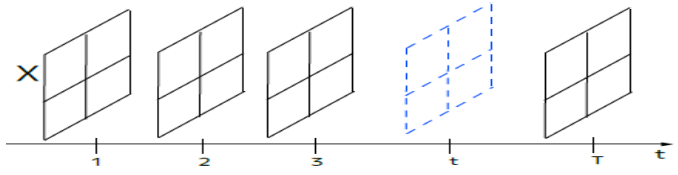
The experimental results with remote sensing satellite imagery data show a validity of the proposed method. As an example of the proposed method, prediction of missing ASTER/VNIR data [7] based on Kalman filter using simultaneously acquired MODIS data [8] as a mean value of time series data in revision of filter status is attempted together with a comparative study of prediction errors for both conventional Kalman filter and the proposed modified Kalman filter which utilizes mean value of time series data derived from the other sources.

II. Proposed Method

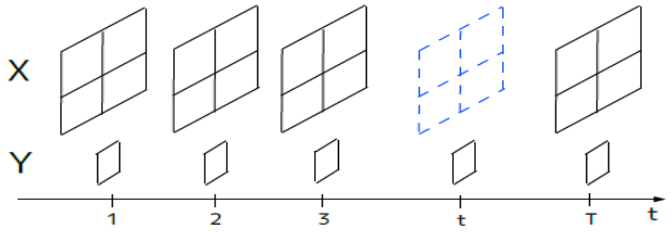
A. Kalman Filter

B. Problem Statement

Prediction error of the conventional Kalman filter is not so small in particular for the cases that there are a long time periods of contiguous missing data. There are some cases that there is time series of mean data of the original time series of data which includes missing data as shown in Figure 1. Not only time series of data X, but also its mean of time series data is also available to use for recovering the missing data. It may be possible to improve prediction accuracy by using mean of time series of data in addition to the available data of the original time series of data.



(a)Typical case



(b)Situation in concern

Fig. 1 Prediction of missing data in time series of data, ***Xt*** with known time series of data, ***Xi*** in the conventional Kalman filter and in the proposed modified Kalman filter.

C. *Proposed Modefied Kalman Filter*

Based on the well known Kalman filter, the missing data can be predicted in accordance with the flow chart of Figure 2. Meanwhile, the flow chart of the proposed modified Kalman filter is shown in Figure 3. Prediction is done by using available data during updating process for estimation of coefficients of the Kalman filter while no update is done when the data is missing for the conventional Kalman filter. Meanwhile, update is done by using mean of the data in concern even for the missing data.

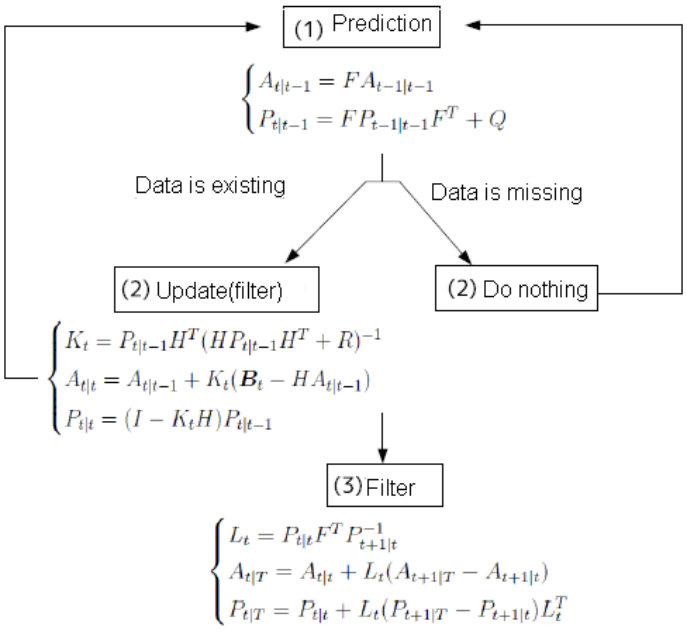


Fig. 2 Prediction of missing data based on the conventional Kalman filter

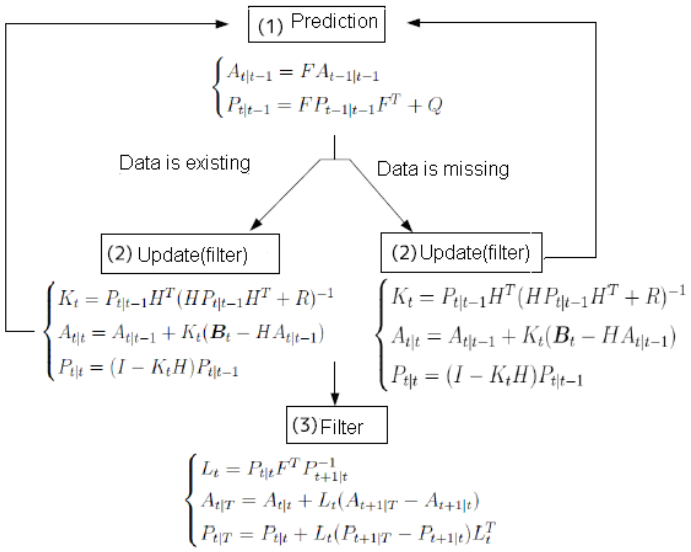


Fig. 3 Process flow of prediction of missing data based on the proposed modified Kalman filter

Where the time series of data is followed by the multi variant auto-regressive model [9] as shown in equation (1)

$$\boldsymbol{X}_t = \sum_{j=1}^{P} \Phi_j \boldsymbol{X}_{t-j} + \boldsymbol{N}_t \quad \{\boldsymbol{N}_t\} \sim WN(0, \Sigma) \tag{1}$$

where $\boldsymbol{\Phi}$ can be estimated with RLS method while $\boldsymbol{\Phi_1}$, … ,$\boldsymbol{\Phi_p}$ is regressive coefficients for $\boldsymbol{X_n}$. On the other hand, *p* denotes is order while ***N*** denotes residual and is followed by multi variant normal distribution with zero mean and covariance matrix of Σ [10]. These can be rewritten with equation (2).

$$\boldsymbol{X}_t = \begin{bmatrix} x_t(1) \\ \vdots \\ x_t(m) \end{bmatrix}, \Phi_j = \begin{bmatrix} \phi_j(11) & \cdots & \phi_j(1m) \\ \vdots & \vdots & \vdots \\ \phi_j(m1) & \cdots & \phi_j(mm) \end{bmatrix}, \boldsymbol{N}_t = \begin{bmatrix} n_t(1) \\ \vdots \\ n_t(m) \end{bmatrix} \tag{2}$$

Multi variant auto-regressive model of the parameters are estimated as follows,

1) *Covariance matrix is estimated at first,*
2) *The following Darvin-Levinsone algorithm [11] is then used introducing the following conditional mean and covariance,*

$$\begin{cases} A_{t|j} \equiv E(\boldsymbol{A}_t|\boldsymbol{A}_j) \\ P_{t|j} \equiv E[(\boldsymbol{A}_t - A_{t|j})(\boldsymbol{A}_t - A_{t|j})^T] \end{cases} \quad (3)$$

At is estimated in accordance with equation (4).

$$\text{Prediction} \begin{cases} A_{t|t-1} = F_t A_{t-1|t-1} \\ P_{t|t-1} = F_t P_{t-1|t-1} F_t^T + Q_t \end{cases}$$
$$\text{Update (Filter)} \begin{cases} K_t = P_{t|t-1} H_t^T (H_t P_{t|t-1} H_t^T + R_t)^{-1} \\ A_{t|t} = A_{t|t-1} + K_t(\boldsymbol{B}_t - H_t A_{t|t-1}) \\ P_{t|t} = (I - K_t H_t) P_{t|t-1} \end{cases} \quad (4)$$

Then the following smoothing filter is applied,

$$\begin{cases} L_t = P_{t|t} F_{t+1}^T P_{t+1|t}^{-1} \\ A_{t|T} = A_{t|t} + L_t(A_{t+1|T} - A_{t+1|t}) \\ P_{t|T} = P_{t|t} + L_t(P_{t+1|T} - P_{t+1|t}) L_t^T \end{cases} \quad (5)$$

State equation of $\boldsymbol{X_n}$ can be written as follows,

$$\begin{aligned} \boldsymbol{A}_{t+1} &= F_t \boldsymbol{A}_t + G_t \boldsymbol{V}_t, \quad \{\boldsymbol{V}_t\} \sim WN(0, \{Q_t\}) \\ \boldsymbol{B}_t &= H_t \boldsymbol{A}_t + \boldsymbol{W}_t, \quad \{\boldsymbol{W}_t\} \sim WN(0, \{R_t\}) \end{aligned} \quad (6)$$

The former equation is called as state equation while the later equations is called as observation equation and are rewritten as follows,

$$\boldsymbol{A}_t = \begin{bmatrix} \boldsymbol{X}_t \\ \boldsymbol{X}_{t-1} \\ \vdots \\ \boldsymbol{X}_{t-p+1} \end{bmatrix}, \boldsymbol{B}_t = \boldsymbol{X}_t \quad (17)$$

$$\begin{bmatrix} \boldsymbol{X}_{t+1} \\ \boldsymbol{X}_t \\ \vdots \\ \boldsymbol{X}_{t-p+3} \\ \boldsymbol{X}_{t-p+2} \end{bmatrix} = \begin{bmatrix} \Phi_p & \Phi_{p-1} & \cdots & \Phi_2 & \Phi_1 \\ \boldsymbol{I} & 0 & \cdots & 0 & 0 \\ 0 & \boldsymbol{I} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{I} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_t \\ \boldsymbol{X}_{t-1} \\ \vdots \\ \boldsymbol{X}_{t-p+2} \\ \boldsymbol{X}_{t-p+1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{I} \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} V_t \quad (7)$$

$$\boldsymbol{B}_t = \begin{bmatrix} \boldsymbol{I} & 0 & \cdots & 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_t \\ \boldsymbol{X}_{t-1} \\ \vdots \\ \boldsymbol{X}_{t-p+2} \\ \boldsymbol{X}_{t-p+1} \end{bmatrix} \quad (8)$$

Thus equation (1) is represented with the above equation (7) and (8) which is called as state space model. State space model can be updated when the data is available while it cannot be updated when the data is missing.

Prediction of the missing data can be done in accordance with Figure 2 based on the conventional Kalman filter. Meanwhile, $\boldsymbol{B_t}$ can be estimated based on the proposed method assuming $\boldsymbol{B_t}$=$\boldsymbol{Y_t}$ と with equation (9),

$$\boldsymbol{A}_t = \begin{bmatrix} \boldsymbol{X}_t \\ \boldsymbol{X}_{t-1} \\ \vdots \\ \boldsymbol{X}_{t-p+1} \end{bmatrix}, \boldsymbol{B}_t = \boldsymbol{Y}_t \quad (9)$$

then

$$\begin{bmatrix} \boldsymbol{X}_{t+1} \\ \boldsymbol{X}_t \\ \vdots \\ \boldsymbol{X}_{t-p+3} \\ \boldsymbol{X}_{t-p+2} \end{bmatrix} = \begin{bmatrix} \Phi_p & \Phi_{p-1} & \cdots & \Phi_2 & \Phi_1 \\ \boldsymbol{I} & 0 & \cdots & 0 & 0 \\ 0 & \boldsymbol{I} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{I} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_t \\ \boldsymbol{X}_{t-1} \\ \vdots \\ \boldsymbol{X}_{t-p+2} \\ \boldsymbol{X}_{t-p+1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{I} \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} V_t \quad (10)$$

$$\boldsymbol{B}_t = \begin{bmatrix} scalar & 0 & \cdots & 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_t \\ \boldsymbol{X}_{t-1} \\ \vdots \\ \boldsymbol{X}_{t-p+2} \\ \boldsymbol{X}_{t-p+1} \end{bmatrix} + W_t \quad (11)$$

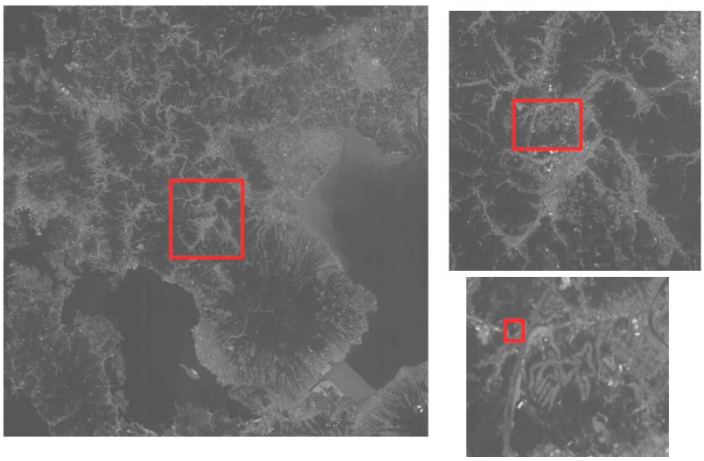
Thus state space model can be updated even for the data is missing. Namely, updating is skipped for the conventional Kalman filter while updating is performed even for the data is missing for the proposed method. Smoothing filter is common for both the conventional and the proposed Kalman filter [12]. In other word, time series of mean data is used as time series of $\boldsymbol{B}$, mean of $\boldsymbol{A}$ is used as coefficients of $\boldsymbol{H_t}$ of the observation equation. The *scalar* in the equation (11) is averaged vector.

III. Experiment

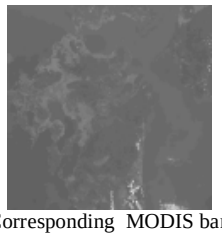
A. Preliminary Simulation

As an example of the proposed method, prediction of missing ASTER/VNIR data based on Kalman filter using simultaneously acquired MODIS data as a mean value of time series data in revision of filter status is attempted together with a comparative study of prediction errors for both conventional Kalman filter and the proposed modified Kalman filter which utilizes mean value of time series data derived from the other sources.

An example of the ASTER/VNIR band 2 image used (Red square on the right bottom image shows 16x16 pixels of ASTER/VNIR band 2 which corresponds to 1 pixel of MODIS band 1) is shown in Figure 4.



(a) ASTER/VNIR band 2



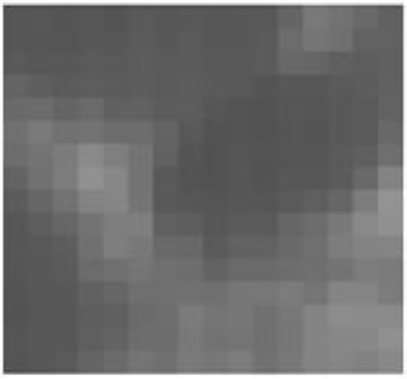
(b)Corresponding MODIS band 1

Fig. 4 An example of the ASTER/VNIR band 2 image used (Red square on the right bottom image shows 16x16 pixels of ASTER/VNIR band 2 which corresponds to 1 pixel of MODIS band 1).

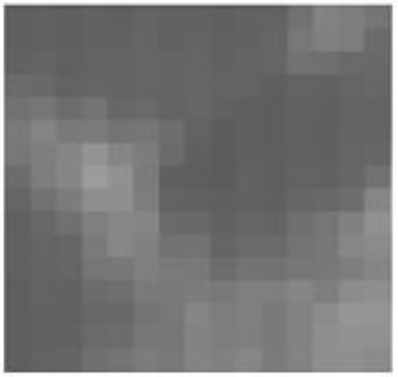
Although these ASTER and MODIS sensors are onboard same satellite, Terra, repeat cycle is different each other. Revisit cycle of the ASTER/VNIR is 16 days while that of the MODIS is almost one day due to the fact that swath width of MODIS is wide enough (2000 km) to achieve one day coverage for entire globe. There are missing data when both sensors acquire cloudy and rainy scenes because these are visible to thermal infrared radiometers. Spectral characteristics of ASTER/VNIR band 2 and MODIS band 1 are almost similar. Instantaneous Field of View: IFOV of ASTER/VNIR is 15 m while that of MODIS is 1 km. Therefore, approximately 16 by 16 pixels of ASTER/VNIR are corresponding to one MODIS pixel. Relatively clear scenes of Ureshino city in Kyushu island in Saga prefecture, Japan are acquired with ASTER/VNIR on December 15 2004, September 29 2005, April 12 2007, May 14 2007, August 18 2007, January 9 2008, April 14 2008, May 3 2008, and May 19 2008 during the period from December 2004 and May 2008. Although ASTER/VNIR acquires imagery data every 16 days in the period, just 10 clear scenes are acquired. By using 10 of the acquired ASTER/VNIR band 2 of imagery data, 8 of the imagery data can be predicted based on Kalman Filter: KF without the corresponding MODIS band 1 data. Also 8 of the imagery data can be predicted based on the proposed modified KF with the corresponding MODIS band 1 data. Figure 5 shows example of small portion of the predicted VNIR band 2 images based on the conventional Kalman filter (KF) and the proposed Kalman filter with MODIS band 1 as a mean value.

Observation equation in equation (6) requires appropriate white noise with zero mean and appropriate standard deviation. In order to determine appropriate standard deviation, prediction error is evaluated with the different standard deviations then an appropriate standard deviation is determined at which prediction error is minimum value.

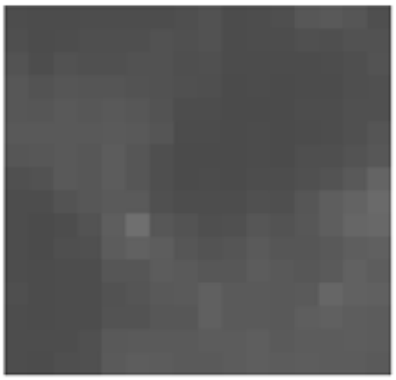
(a) Missing data of 2007/4/12



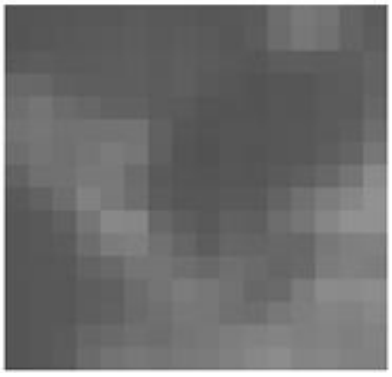
(b) Conventional KF



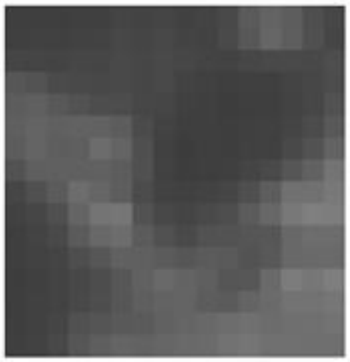
(c) Proposed KF



(d) Missing data of 2008/1/9



(e) Conventional KF



(f) Proposed KF

Fig. 5 Examples of the predicted VNIR band 2 images based on the conventional Kalman filter (KF) and the proposed Kalman filter with MODIS band 1 as a mean value

Table 1 shows the difference between averaged pixel value of VNIR band 2 and the corresponding pixel value of MODIS band 1. It is found that the difference between both ranges from -7 to +11%.

TABLE I. Difference between averaged pixel value of VNIR band 2 and the corresponding pixel value of MODIS band 1

Observation Date	VNIR B2	MODIS B1	% Diff.
2004/12/15	65.4	61	6.728
2005/9/29	53.3	50	6.191
2007/4/12	69.1	64	7.381
2007/4/28	65.1	62	4.762
2007/5/14	67.7	60	11.374
2007/8/18	55.9	60	-7.335
2008/1/9	39.9	38	4.762
2008/4/14	64.3	60	6.687
2008/5/3	71.3	64	10.238
2008/5/19	70.9	66	6.911
		Average	5.7699
		St.Dev.	5.0659

Figure 6 shows prediction error as a function of standard deviation for the data of May 14 2007 and August 18 2007. It is found that the most appropriate standard deviation depends on the time series data in concern.

Noise dependency of the proposed modified KF is evaluated with additive noise (zero mean and standard deviation of 5, 10, and 15 % of the mean value). Figure 7 shows influence due to the difference between 16 by 16 pixel average of VNIR band 2 data and the corresponding 1 pixel of MODIS band 1 data on prediction error.

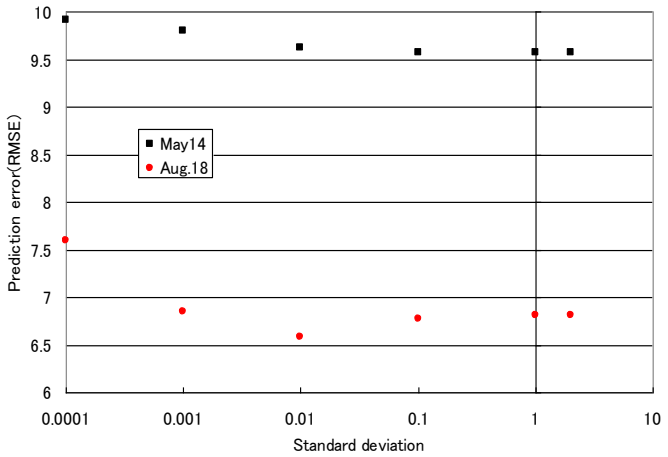


Fig. 6 Experimentally determined standard deviation of R minimizing prediction error (RMSE) for the data of May 14 2007 and August 18 2007.

Table 2 shows prediction errors for the conventional Kalman filter and the proposed modified Kalman filter. Experimental data shows that 4 to 111% of prediction error reduction can be achieved by the proposed modified Kalman filter in comparison to the conventional Kalman filter.

It is found that the reduction rate depends on the mean value accuracy of time series data derived from the other data sources. The experimental results with remote sensing satellite imagery data show a validity of the proposed method

TABLE II. Prediction errors for the conventional Kalman filter and the proposed modified Kalman filter

Missing data	Kalman filter(KF)	Modified KF	%Difference
2005/9/29	21.587	16.67	29.496
2007/4/12	16.78	13.428	24.963
2007/4/28	7.851	7.16	9.651
2007/5/14	9.339	6.289	48.498
2007/8/18	6.443	6.174	4.357
2008/1/9	22.368	10.584	111.338
2008/4/14	14.198	11.509	23.364
2008/5/3	7.901	6.802	16.157
Average	13.308	9.827	33.478
St.Dev.	6.385	3.879	34.214

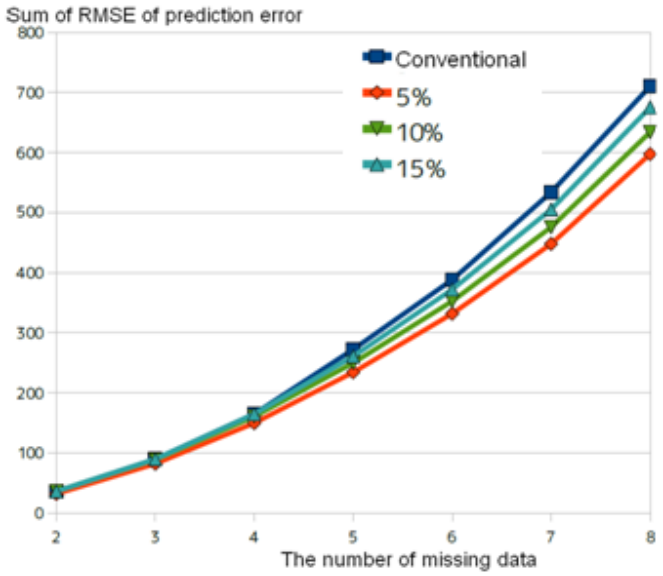


Fig. 7 Influence due to the difference between 16 by 16 pixel average of VNIR band 2 data and the corresponding 1 pixel of MODIS band 1 data on prediction error.

IV. Conclusion

Recovering method of missing data based on the proposed modified Kalman filter for the case that the time series of mean data is know is proposed. The proposed method allows creation of fine resolution sensor data with the aforementioned method based on a modified Kalman filter. As an example of the proposed method, prediction of missing ASTER/VNIR data based on Kalman filter using simultaneously acquired MODIS data as a mean value of time series data in revision of filter status is attempted together with a comparative study of prediction errors for both conventional Kalman filter and the proposed modified Kalman filter which utilizes mean value of time series data derived from the other sources. Experimental data shows that 4 to 111% of prediction error reduction can be achieved by the proposed modified Kalman filter in comparison to the conventional Kalman filter. It is found that the reduction rate depends on the mean value accuracy of time series data derived from the other data sources. The experimental results with remote sensing satellite imagery data show a validity of the proposed method

Acknowledgment

The author would like to thank Mr. Tetsuo Yamaguchi for his effort to conduct the experiments.

References

[1] Stratonovich, R.L., *Application of the Markov processes theory to optimal filtering*. Radio Engineering and Electronic Physics, 5:11, pp.1–19、1960.

[2] K.Seto and K.Arai, Prediction method for time series data analysis with relatively large missing data based on RLS method, Journal of Japan Society for Photogrammetry and Remote Sensing, 38, 5, 20-27, 1999.

[3] K.Seto and K.Arai, Applicability of RLS method for parameter estimation of Kalman filter, Journal of Japan Society for Photogrammetry and Remote Sensing, 39, 1, 48-54, 2000.

[4] K.Seto and K.Arai, Prediction method for time series of data with comparatively large missing data, Journal of Remote Sensing Society of Japan, 20, 1, 43-54, 2000

[5] K.Seto and K.Arai, Effectiveness of the proposed prediction method for nonlinear and nonstationary time series analysis method with dynamic charcteristic extraction, Journal of Japan Society for Photogrammetry and Remote Sensing, 39, 4, 4-12, 2000.

[6] .Arai K. (1996), Fundamental theory for image processing, Gakujutsu-Tosho Shuppan Publishing Co., Ltd.

[7] P.Slater, K.Thome, A.Ono, F.Sakuma, K. Arai, F.Palluconi, H.Fujisada, Y.Yamaguchi and H.Kieffer, *Radiometric Calibration of ASTER Data*, Journal of Remote Sensing Society of Japan, Vol.15, No.2, pp.16-23, Jun.1994.

[8] Barnes, W. L., and V. V. Salomonson, *MODIS: A Global Image Spectroradiometer for the Earth Observing System*, Crit. Rev. Opt. Sci. Technol., CR47, 285-307, 1993.

[9] George Box and Gwilym M. Jenkins. *Time Series Analysis: Forecasting and Control*, second edition. Oakland, CA: Holden-Day, 1976.

[10] Hayes, Monson H., Section 9.4: Recursive Least Squares, *Statistical Digital Signal Processing and Modeling*. Wiley. p.541, 1996.

[11] N.G. van Kampen, *Stochastic processes in physics and chemistry*. New York: North-Holland, 1981.

[12] Japan Medical Electronics Society of Japan Edt, Foundametals of Biological Signal Processing, ISBN : 978-4-339-07132-0 、 Corona Publishing Co. Ltd.,2007.

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.

3D Skeleton model derived from Kinect Depth Sensor Camera and its application to walking style quality evaluations

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

Rosa Andrie Asmara[1,2]
[2)] Informatics Management Department
State Polytechnics of Malang
Malang, Indonesia

***Abstract*—Feature extraction for gait recognition has been created widely. The ancestor for this task is divided into two parts, model based and free-model based. Model-based approaches obtain a set of static or dynamic skeleton parameters via modeling or tracking body components such as limbs, legs, arms and thighs. Model-free approaches focus on shapes of silhouettes or the entire movement of physical bodies. Model-free approaches are insensitive to the quality of silhouettes. Its advantage is a low computational costs comparing to model-based approaches. However, they are usually not robust to viewpoints and scale. Imaging technology also developed quickly this decades. Motion capture (mocap) device integrated with motion sensor has an expensive price and can only be owned by big animation studio. Fortunately now already existed Kinect camera equipped with depth sensor image in the market with very low price compare to any mocap device. Of course the accuracy not as good as the expensive one, but using some preprocessing we can remove the jittery and noisy in the 3D skeleton points. Our proposed method is part of model based feature extraction and we call it 3D Skeleton model. 3D skeleton model for extracting gait itself is a new model style considering all the previous model is using 2D skeleton model. The advantages itself is getting accurate coordinate of 3D point for each skeleton model rather than only 2D point. We use Kinect to get the depth data. We use Ipisoft mocap software to extract 3d skeleton model from Kinect video. From the experimental results shows 86.36% correctly classified instances using SVM.**

***Keywords*—Disable gait classification; 3D Skeleton Model; SVM; Biometrics**

I. INTRODUCTION

In recent years, there has been an increased attention on effectively identifying individuals for prevention of terrorist attacks. Many biometric technologies have emerged for identifying and verifying individuals by analyzing face, fingerprint, palm print, iris, gait or a combination of these traits [1–3].

Human Gait as the classification and recognition object is the famous biometrics system recently. Many researchers had focused this issue to consider for a new recognition system [4–11]. Human Gait classification and recognition giving some advantage compared to other recognition system. Gait classification system does not require observed subject's attention and assistance. It can also capture gait at a far distance without requiring physical information from subjects.

There is a significant difference between human gait and other biometrics classification. In human gait, we should use video data instead of using image data as other biometrics system used widely. In video data, we can utilize spatial data as well as temporal data compare to image data.

There are 2 feature extraction method to be used in gait classification: model based and free model approach [12]. Model-based approaches obtain a set of static or dynamic skeleton parameters via modeling or tracking body components such as limbs, legs, arms and thighs. Gait signatures derived from these model parameters employed for identification and recognition of an individual. It is obvious that model-based approaches are view-invariant and scale-independent. These advantages are significant for practical applications, because it is unlikely that reference sequences and test sequences taken from the same viewpoint. Model-free approaches focus on shapes of silhouettes or the entire movement of physical bodies. Model-free approaches are insensitive to the quality of silhouettes. Its advantage is a low computational costs comparing to model-based approaches. However, they are usually not robust to viewpoints and scale.

Gait therapist have a problem to calculate the quality improvement of the therapy that they did. They could calculate the gait quality using some device in the lab and in practical not too efficient. We propose a method that can measure gait disable quality and classify the result by only capturing the object walking in front of camera.

Imaging technology developed quickly this decades. Motion capture (mocap) device integrated with motion sensor has an expensive price and can only be owned by big animation studio. Fortunately now already existed Kinect camera equipped with depth sensor image in the market with very low price compare to any mocap device. Of course the accuracy not as good as the expensive one, but using some preprocessing we can remove the jittery and noisy in the 3D skeleton points. Our proposed method is part of model based feature extraction and we call it 3D Skeleton model. 3D skeleton model for extracting gait itself is a new model style considering all the previous model is using 2D skeleton model. The advantages itself is getting accurate coordinate of 3D point for each skeleton model rather than only 2D point. We use Kinect to get the depth data. We use Ipisoft mocap software to extract 3d skeleton model from Kinect video. Those 3D skeleton model exported to BVH

animation standard format file and imported to our programming tool which is Matlab. We use Matlab to extract the feature and use a classifier. We create our own gait disable dataset in 3D environment since there are not exist such a dataset before.

II. PROPOSED METHOD

The classification of disable gait quality in this paper consists of three part, preprocessing, feature extraction, and classification. Figure 1 shows the complete overview of proposed human disable gait quality classification.

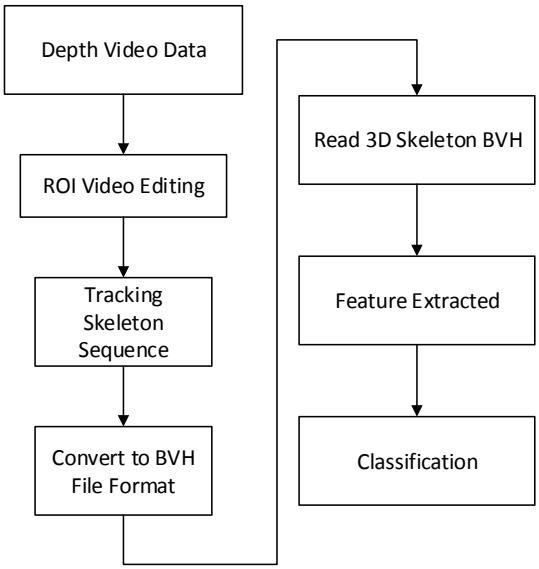


Fig.1. Proposed human disable gait quality classification

A. Preprocessing

First, the Video data using Kinect and IpiRecorder to record the depth data along with RGB video data is captured. To get the video data, there are some recommendation should be considered:

1) using 9 by 5 feet room space, to get best capture.

2) Object should be dressed in casual slim clothing, avoid shiny fabrics.

3) We should ensure that the whole body including arms and legs is visible during the recording states. beginning from T-Pose and the recording can be started.

Second, the depth video data in IPISoft motion capture application is processed. IPISoft will create the 3D skeleton model from video depth recorded using some tracking motion method. The first step is to take only the gait scene, and remove unimportant video scene or we call the Region of Interest (ROI) video. Figure 3 below show the example of video recording.

Third, Create the skeleton 3d model using the tracking motion method, remove the jittery and noises, and export the skeleton model to BVH file format in IPISoft.

Fourth, Read the BVH file, extracted the feature, and classify the feature.

B. Dataset

Unfortunately, there are no Kinect Video Depth gait dataset exists until now. All exist gait dataset is using ordinary camera like USF gait dataset, SOTON gait dataset, and CASIA gait dataset. Figure 2 show the example of CASIA gait dataset.

Fig.2. Example of CASIA gait dataset

To conduct the experiment, we should prepare the dataset. We will use the Kinect Gait Dataset to measure gait quality in disable person. The proposed research will analyze the capability of Kinect and 3D Skeleton model accuracy for gait classification. We also use fake gait patient to be the subject of disable gait person. The subject is analyzing neuropathic patient and classify the gait quality into 5 classes. The first class is for normal gait and calls it class A. The worst quality for neuropathy gait is called class E. The dataset provide 18 Videos each class, thus total data is 90. The dataset will provide knee left angle feature only because there will only left foot simulated.

Figure 3 shows the T-pose position before the video recording start. The top right image showing the RGB video sequence. The color gradient used to represents the depth in video data. Blue color means the object is close to the camera and red color means the object is far from camera.

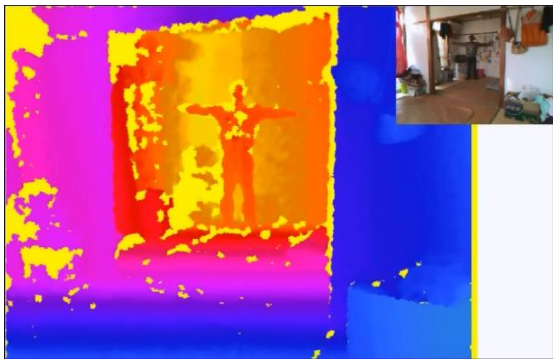


Fig.3. T-Pose Position before the recording begin

Figure 4 shows the 3D skeleton tracking motion sequence. First task is specifying subject's physical parameter like gender and height. IpiSoft will detect the ground plane automatically and provide the 3D skeleton in T-Pose position. Our next job is try to put the T-Pose skeleton in the same position with the subject T-Pose position in the first sequence of video. This time also we should determine the Region of Interest video to be processed. Instead of all the video sequence that we use, we could only take the most important part of the video sequence. Once we put the skeleton to the same position with the subject,

we can refitting pose using the application and start tracking. Jittery removal and Trajectory filtering can be done after the tracking finished. The skeleton sequence result can be import to BVH file standard. Figure 5 below show the BVH file result and preview in BVH file viewer.

Fig.4. Skeleton motion tracking sequence

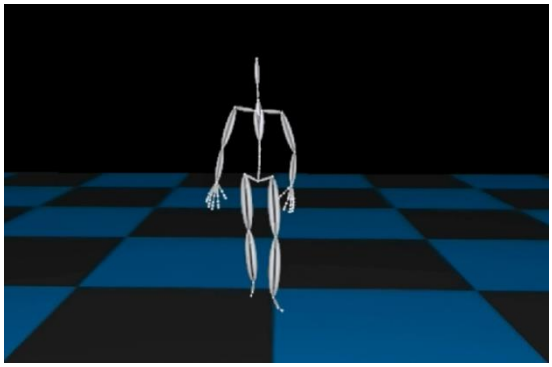


Fig.5. BVH skeleton results

To get the 3D coordinates from skeleton model, we need to get the tree structure and the channel data first. Those data provided by BVH File. Figure 6 show the result of the skeleton model in 3D coordinates. Once we got the coordinates, we can calculate the angle of knee using 3 coordinate of Hip, Left Thigh, and Left Shin from the skeleton.

$$A = P_2 - P_1 = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad [1]$$

$$B = P_3 - P_2 = \sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2} \quad [2]$$

$$C = P_3 - P_1 = \sqrt{(x_3 - x_1)^2 + (y_3 - y_1)^2} \quad [3]$$

Thus the angle of θ is represented as follows:

$$\theta = \cos^{-1}\left(\frac{B^2 - A^2 - C^2}{2AC}\right) \quad [4]$$

C. Feature Extraction

Using the knee angle feature that we created, we can extract the features required for human gait classifications.

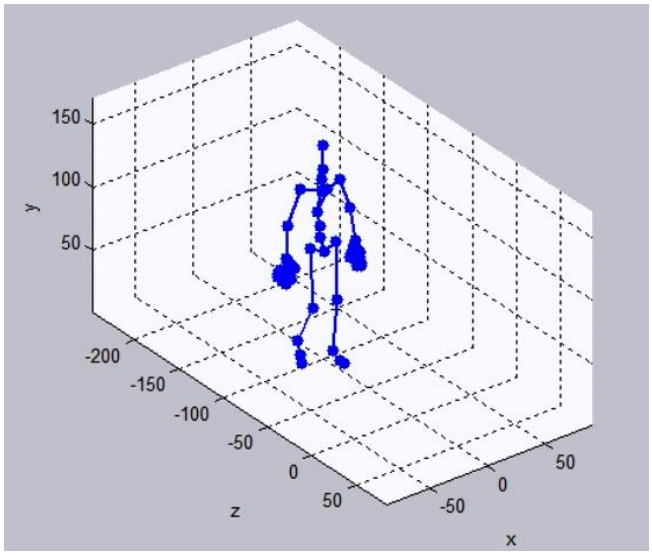


Fig.6. Skeleton model imported in MATLAB

The angles of knee angle can be calculated using simple trigonometry formula as shown in Figure 7 and equation [1] – [4].

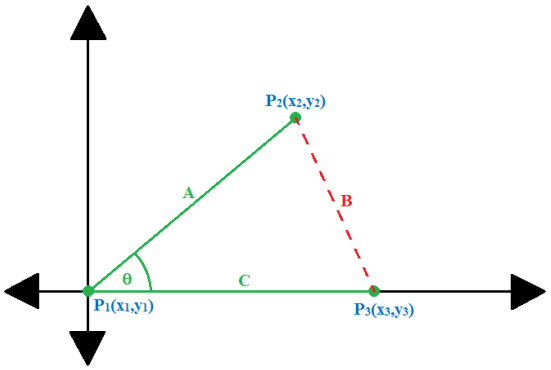


Fig.7. Illustration to calculate angle between two lines using points coordinates

Figure 8 shows examples of the extracted features. In the Figure 8 there are seven peoples' knee angles of features. Their human gaits are evaluated with five grades from A to E. Gait cycles are different each other. Therefore, normalization is required with the longest human gait. Figure 9 shows the normalized human gait derived knee angles of features.

D. Classification

SVM (Support Vector Machine) are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier.

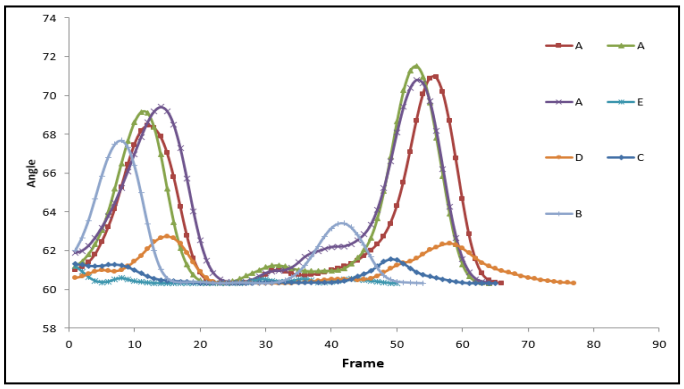


Fig.8. Sample from knee angle feature

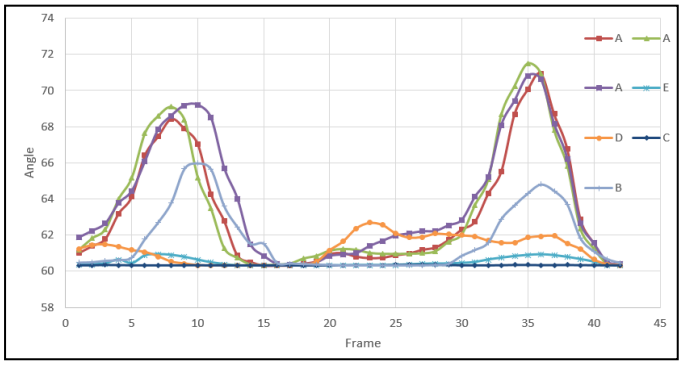


Fig.9. Figure 7 after normalization

Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

III. EXPERIMENTAL RESULT

Using C4.5 pruned tree, we got 77.27% correctly classified instances. Table 1 shows Confusion Matrix from C4.5 Classifier. Also Table 1 shows the detailed accuracy by class using C4.5 Classifier. The decision tree results from C4.5 is as follows,

```
8 <= 62.661977
|  43 <= 61
|  |  39 <= 60.577181
|  |  |  43 <= 51
|  |  |  |  8 <= 60.716972: B (5.0)
|  |  |  |  8 > 60.716972: D (2.0)
|  |  |  43 > 51: D (15.0/1.0)
|  |  39 > 60.577181: B (13.0)
|  43 > 61
|  |  35 <= 60.897472: D (2.0)
|  |  35 > 60.897472: C (17.0)
8 > 62.661977
|  36 <= 68.967621
|  |  32 <= 61.451288: A (3.0)
|  |  32 > 61.451288: E (19.0/1.0)
|  36 > 68.967621: A (12.0)
```

TABLE I. Confusion Matrix from C4.5 Classifier

	A	B	C	D	E
A	11	0	0	0	5
B	0	14	1	3	0
C	0	0	16	1	1
D	0	4	3	11	0
E	2	0	0	0	16

TABLE II. Detailed accuracy by class using C4.5 Classifier

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
A	0.688	0.028	0.846	0.688	0.759	0.858
B	0.778	0.057	0.778	0.778	0.778	0.937
C	0.889	0.057	0.8	0.889	0.842	0.924
D	0.611	0.057	0.733	0.611	0.667	0.833
E	0.889	0.086	0.727	0.889	0.8	0.879
Weighted Avg	0.773	0.058	0.775	0.773	0.769	0.887

Using SVM, we got 86.36% correctly classified instances.

TABLE III. Confusion Matrix from SVM Classifier

	A	B	C	D	E
A	14	0	0	0	2
B	0	15	0	1	2
C	0	0	17	1	0
D	0	3	2	13	0
E	1	0	0	0	17

TABLE IV. Detailed accuracy by class using SVM classifier

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
A	0.875	0.014	0.933	0.875	0.903	0.931
B	0.833	0.043	0.833	0.833	0.833	0.895
C	0.944	0.029	0.895	0.944	0.919	0.958
D	0.722	0.029	0.867	0.722	0.788	0.847
E	0.944	0.057	0.81	0.944	0.872	0.944
Weighted Avg	0.864	0.035	0.866	0.864	0.862	0.914

IV. CONCLUSION

The proposed method uses Kinect depth sensor camera and Ipisoft motion capture software to generate 3D skeleton model. Ipisoft itself is special purpose application to create skeleton so user can use the motion to their computer generated character motion. The skeleton generated will then extract the knee angle feature and use the feature to measure the gait disable quality. The purpose of this research is to analyze whether Kinect and Ipisoft can be used for extracting feature from 3D skeleton in gait related research. Experiments are done using knee angle feature and 18 video dataset in each class. From the experimental results shows 86.36% correctly classified instances using SVM.

ACKNOWLEDGMENT

Portions of the research in this paper use the CASIA Gait Database collected by Institute of Automation, Chinese Academy of Sciences. In his connection, authors would like to thank to Chinese Academy of Sciences for their providing of the Gait database.

REFERENCES

[1] X. Qinhan, "Technology review - Biometrics-Technology, Application, Challenge, and Computational Intelligence Solutions," *IEEE Computational Intelligence Magazine*, vol. 2, pp. 5–25, 2007.

[2] E. Yih, G. Sainarayanan, and A. Chekima, "Palmprint Based Biometric System: A Comparative Study on Discrete Cosine Transform Energy, Wavelet Transform Energy and SobelCode Methods," *Biomedical Soft Computing and Human Sciences*, vol. 14, no. 1, pp. 11–19, 2009.

[3] Z. Yang, M. Li, and H. Ai, "An experimental study on automatic face gender classification," *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 3, pp. 1099 – 1102, 2006.

[4] N. V. Boulgouris, D. Hatzinakos, and K. N. Plataniotis, "Gait recognition: A Challenging Signal Processing Technology For Biometric Identification," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 78–90, 2005.

[5] M. S. Nixon and J. N. Carter, "Automatic Recognition by Gait," *Proceedings of the IEEE*, vol. 94, pp. 2013–2024, 2006.

[6] D. Cunado, M. S. Nixon, and J. N. Carter, "Automatic Extraction and Description of Human Gait Models for Recognition Purposes," *Computer Vision and Image Understanding*, vol. 90, no. 1, pp. 1–41, 2003.

[7] X. Li, S. Maybank, and S. Yan, "Gait components and their application to gender recognition," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 38, no. 2, pp. 145–155, 2008.

[8] K. Arai and R. Asmara, "Human Gait Gender Classification using 2D Discrete Wavelet Transforms Energy," *IJCSNS International Journal of Computer Science and Network Security*, pp. 62–68, 2011.

[9] K. Arai and R. Asmara, "Human Gait Gender Classification in Spatial and Temporal Reasoning," *IJARAI International Journal of Advanced Research in Artificial Intelligence*, vol. 1, no. 6, pp. 1–6, 2012.

[10] G. Huang and Y. Wang, "Gender classification based on fusion of multi-view gait sequences," *Computer Vision–ACCV 2007*, vol. 4843, pp. 462–471, 2007.

[11] L. Lee and W. Grimson, "Gait Analysis for Recognition and Classification," in *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, 2002, pp. 148–155.

[12] J. Wang, M. She, S. Nahavandi, and A. Kouzani, "A Review of Vision-Based Gait Recognition Methods for Human Identification," *2010 International Conference on Digital Image Computing: Techniques and Applications*, pp. 320–327, Dec. 2010.

AUTHIORS PROFILE

Kohei Arai 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 April 1974 to December 1978 and also was with National Space Development Agency of Japan from January, 1979 to March, 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 moved to Saga University as a Professor in Department of Information Science on April 1990. He was a 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 a councilor of Saga University for 2002 and 2003. He also was an executive councilor for the Remote Sensing Society of Japan for 2003 to 2005. He is an Adjunct Professor of University of Arizona, USA since 1998. He also is Vice Chairman of the Commission A of ICSU/COSPAR since 2008. He wrote 29 books and published 290 journal papers.

Rosa A. Asmara received the B.E. degree in electronics engineering from Brawijaya University, and the M.S. degree in Multimedia engineering, from Institute of Technology Sepuluh Nopember, Surabaya, Indonesia, in 2004 and 2009, respectively. He is currently a PhD Student at Information Science in Saga University, Japan. His research interests include signal processing, image processing, parallel processing, pattern recognition, and computer vision.

Contradiction Resolution between Self and Outer Evaluation for Supervised Multi-Layered Neural Networks

Ryotaro Kamimura
IT Education Center and Graduate School of Science and Technology
Tokai University, 1117 Kitakaname, Hiratsuka, Kanagawa, Japan
ryo@keyaki.cc.u-tokai.ac.jp

***Abstract*—In this paper, we propose a new type of information-theoretic method. We suppose that a neuron should be evaluated from different points of view to precisely discern its properties. In this paper, we restrict ourselves to two types of evaluation methods for neurons, namely, self and outer-evaluation. A neuron fires only as a result of evaluating itself, while the neuron can fire as a result of evaluation by all surrounding neurons. Self- and outer-evaluation should be equivalent to each other. When contradiction between two types of evaluation exists, the contradiction should be as small as possible. Contradiction between self- and outer-evaluations is realized in terms of the Kullback-Leibler divergence between two types of neurons. Contradiction between self- and outer-evaluation can be resolved by decreasing the contradiction ratio between the two types of evaluation in terms of KL divergence. This method is expected to extract the main features in input patterns, if those are shared by two types of evaluation. We applied the method to two data sets, namely, the logistic and dollar-yen exchange rate data. In both problems, experimental results showed that visualization performance could be improved, leading to clearer class structure for both problems. In addition, when visualization was improved, generalization performance did not necessarily degrade, showing the possibility of networks with better visualization and prediction performance.**

***Keywords*—contradiction resolution; self- and outer-evaluation; visualization; self-organizing maps**

I. Introduction

We here introduce the necessity of using multiple types of evaluation on a neuron to fully understand its main mechanism. We restrict ourselves to two types of evaluation for neurons, namely, self- and outer-evaluation for actual implementation. Then, the necessity is more concretely explained in terms of explicit class structure in the self-organizing maps and the interpretation of internal representations in neural networks.

A. Contradiction Resolution

In this paper, we suppose that a neuron can be evaluated from multiple points of view. For example, a neuron can be evaluated individually or as a member of a neuron group. The neuron may be seen differently when it is evaluated individually or as a member of a group. If evaluation as an individual neuron is contradictory to evaluation as a member of a neuron group, the neuron should take some action to solve this difficulty. This behavior cannot be fully explained without considering the effects of other neurons, meaning that a neuron cannot be understood from a single type of evaluation alone; we must evaluate a neuron from multiple points of view to fully understand the main mechanism of the neuron.

For the simplification and actual application of this necessity of multiple types of evaluation, we suppose two types of neuron evaluation [1]. One type of evaluation is realized by evaluating a neuron for itself without any consideration on other neurons. This means that a neuron fires to input patterns, taking into account a relation between the neuron and the input patterns. This evaluation is called ”self-evaluation”. On the other hand, the other one is obtained by evaluating a neuron by taking into account other neurons’ relations between the neurons and the input patterns. This means that the neuron’s firing rates are determined only by those of other neurons. This evaluation is called ”outer evaluation”. If the self-evaluation of a neuron is different from the outer-evaluation of the same neuron, the neuron has very specific characteristics which are not shared by other neurons. On the other hand, if the self-evaluation is equivalent to the outer-evaluation, the characteristics inherent to the neuron are shared by the other neurons. Our hypothesis is that the contradiction between the two types of evaluation should be reduced as much as possible. It is desirable that all neurons are in harmony with each other.

B. Explicit Class Structure

The contradiction resolution can be applied to the clarification of class structure in self-organizing maps. Self-organizing maps [2], [3], [4] are a well-known technique in neural networks. In particular, they have been used to visualize complex data on simpler form. However, it has been difficult to fully visualize knowledge obtained by the self-organizing maps. Thus, there have been many different kinds of visualization techniques so far developed for self-organizing maps [5], [6], [7], [8], [9], [10], [11], [12]. [13], [14], [15], [16], [17], [18], [19], [20], [21], [22]. However, we can say that those methods did not necessarily succeed in visualizing the knowledge obtained by conventional self-organizing maps.

We think that one of the main reasons for the difficulty in visualization is due to the characteristics of self-organizing maps. Self-organizing maps are based upon competition and cooperation between neurons. In particular, cooperation plays the most important role. This cooperation is based upon the hypothesis that neighboring neurons behave in the same way as a particular neuron. A winner neuron is selected by the process

of competition, and the neighboring neurons of the winner are updated in the same way, in proportion to distance from the winner neuron. Now, to clarify class structure, it is necessary to make class boundaries as clear as possible. However, self-organizing maps go through a process of making class boundaries unclear. This is because neighboring neurons must be as similar as possible, due to the process of cooperation. The class boundaries are naturally based upon dissimilarity between neurons. This shows the necessity to attenuate cooperation between neurons for explicit class structure.

In our contradiction resolution, cooperation between neurons corresponds to the outer-evaluation, because the results obtained by the outer-evaluation are ones realized only by all the other neurons. On the other hand, self-evaluation is a form of evaluation realized only within a neuron without considering any other neurons. The influence of self-evaluation can be used to attenuate the effect of cooperation between neurons and to clarify dissimilarity between neurons. This dissimilarity is related to the clarification of class boundaries in self-organizing maps.

C. Interpretation of Internal Representation

Our contradiction resolution is used to provide supervised neural networks with more interpretable internal representations. One of the most important things to do in neural networks is to explain why such neural mechanisms can be used to produce outputs from inputs through obtained internal representations [23]. Though there have been many attempts to interpret final representations [24], [25], [26], [27], [28] [29], [30], [31], the problem has remained unsolved. Furthermore, we are as yet unaware of the relationship between interpretable representations and generalization performance.

Our method aims to provide neural networks with interpretable internal representations in two ways, namely, through visualization performance due to self-organizing maps and the characteristics of self- and outer-evaluation. First, because our method is applied to the production of self-organizing maps, the outputs from our method are inherited from interpretable knowledge in the self-organizing maps. As above mentioned, the introduction of self- and outer-evaluation can be used to make class boundaries clearer and to make it easier to interpret final knowledge.

Second, the self- and outer-evaluation aim to enhance the characteristics shared by two types of evaluation. A neuron is evaluated for itself and by all the other ones. When the characteristics obtained by the self-evaluation are equivalent to those obtained by the outer-evaluation, those characteristics should be emphasized or enhanced. On the other hand, if the characteristics obtained by self-evaluation are different from those obtained by outer-evaluation, those characteristics should be weakened. Thus, we can expect that gradually, only important and shared characteristics will become stronger. Our contradiction resolution should finally produce interpretable knowledge by enhancing important characteristics.

D. Outline

In Section 2, we first explain the concept of contradiction between self and outer-evaluation. Then, we present how to compute the firing rates by the self and outer-evaluation. We introduce the Kullback-Leibler divergence between firing rates by self- and outer-evaluation. This Kullback-Leibler divergence is called the ”contradiction ratio” to show how self- and outer-evaluation are different from each other. Finally, we explain how to compute connection weights by reducing the contradiction ratio.

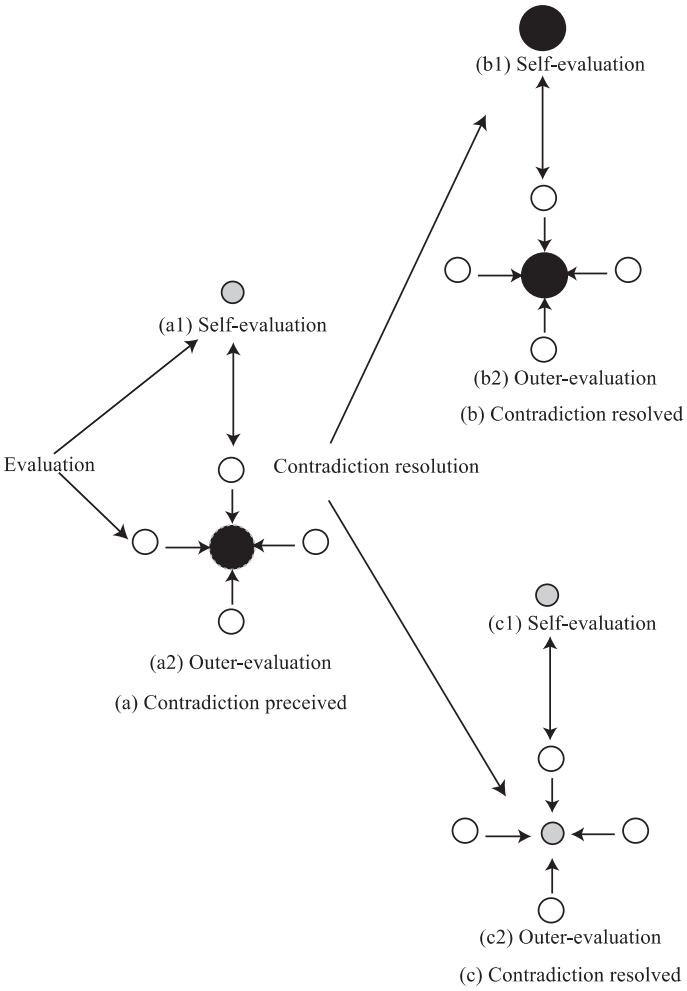


Fig. 1. Concept of contradiction resolution: perception (a) and resolution (b) and (c).

In Section 3, we present two experimental results, namely, the logistic data and dollar-yen exchange rate estimation. In the logistic data, we explain several evaluation measures and how to implement them, showing experimental results. While the results show that the conventional SOM failed to produce explicit class boundaries, the contradiction resolution produced from two to five class boundaries, depending on the parameter. In addition, generalization performance was the same as that by the conventional self-organizing maps and radial-basis function networks. In the dollar-yen exchange rate prediction, class structure became gradually complicated when the parameter was increased. The generalization error gradually decreased and seemed to reach a stable state when the parameter was increased. The generalization errors were sufficiently small compared to those generated by the conventional self-organizing maps and the radial-basis function networks.

II. Theory and Computational Methods

A. Concept of Contradiction Resolution

We suppose that a neuron can be evaluated from many different points of view to fully understand its main mechanism. The characteristics of a neuron can be changed by different types of evaluation. A neuron is defined by the unification of all the characteristics evaluated from different points of view. For simplification, we suppose that there are only two types of evaluation, namely, self- and outer-evaluation. In self-evaluation, a neuron is evaluated by itself and for itself, as represented in Figure 1 (a1). On the other hand, in outer-evaluation, a neuron is evaluated and viewed by all the other neurons as shown in Figure 1 (a2). It is desirable that the results of self-evaluation are equivalent to those of outer-evaluation. However, it happens that the results of self-evaluation are contrary to those by outer-evaluation. As shown in Figure 1 (a), the firing rate obtained via self-evaluation is small while that obtained via outer-evaluation is large. Because this contradiction between self- and outer-evaluation exists, we should make it as small as possible. In Figure 1 (b1), the firing rate obtained by self-evaluation becomes larger and closer to that obtained by outer-evaluation. This is a case where outer-evaluation is more influential. On the other hand, in Figure 1 (c), the effect of self-evaluation becomes apparent. The firing rate obtained by the outer-evaluation in Figure 1 (c2) becomes smaller by the contradiction resolution, because the effect of self-evaluation has some influence on the outer-evaluation.

By using this property of contradiction resolution, the characteristics shared by both self- and outer-evaluation can be enhanced, while those specific to each evaluation can be inhibited. Thus, we can expect that important characteristics are enhanced by the contradiction resolution.

B. Self and Outer Evaluation

Let us explain how to compute outputs from competitive neurons and input patterns in Figure 2. The sth input pattern of total S patterns can be represented by $\mathbf{x}^s = [x_1^s, x_2^s, \cdots, x_L^s]^T, \quad s = 1, 2, \cdots, S$. Connection weights into the jth competitive unit of total M units are computed by $\mathbf{c}_j = [c_{1j}, c_{2j}, \cdots, c_{Lj}]^T, \quad j = 1, 2, \ldots, M$. Now, we can compute the firing rates by self and outer-evaluation. A neuron's firing rates by self-evaluation can be defined by using the outputs from the neuron. Then, the jth competitive neuron output, without considering the other neurons, can be computed by

$$v_j^s = \exp\left(-\frac{\| \mathbf{x}^s - \mathbf{c}_j \|^2}{2\sigma_\beta^2}\right). \tag{1}$$

where $\mathbf{x}^s$ and $\mathbf{w}_j$ are supposed to represent L-dimensional input and weight column vectors, where L denotes the number of input units. The spread parameter σ_β is computed by $1/\beta$, where $\beta > 0$. Thus, the firing rate obtained by the self-evaluation can be computed by

$$p(j \mid s) = \frac{\exp\left(-\frac{\|\mathbf{x}^s - \mathbf{c}_j\|^2}{2\sigma_\beta^2}\right)}{\sum_{m=1}^{M} \exp\left(-\frac{\|\mathbf{x}^s - \mathbf{c}_m\|^2}{2\sigma_\beta^2}\right)}. \tag{2}$$

The firing rate of a neuron evaluated by outer-evaluation is determined by considering all the firing rates of all the other neurons. We approximate the firing rates by summing all firing rates of all neighboring neurons, excluding the target neuron. In addition, the rates are weighted by the distances between those neurons and the target neuron. Then, the outputs by the outer evaluation are defined by

$$y_j^s = \sum_{m=1}^{M} (1 - \delta_{jm}) \phi_{jm} p(m|s), \tag{3}$$

where ϕ_{jm} represent relations between the jth and mth neuron. The output obtained by the evaluation is the sum of all neighboring neurons' firing rates weighted by the relations between them. Then, the firing rate of the outer-evaluation can be defined by

$$q(j \mid s) = \frac{\sum_{m=1}^{M} (1 - \delta_{jm}) \phi_{jm} p(m|s)}{\sum_{r=1}^{M} \sum_{m=1}^{M} (1 - \delta_{rm}) \phi_{rm} p(m|s)}. \tag{4}$$

C. Contradiction Resolution

Contradiction resolution aims to reduce contradiction between self and outer-evaluation [32]. We use the Kullback-Leibler divergence to represent the difference between self- and outer-evaluation. Using the Kullback-Leiber divergence, the contradiction ratio is defined by

$$CR = \sum_{s=1}^{S} p(s) \sum_{j=1}^{M} p(j \mid s) \log \frac{p(j \mid s)}{q(j \mid s)}. \tag{5}$$

When the KL divergence is minimized, we have

$$p^*(j \mid s) = \frac{q(j \mid s) \exp\left(-\frac{\|\mathbf{x}^s - \mathbf{c}_j\|^2}{2\sigma_\beta^2}\right)}{\sum_{m=1}^{M} q(m \mid s) \exp\left(-\frac{\|\mathbf{x}^s - \mathbf{c}_m\|^2}{2\sigma_\beta^2}\right)}. \tag{6}$$

By substituting this optimal firing rates $p^*(j \mid s)$ for $p(j \mid s)$, we have the free energy:

$$\begin{aligned} F &= -2\sigma^2 \sum_{s=1}^{S} p(s) \log \sum_{j=1}^{M} q(j \mid s) \\ &\quad \times \exp\left(-\frac{\| x^s - c_j \|^2}{2\sigma_\beta^2}\right). \end{aligned} \tag{7}$$

This equation can be expanded as

$$\begin{aligned} F &= \sum_{s=1}^{S} p(s) \sum_{j=1}^{M} p^*(j \mid s) \|\mathbf{x}^s - \mathbf{w}_j\|^2 \\ &\quad + 2\sigma_\beta^2 \sum_{s=1}^{S} p(s) \sum_{j=1}^{M} p^*(j \mid s) \log \frac{p(j \mid s)}{q(j \mid s)}. \end{aligned} \tag{8}$$

Thus, the free energy can be used to decrease KL divergence as well as quantization errors. By differentiating the free energy, we can obtain re-estimation formula:

$$\mathbf{w}_j = \frac{\sum_{s=1}^{S} p^*(j \mid s) \mathbf{x}^s}{\sum_{s=1}^{S} p^*(j \mid s)}. \tag{9}$$

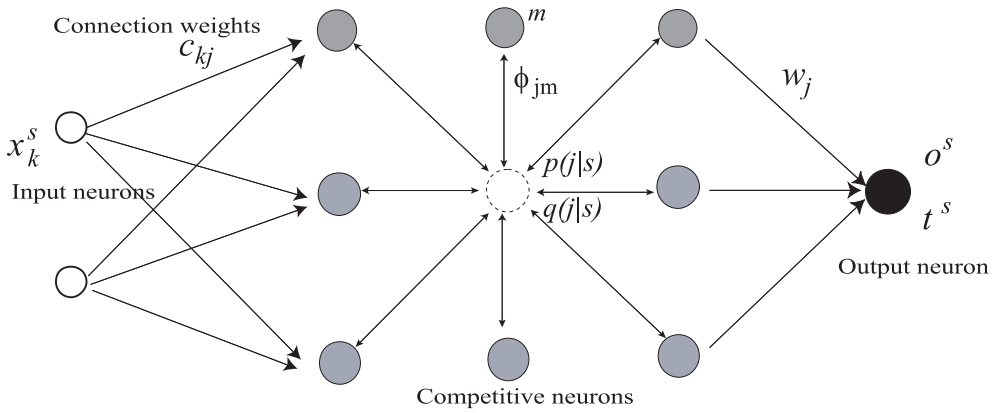


Fig. 2. A neural network for two types of evaluation where some connection weights are eliminated for simplification purposes.

III. Results and Discussion

We here present two experimental results by using the artificial data and the dollar-yen exchange rates. We used the logistic data as the artificial one, where in addition to the presentation of experimental results, we show how to compute several computational measures to evaluate the performance quantitatively and visually. In the dollar-yen exchange rate estimation, we in particular show how visualization performance is related to prediction performance. The network size was set to a much larger one for the intuitive interpretation of our results, namely, 20 by 15 neurons.

A. Logistic Function Identification

1) Experiment Outline: We first used artificial data generated by the logistic function $y = (1 - \exp(-x))/(1 + \exp(-x))$. As shown in Figure 3(a), the training data was generated by the function added to the normal random values with a standard deviation of 0.1. The number of training and testing patterns was 100 and 1,000, respectively. Figure 3(b) shows testing data (in black) and predicted values (in red) by the RBF network with forward selection with the Bayesian information criterion [33], [34], [35]. The input data increased linearly from -10 to 10, meaning that no boundaries inside could be identified. Thus, we tried to show how the contradiction resolution divides this linear data into classes. Then, we examine to what extent this classification is related to the prediction performance.

2) Quantitative Evaluation: We quantitatively evaluated the performance of contradiction resolution in terms of the property of topological maps and prediction performance. For the property of the self-organizing maps, we used the well-known and simple error measures for quantification evaluation, namely, quantization and topographic errors. This is because much importance was placed on the easy reproduction of our results. The quantization error is simply the average distance from each data vector to its BMU (best-matching unit). The topographic error is the percentage of data vectors for which the BMU and the second-BMU are not neighboring units [36]. In addition, to measure how the maps are organized, we computed mutual information (INF)

$$INF = \sum_{s=1}^{S} \sum_{j=1}^{M} p(s)p(j \mid s) \log \frac{p(j \mid s)}{p(j)}. \qquad (10)$$

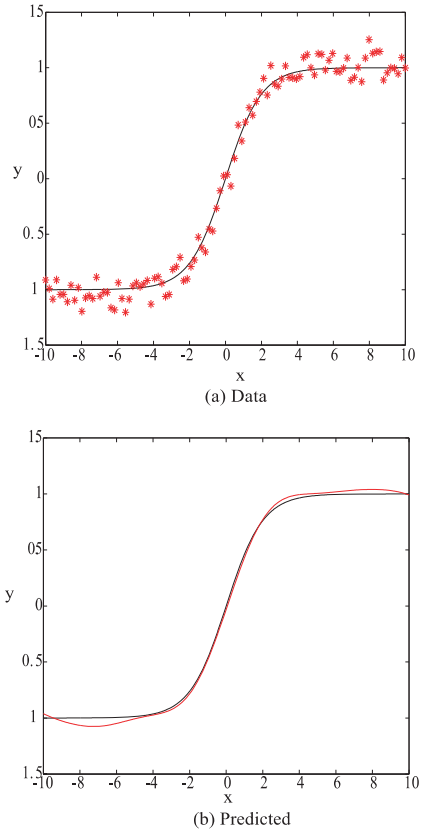


Fig. 3. Training data in red (a) and testing data in black with predicted values in red (b) for the logistic function $y = (1 - \exp(-x))/(1 + \exp(-x))$.

As this mutual information increases, the organization of the maps increases accordingly.

Table I shows the summary of the experimental results. The mean squared errors (MSE) between the targets and outputs on the output layer decreased from 0.11346 (β=1) to 0.00126 (β=5). Then, the mean squared errors (MSE) slightly increased and reached a stable value. The quantization errors (QE) decreased from 5.050 (β=1) to 0.073 (β=20). The topographic errors (TE) decreased from one (β=1) to 0.2 (β=30). Mutual information increased from 0 (β=1) to 1.654 (β=30). The correlation coefficients between mutual information and MSE, QE and TE were -0.962, -0.999 and -0.419, respectively.

TABLE I. MSE FOR THE TESTING DATA, QUANTIZATION ERRORS (QE), TOPOGRAPHIC ERRORS (TE) AND MUTUAL INFORMATION WHEN THE PARAMETER β WAS CHANGED FROM 1 TO 19 FOR THE 20 BY 15 MAP. THE SYMBOL CC IN THE FINAL ROW REPRESENTS THE CORRELATION COEFFICIENTS BETWEEN INFORMATION (INF) AND THE OTHER MEASURES.

β	MSE	QE	TE	INF
1	0.11346	5.050	1.000	0.000
3	0.00159	1.580	1.000	1.146
5	0.00126	0.633	1.000	1.428
7	0.00127	0.317	1.000	1.533
10	0.00128	0.151	1.000	1.605
15	0.00126	0.094	0.720	1.638
20	0.00126	0.073	0.660	1.649
25	0.00126	0.075	0.420	1.653
30	0.00126	0.080	0.200	1.654
CC	-0.962	-0.999	-0.419	

Thus, the quantization errors and topographic errors could be decreased when mutual information was increased.

3) Visual Evaluation: We evaluated visualization performance in terms of a U-matrix and contradiction ratios. For visualization, we computed the contradiction ratio for each neuron by the average of the contradiction ratios over all input patterns

$$R_j = \sum_{s=1}^{S} p(s) \left| \log \frac{p(j \mid s)}{q(j \mid s)} \right|. \qquad (11)$$

We used the absolute values of the contradiction ratios only for better visualization, namely, to show a tendency similar to that of the U-matrix.

Figure 4 shows U-matrices (a) and contradiction ratios (b). When the parameter β was increased from 10 to 20 in Figure 4 (a1)-(a3), two class boundaries were gradually unfolded. When the parameter β was 25 in Figure 4 (a4), the number of classes increased to three. Finally, when the parameter was increased to 30 in Figure 4 (a5), the number of class boundaries increased to four. In addition, the contradiction ratios gradually detected five classes corresponding to those detected by the U-matrix in Figure 4 (b1)-(b5).

4) Interpretation: Figure 5 shows labels (1) and data with boundaries (2) when the parameter β was changed from 20 (a) to 30 (c). When the parameter β was 20, two class boundaries were detected and three classes were identified. As shown in Figure 5 (a), the logistic function was separated into flat higher and lower areas, and the intermediate areas between them. When the parameter β was increased to 25 in Figure 5 (b), we could see three classes boundaries dividing four classes, see Figure 5 (b1). As shown in Figure 5 (b2), in addition to two class boundaries in Figure 5 (b1), a class boundary in the middle of the U-matrix could be identified. Finally, when the parameter β was 30 in Figure 5(c), four class boundaries with five classes were identified, see Figure 5(c1). The slope between the lower and higher areas were further subdivided into three classes in Figure 5 (c2).

5) Using Conventional Methods: For comparison, we used the conventional SOM for the same logistic function. For producing the self-organizing maps, the well-known SOM toolbox of Vesanto et al. [37] was used, because the final results of SOMs have been very different given the small changes in implementation such as initial conditions. We have

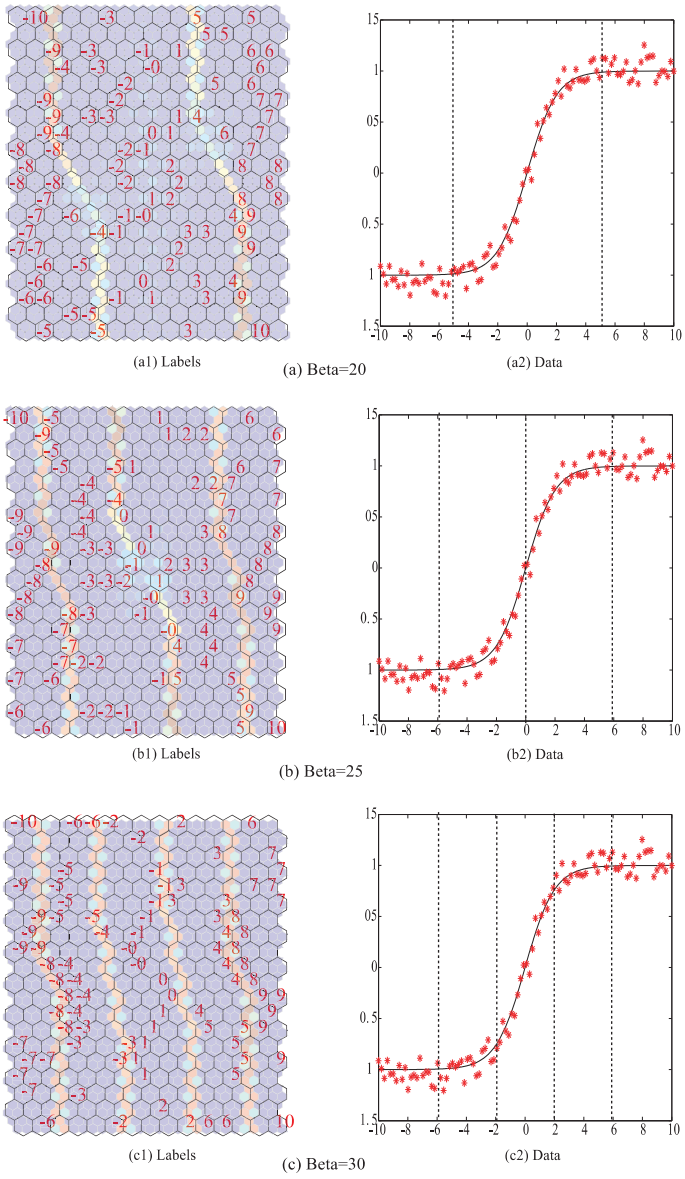


Fig. 5. U-matrices with 20 by 15 when the parameter β was changed from 20 (a) to 30 (c).

confirmed the reproduction of stable final results by using this package. In the SOMs, the Batch method was used, which has shown better performance than the popular real-time method in terms of visualization, quantization and topographic errors.

The mean squared error (MSE) was the same (0.00126) as that of the contradiction ratio in Table II. The quantization error was 0.048, see Table II. On the other hand, by the contradiction resolution, the lowest error was 0.073, as shown in Table I. Thus, the conventional SOM showed lower quantization errors. The topographic error was 0.580, shown in Table II. Thus, the topographic error by the contradiction ratio in Table I was much lower than that obtained by the conventional SOM. Mutual information was 1.635 by the SOM, while it increased to 1.654 via the contradiction resolution, see Table I. Thus, mutual information by contradiction resolution was slightly higher than that by the conventional SOM. The RBF networks with the Ridge regression and generalized cross validation

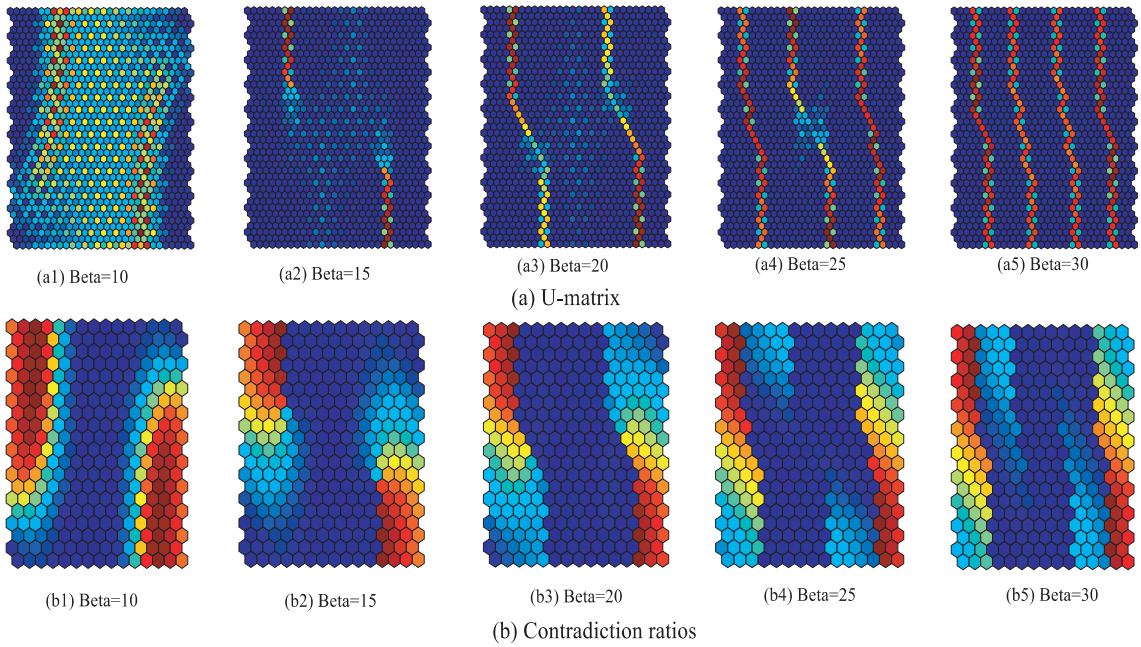


Fig. 4. U-matrices and contradiction ratios with 20 by 15 when the parameter β was changed from 10 (a) to 30 (c).

TABLE II. MSE FOR TESTING DATA, QUANTIZATION ERRORS (QE), TOPOGRAPHIC ERRORS (TE) AND MUTUAL INFORMATION (INF) BY THE SOM AND RADIAL-BASIS NETWORKS. THE SYMBOL "FS" DENOTES THE FORWARD SELECTION RBF WITH THE BAYESIAN INFORMATION CRITERION AND "RR" DENOTES THE RIDGE REGRESSION RBF WITH THE GENERALIZED CROSS VALIDATION.

Size or methods	MSE	QE	TE	INF
SOM	0.00126	0.048	0.580	1.635
FS(bic)	0.00142			
RR(gcv)	0.00124			

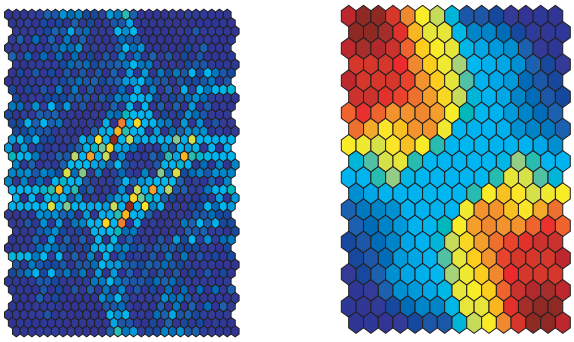


Fig. 6. U-matrices (a) and contradiction ratios (b) by the conventional SOM when the network size was 20 by 15.

[33], [34], [35] showed the best result of 0.00124 (MSE). However, by the RBF with forward selection and the Bayesian information criterion, the worst error of 0.00142 was obtained. By visual inspection, we could see that two class boundaries, which were rather weak, moved to the center of the U-matrix in Figure 6 (a). By the contradiction ratios, two classes on the upper right and lower left hand sides of the map became larger, as in Figure 6(b).

TABLE III. MSE FOR TESTING DATA, QUANTIZATION ERRORS (QE), TOPOGRAPHIC ERRORS (TE) AND MUTUAL INFORMATION WHEN THE TIME LAG WAS CHANGED FROM 1 TO 10 AND THE PARAMETER β WAS 10.

Lag	MSE	QE	TE	INF
1	0.068	0.058	0.776	0.794
2	0.077	0.286	0.385	0.802
3	0.066	0.370	0.304	0.800
4	**0.050**	0.450	0.194	0.793
5	0.058	0.447	0.133	0.794
6	0.060	0.495	0.257	0.794
7	0.061	0.536	0.174	0.791
8	0.064	0.559	0.181	0.789
9	0.063	0.571	0.130	0.790
10	0.065	0.608	0.097	0.786

B. Dollar-Yen Exchange Rate Prediction

1) Time Lag: In the dollar-yen exchange rate prediction, the rate at time t must be estimated by the previous q rates $x_{t-1,\ldots,t-q}$. The time lag q must be determined before the experiment. We determined the time lag based on the errors (MSE) for the testing data when the parameter β was ten. Table III shows the summary of the experimental results. The mean squared error (MSE) for the testing data decreased and increased from 0.068 (lag=1) to 0.065 (lag=10). When the time lag was four, the minimum value of 0.050 was obtained. The quantization errors (QE) increased from 0.058 to 0.608, while the topographic errors (TE) decreased from 0.776 to 0.097. Mutual information decreased slightly from 0.802 (lag=2) to 0.786 (lag=10). The experimental results showed that the mean squared error (MSE) was the lowest when the time lag was four. The quantization errors (QE) increased gradually when the lag increased, while the topographic errors decreased. In addition, mutual information remained almost unchanged. From these results, in our experiments, the time lag was determined to be four.

TABLE IV. MSE FOR THE TESTING DATA, QUANTIZATION ERRORS (QE), TOPOGRAPHIC ERRORS (TE) AND MUTUAL INFORMATION WHEN THE PARAMETER β WAS CHANGED FROM 1 TO 19 FOR THE 20 BY 15 MAP. . THE SYMBOL CC IN THE FINAL ROW REPRESENTS CORRELATION COEFFICIENTS BETWEEN INFORMATION (INF) AND THE OTHER MEASURES.

β	MSE	QE	TE	INF
1	5.024	3.977	0.000	0.000
3	0.108	0.861	0.161	0.715
5	0.155	0.611	0.300	0.787
7	0.109	0.573	0.294	0.794
9	**0.049**	0.508	0.317	0.796
11	0.049	0.401	0.161	0.796
13	0.055	0.326	0.194	0.797
15	0.054	0.298	0.244	0.797
17	0.055	0.281	0.228	0.797
19	0.054	0.263	0.206	0.795
CC	-0.973	-0.989	0.709	

2) Prediction Performance: Table IV shows the summary of the experimental results. The minimum value for MSE was 0.049 when the parameter β was nine and eleven. The quantization errors (QE) decreased gradually to 0.263 (β=19). The topographic errors (TE) fluctuated from zero (β=1) to 0.317(β=9). Mutual information immediately reached the level of 0.794 when the parameter β was seven. Then, mutual information seemed to be stable. The correlation coefficients between MSE, QE, TE and mutual information were -0.973, -0.989 and 0.709, respectively.

The experimental results showed that the mean squared errors for the testing data was the lowest when the parameter β was around 10. The quantization errors decreased gradually when the parameter β was increased. The topographic errors fluctuated when the parameter β was increased. In addition, mutual information had a strong negative correlation with the MSE and quantization errors. This means that if we want to decrease the MSE and quantization errors, we have only to increase mutual information, meaning that all we have to do given decreasing MSE and quantization errors is to increase the parameter β.

3) Visual Performance: Figure 7 shows the U-matrices and contradiction ratios when the parameter β was increased from 3 to 19. When the parameter β was increased from three in Figure 7 (a1) and (b1) to seven in Figure 7 (a3) and (b3), one clear class boundary could be seen and neurons with higher contradiction ratios could be seen on the upper side of the map. When the parameter β was increased to nine in Figure 7 (a4) and (b4), the class boundary on the upper side was twisted, and another boundary appeared at the bottom. Neurons with the contradiction ratios were also differentiated on the upper side, and a small group of higher contradiction ratios could be seen on the lower side of the map. When the parameter β was increased to 19 in Figure 7 (a5) and (b5), the class boundaries on the upper side of the map became further complicated, and other boundaries at the bottom appeared. The contradiction ratios with high values scattered to the four corners of the map, as in Figure 7 (b5).

4) Interpretation: We tried to interpret how contradiction resolution classified the entire period. We interpreted the maps when the parameter β was fixed so as to minimize the MSE for the testing data. Figure 8 shows maps with labels based on the U-matrices when the network size was 20 by 15 (c)

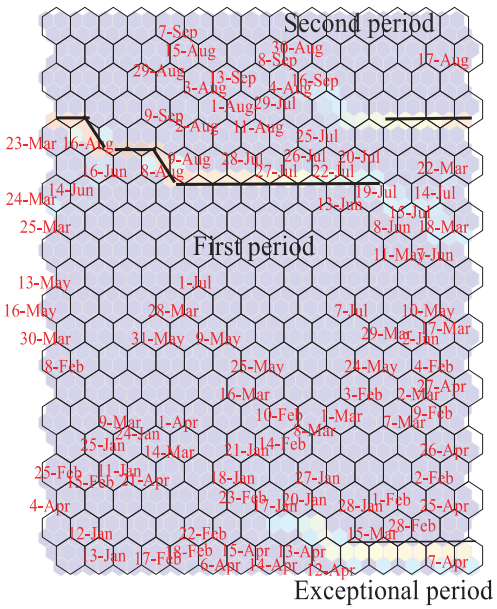


Fig. 8. Map with labels and class boundaries based on the U-matrices and with 20 by 15 neurons for the dollar-yen exchange rates.

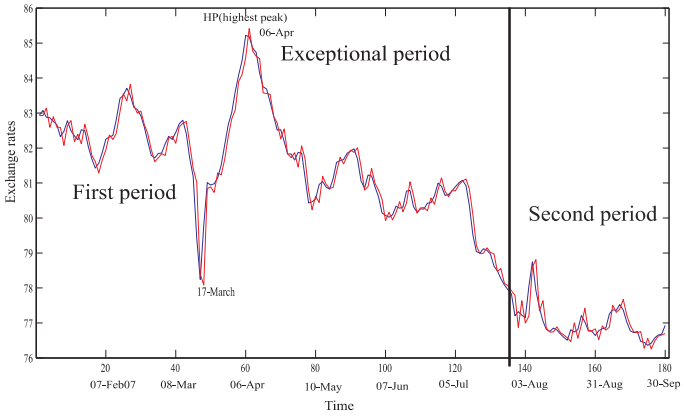


Fig. 9. Dollar-yen exchange rates during 2011.

and the parameter β was nine. The entire period seemed to be divided into three. The first period saw the rates gradually decreasing from January to July in Figure 9. The second period saw the dollar-yen rates remain relatively stable during August and September, as in Figure 9. In addition, a period with the highest rates of April was separated by the class boundary on the lower side.

5) Conventional Methods: For comparison, we used the conventional SOM and RBF networks. Table V shows the summary of experimental results when we used the conventional SOM and the radial basis networks. By the SOM, the MSE was 0.052, which was higher than that obtained by the contradiction ratio in Table IV.

Using the RBE network with forward selection and the Bayesian information criterion, the obtained error was 0.055. When the RBF network with Ridge regression was used, the MSE increased to 0.063, which was larger than that obtained by our method in Table IV. Thus, our method showed the possibility of better prediction performance. For the other measures, the quantization and topographic errors were smaller

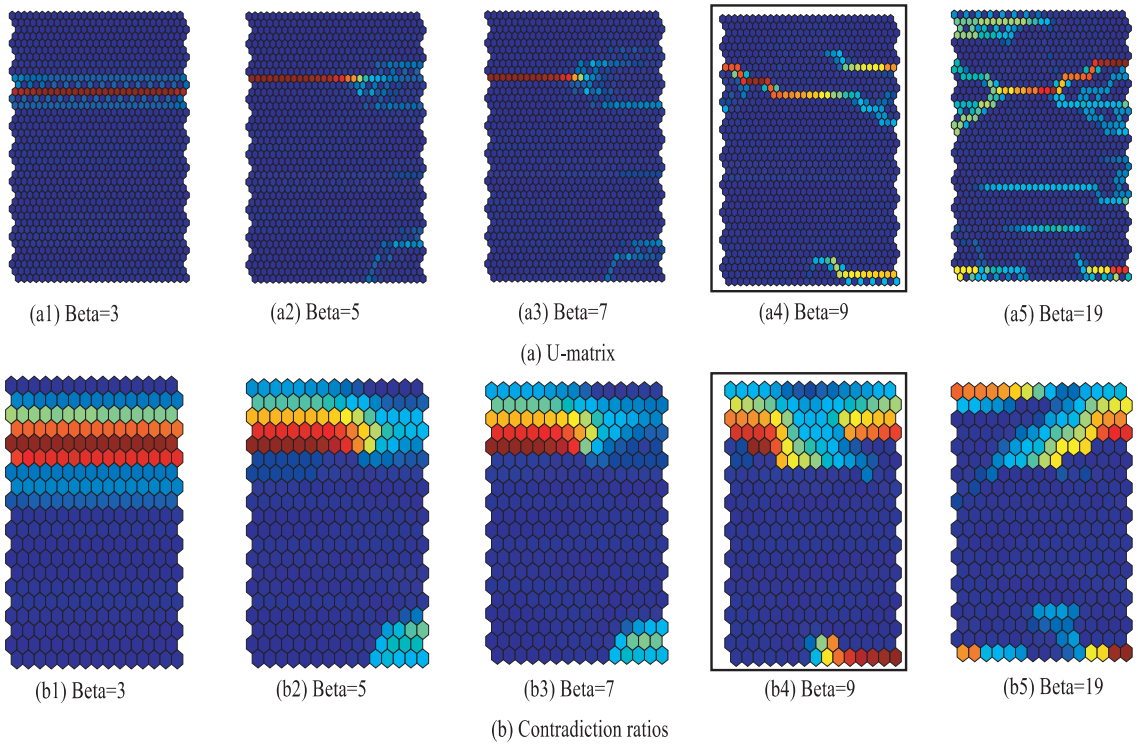


Fig. 7. U-matrices with 20 by 15 when the parameter β was increased from 3 to 19.

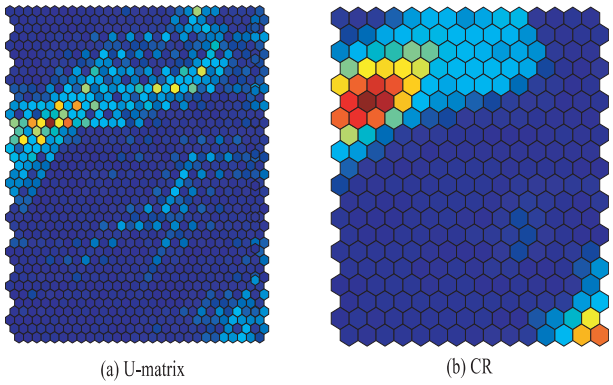


Fig. 10. Contradiction ratios and U-matrices by the conventional methods.

TABLE V. MSE FOR TESTING DATA, QUANTIZATION ERRORS (QE), TOPOGRAPHIC ERRORS (TE) AND MUTUAL INFORMATION BY THE CONVENTIONAL SOM AND RBF NETWORKS. "FS" DENOTES THE FORWARD SELECTION RBF WITH THE BAYESIAN INFORMATION CRITERION AND "RR" DENOTES THE RIDGE REGRESSION RBF WITH THE GENERALIZED CROSS VALIDATION.

Methods	MSE	QE	TE	INF
SOM	0.052	0.235	0.117	0.773
FS(bic)	0.055			
RR(gcv)	0.063			

than those obtained by our method in Table IV. In addition, mutual information was also smaller than that obtained by the contradiction ratio in Table IV.

Figures 10 (a) and (b) show the contradiction ratios and U-matrices obtained by the conventional SOM. Class boundaries on the upper side of the matrix seemed to be present, but were weaker than those obtained by our method, as shown in Figure 7. In addition, a group of neurons with higher contradiction ratios was smaller and weaker compared to those present in our method, see Figure 7.

C. Discussion

1) Validity of Methods and Experimental Results: In this paper, we proposed a new information-theoretic method to resolve contradiction in neural networks. We supposed that a neuron can be evaluated from multiple points of view. To fully understand the characteristics of the neuron, we should examine the characteristics obtained from these different types of evaluation. For simplification, we suppose that a neuron can be viewed from two forms of evaluation. These two types of evaluation are self- and outer-evaluation, respectively. For the self-evaluation, a neuron's firing rate is determined by evaluating only relations between the neuron and the incoming input patterns. For the outer-evaluation, a neuron's firing rate is determined by evaluating relations between the incoming input patterns and all its surrounding neurons. If contradiction between self- and outer-evaluation exists, this contradiction should be reduced as much as possible. In this way, we expect that the characteristics shared by self- and outer-evaluation will be enhanced, and eventually, that important characteristics can be intensified. Thus, the contradiction resolution can be expected to improve visualization and prediction performance by extracting important characteristics in input patterns.

First, the experimental results confirmed that interpretation by visualization could be improved. In the logistic data, weak class boundaries were produced in terms of the U-matrix by the conventional SOM, as in Figure 6. By contradiction resolution,

depending on the parameter β, we were able to explicitly identify three to five classes in the logistic data in Figure 4. For the dollar-yen exchange rates, the number of class boundaries differed depending on the parameter β, see Figure 7. The conventional SOM failed to produce explicit class structure in terms of the U-matrix in Figure 10. Compared with the class structure obtained by the conventional SOM, the class structure obtained by contradiction resolution was clearer, as in Figure 7.

As mentioned in the introduction section, many attempts have been made to visualize final representations in self-organizing maps [5], [6], [7], [8], [9], [10], [13], [14], [16]. However, it has been difficult to visualize SOMs' knowledge at the present of stage of techniques. Our method of contradiction resolution succeeded in producing explicit class structure in two experimental results. These experimental results showed that distinction between self- and outer-evaluation had effects to clarify class structure.

However, the quality of visualized maps was not necessarily improved. For the logistic data, quantization errors were not necessarily smaller than those obtained by the conventional SOM in Table I and II. For the dollar-yen exchange rates, quantization and topographic errors were not smaller than those obtained by the conventional SOM in Tables IV and V. Only for the topographic errors of the logistic data, it was possible that the errors were smaller than those by obtained the conventional SOM, see Tables I and II. Though improved predication performance was obtained, there is the possibility that quantization and topographic errors were scarified. Thus, careful attention should be paid to final representations for interpretation.

Favorable results were obtained for prediction performance. For the logistic data, contradiction resolution showed that the MSE for the testing data was almost equivalent to that obtained by the conventional SOM and RBF networks, as in Tables I and II. In the dollar-yen exchange rates, the MSE by the contradiction reduction was the lowest in Tables IV and V. In both cases, clearer class structures could be identified as in Figures 7 and 10. This shows that contradiction resolution could be used to improve visualization while keeping generalization errors low.

Because the evaluation criteria between our method and the conventional RBF were different from each other, we could not say definitely that contradiction resolution showed better prediction performance. However, the experimental results showed the possibility that prediction performance by contradiction resolution is not necessarily contrary to visualization performance.

2) Limitation of the Method: One of the main problems is that we have not yet determined criteria to choose optimal representations for visualization and prediction. For visualization, the experimental results showed that when the parameter β was increased, class structure became complicated. Because we do not yet have criteria to measure the clarity of class structure, all we have to do is to visually inspect the final representations and to choose the best possible representations for visualization. Thus, it is apparent that we need objective criteria to choose optimal representations for better visualization.

Objective criteria have not existed to choose the optimal representations for prediction performance either. One of the ways to solve this problem is to use mutual information. In the two experimental results, mutual information was correlated with quantization errors and MSE. The correlation coefficients were close to one for the two data sets in Tables I and IV. This means that quantization and MSE can be decreased by increasing mutual information. Thus, one way to obtain optimal representations is to monitor mutual information. Then, the optimal point should be where mutual information ceases to increase, namely, where mutual information reaches a stable point.

3) Possibility of the Method: The possibility of the method was explained in terms of a new SOM with a new visualization method for self-organizing maps and an extension to multiple types of evaluation. First, our method is related to a new type of self-organizing maps which produce more explicit class structure with a new visualization method. As mentioned in the introduction section, it has been very difficult to visualize SOM knowledge because final representations are usually not easily interpretable. There have been many attempts to visualize SOM knowledge more clearly. However, even if sophisticated visualization methods are used, it is still difficult to interpret SOM knowledge. In our experiments, by using the conventional SOM, weak boundaries were detected both for the logistic data in Figure 6 and for the dollar-yen exchange rates in Figure 10. However, even if class boundaries were weak in terms of a U-matrix, clear contradiction ratios could be obtained in the above figures. In addition, the contradiction ratios for each neuron showed clear class structure for both cases in Figures 5 and 8. These results show the possibility of a new SOM with a visualization method by using the contradiction ratios.

Second, we have restricted ourselves to two types of evaluation, namely, self- and outer-evaluation. This is because they are easily implemented when we suppose the two types of evaluation. However, as mentioned in the introduction section, we think that a neuron should be evaluated from as many different viewpoints as possible. Contradiction should be computed among many different types of evaluation, and this contradiction should be reduced as much as possible. If it is possible to take into account many types of evaluation for neurons, the characteristics of the neurons can be more exactly interpreted.

IV. Conclusion

In this paper, we proposed a new type of information-theoretic method called "contradiction resolution." We supposed that a neuron should be evaluated differently and those different types of evaluation should be unified. For simplification, we restricted ourselves to only two types of evaluation, namely, self- and outer-evaluation. In the self-evaluation, a neuron is evaluated for itself, while in the outer-evaluation, a neuron is evaluated by all its neighboring neurons. Contradiction between the two types of neurons is represented by the Kullback-Leibler divergence. Contradiction ratio in terms of the Kullback-Leibler divergence is reduced as much as possible.

We applied the method to the logistic and dollar-yen exchange rates. For the logistic data, experimental results confirmed that several explicit class boundaries which could

not be detected by the conventional self-organizing maps were detected by our method. In addition, better generalization performance was obtained without significantly degrading topographic preservation. In the dollar-yen exchange rates, class structure obtained by contradiction resolution was much better than that obtained by the conventional self-organizing maps. The best MSE was obtained by using contradiction resolution, though different criteria were used for comparison. At the least, the experimental results showed a possibility of better prediction performance. However, this improved performance was obtained by sacrificing quantization and topographic errors.

Though there are several problems in our method, such as the selection of optimal representations, it still, according to the experimental results, has shown its potential for visualizing data without sacrificing prediction performance. This shows that we can develop neural networks with better interpretation while keeping better predication performance.

REFERENCES

[1] R. Kamimura, "Contradiction resolution and its application to self-organizing maps," in *SMC, The 2012 International Joint Conference on*, IEEE, 2012.

[2] T. Kohonen, *Self-Organizing Maps*. Springer-Verlag, 1995.

[3] T. Kohonen, "The self-organization map," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464–1480, 1990.

[4] T. Kohonen, "Self-organized formation of topological correct feature maps," *Biological Cybernetics*, vol. 43, pp. 59–69, 1982.

[5] J. W. Sammon, "A nonlinear mapping for data structure analysis," *IEEE Transactions on Computers*, vol. C-18, no. 5, pp. 401–409, 1969.

[6] A. Ultsch and H. P. Siemon, "Kohonen self-organization feature maps for exploratory data analysis," in *Proceedings of International Neural Network Conference*, (Dordrecht), pp. 305–308, Kulwer Academic Publisher, 1990.

[7] A. Ultsch, "U*-matrix: a tool to visualize clusters in high dimensional data," Tech. Rep. 36, Department of Computer Science, University of Marburg, 2003.

[8] J. Vesanto, "SOM-based data visualization methods," *Intelligent Data Analysis*, vol. 3, pp. 111–126, 1999.

[9] S. Kaski, J. Nikkila, and T. Kohonen, "Methods for interpreting a self-organized map in data analysis," in *Proceedings of European Symposium on Artificial Neural Networks*, (Bruges, Belgium), 1998.

[10] I. Mao and A. K. Jain, "Artificial neural networks for feature extraction and multivariate data projection," *IEEE Transactions on Neural Networks*, vol. 6, no. 2, pp. 296–317, 1995.

[11] C. De Runz, E. Desjardin, and M. Herbin, "Unsupervised visual data mining using self-organizing maps and a data-driven color mapping," in *Information Visualisation (IV), 2012 16th International Conference on*, pp. 241–245, IEEE, 2012.

[12] S. Shieh and I. Liao, "A new approach for data clustering and visualization using self-organizing maps," *Expert Systems with Applications*, 2012.

[13] H. Yin, "ViSOM-a novel method for multivariate data projection and structure visualization," *IEEE Transactions on Neural Networks*, vol. 13, no. 1, pp. 237–243, 2002.

[14] M.-C. Su and H.-T. Chang, "A new model of self-organizing neural networks and its application in data projection," *IEEE Transactions on Neural Networks*, vol. 123, no. 1, pp. 153–158, 2001.

[15] S. Wu and T. Chow, "Prsom: A new visualization method by hybridizing multidimensional scaling and self-organizing map," *Neural Networks, IEEE Transactions on*, vol. 16, no. 6, pp. 1362–1380, 2005.

[16] L. Xu, Y. Xu, and T. W. Chow, "PolSOM-a new method for multidimentional data visualization," *Pattern Recognition*, vol. 43, pp. 1668–1675, 2010.

[17] Y. Xu, L. Xu, and T. Chow, "Pposom: A new variant of polsom by using probabilistic assignment for multidimensional data visualization," *Neurocomputing*, vol. 74, no. 11, pp. 2018–2027, 2011.

[18] L. Xu and T. Chow, "Multivariate data classification using polsom," in *Prognostics and System Health Management Conference (PHM-Shenzhen), 2011*, pp. 1–4, IEEE, 2011.

[19] M. Bogdan and W. Rosenstiel, "Detection of cluster in self-organizing maps for controlling a prostheses using nerve signals," in *9th European Symposium on Artificial Neural Networks. ESANN' 2001. Proceedings. D-Facto, Evere, Belgium*, pp. 131–6, 2001.

[20] D. Brugger, M. Bogdan, and W. Rosenstiel, "Automatic cluster detection in kohonen's som," *Neural Networks, IEEE Transactions on*, vol. 19, no. 3, pp. 442–459, 2008.

[21] T. Haraguchi, H. Matsushita, and Y. Nishio, "Community self-organizing map and its application to data extraction," in *Neural Networks, 2009. IJCNN 2009. International Joint Conference on*, pp. 1107–1114, IEEE, 2009.

[22] Z. Li, R. Wang, and L. Chen, "Extracting community structure of complex networks by self-organizing maps," in *Proceedings of the third international symposium on optimization and systems biology (OSB' 09)*, pp. 48–56, 2009.

[23] D. E. Rumelhart, G. E. Hinton, and R. Williams, "Learning internal representations by error progagation," in *Parallel Distributed Processing* (D. E. Rumelhart and G. E. H. et al., eds.), vol. 1, pp. 318–362, Cambridge: MIT Press, 1986.

[24] L. I. Nord and S. P. Jacobsson, "A novel method for examination of the variable contribution to computational neural network models," *Chemometrics and Intelligent Laboratory Systems*, vol. 44, pp. 153–160, 1998.

[25] A. Micheli, A. Sperduti, and A. Starita, "Analysis of the internal representations developed by neural networks for structures applied to quantitative structure-activity relationship studies of benzodiazepines," *Journal of Chemical Information and Computer Sciences*, vol. 41, pp. 202–218, 2001.

[26] G. G. Towell and J. W. Shavlik, "Extracting refined rules from knowledge-based neural networks," *Machine learning*, vol. 13, pp. 71–101, 1993.

[27] M. Ishikawa, "Structural learning with forgetting," *Neural Networks*, vol. 9, no. 3, pp. 509–521, 1996.

[28] M. Ishikawa, "Rule extraction by successive regularization," *Neural Networks*, vol. 13, pp. 1171–1183, 2000.

[29] P. Howes and N. Crook, "Using input parameter influences to support the decisions of feedforward neural networks," *Neurocomputing*, vol. 24, p. 1999, 1999.

[30] R. Feraud and F. Clerot, "A methodology to explain neural network classification," *Neural Networks*, vol. 15, pp. 237–246, 2002.

[31] R. Setiono, W. K. Leow, and J. M. Zurada, "Extracting of rules from artificial neural networks for nonlinear regression," *IEEE Transactions on Neural Networks*, vol. 13, no. 3, pp. 564–577, 2002.

[32] R. Kamimura, "Self-enhancement learning: target-creating learning and its application to self-organizing maps," *Biological cybernetics*, pp. 1–34, 2011.

[33] M. Orr *et al.*, "Introduction to radial basis function networks," *University of Edinburg*, 1996.

[34] M. Orr, "Recent advances in radial basis function networks," *Relatório técnico, Centre for Cognitive Science, University of Edinburgh*, 1999.

[35] M. Orr, "Regularization in the selection of radial basis function centers," *Neural computation*, vol. 7, no. 3, pp. 606–623, 1995.

[36] K. Kiviluoto, "Topology preservation in self-organizing maps," in *In Proceedings of the IEEE International Conference on Neural Networks*, pp. 294–299, 1996.

[37] J. Vesanto, J. Himberg, E. Alhoniemi, and J. Parhankangas, "SOM toolbox for Matlab," tech. rep., Laboratory of Computer and Information Science, Helsinki University of Technology, 2000.

AUTHOR PROFILE

Ryotaro Kamimura is currently a professor of IT Education Center and Graduate School of Science and Technology of Tokai University in Japan. His research interests are information-theoretic approach to neural computing.

Weapon Target Assignment with Combinatorial Optimization Techniques

Asım Tokgöz[1], Serol Bulkan[2]
Department of Industrial Engineering
Marmara University
Göztepe, Istanbul 34722, TURKEY

***Abstract*—Weapon Target Assignment (WTA) is the assignment of friendly weapons to the hostile targets in order to protect friendly assets or destroy the hostile targets and considered as a NP-complete problem. Thus, it is very hard to solve it for real time or near-real time operational needs. In this study, genetic algorithm (GA), tabu search (TS), simulated annealing (SA) and Variable Neighborhood Search (VNS) combinatorial optimization techniques are applied to the WTA problem and their results are compared with each other and also with the optimized GAMS solutions. Algorithms are tested on the large scale problem instances. It is found that all the algorithms effectively converge to the near global optimum point(s) (a good quality) and the efficiency of the solutions (speed of solution) might be improved according to the operational needs. VNS and SA solution qualities are better than both GA and TS.**

Keywords—Weapon Target Assignment; Combinatorial Optimization; Genetic Algorithm; Tabu Search; Simulated Annealing; Variable Neighborhood Search; WTA; WASA; TEWASA

I. INTRODUCTION

In military operations, problems in planning and scheduling often require feasible and close to optimal solutions with the limited computing recourses and within very short time periods [1]. Especially the weapons developed in contemporary technology give very less chance to defend friendly assets as enemy forces execute complex saturated attacks. Therefore, quick and efficient reactions to subsidise these attacks become very vital to survive in the combat arena. Thus, assigning the limited resources (own weapons) to the hostile targets to achieve certain tactical goals [2] becomes an important issue called as Weapon Target Assignment (WTA) problem.

The WTA problem is a classical constrained combinatorial optimization problem arising in the field of military operations research [2] and it is NP-complete [3]. The term "allocation" and "assignment" are used analogously in the literature. The complexity of this problem drastically increases if you add the temporal and spatial constraints of both the friendly forces and hostile targets. The allocation of available capabilities to the correct targets needs complex calculations in a very short time. There are various simple traditional algorithms in literature that solve this problem such as graph search, implicit enumeration algorithms, dynamic programming, branch and bound algorithms,[4, 5] and simulated annealing. Even though these algorithms are simple, it is difficult to implement them especially when the problem scale is large. With the evolution of computer technology, some naive algorithms are proposed like analytical hierarchy process [6], network flow based methods [7], neural networks [8], genetic algorithms, tabu search, ant colony algorithm (ACO) and particle swarm optimization (PCO), very large scale neighborhood (VLSN), and maximum marginal return (MMR).

There are many works in the literature regarding the algorithms mentioned above. All of them almost define the problem similar but there are some differences on the solution approach. These are generally based on many factors, such as capability of targets and weapons, defense strategies, current combat conditions, the dynamic or static solution, layered defense or not, protecting the friendly assets or killing hostile targets or both (asset-based or target based), assessing the threats according to the capability and intent. Therefore the problem formulation differs the way how you approach the solution. All of them state the WTA as hard assignment problem and a comprehensive mathematical problem formulation example can be found at [4, 2]. You can also find a detailed literature overview and algorithm comparison from different perspectives at [9, 2, 10]. Regarding these works, the problem formulation looks similar, but there are some key points that need to be emphasized;

- Lee et al. [11] uses partially mapped crossover (PMX), inversion mutation and simulated annealing as local search.
- Lee et al. [12] uses greedy reformation scheme so as to have locally optimal offspring (greedy eugenics) which is a kind of novel crossover operator (EX) that try to inherit the good genes from the parents. They give information about various eugenics mechanisms orderly simple eugenics, simulated annealing, immune operator, and greedy eugenics.
- Lu et al. [13] uses uniform generation of initial population, roulette wheel selection based on sigma truncation scaling fitness and self-adaptive parameters for genetic operations (dynamic probability of mutation and crossover). Crossover adopts the two-point-crossover operation while mutation employs swap mutation operator.
- Shang et al. [14] added an extra step to the classical GA, which comes after creating new chromosomes by crossover and mutation, which is to apply local search to create new chromosomes and then evaluate the new chromosomes. After that, select the chromosomes with

the best fitness from the population. For LS they propose the immune GA (IGA) which is composed from vaccination and immune selection, of which the former is used for raising fitness and later is for preventing the deterioration. They also used a new crossover operator called as elite preserving crossover operator to enrich a more effective search.

- Dou et al. [15] uses chromosomes in matrix representation and adopts the dynamically adjusting punishment gene and self-regulating punishment rates. Initial population is selected in a reasonable manner so as to satisfy the constraints automatically. Roulette wheel selection and dynamically adjusted crossover and mutation probabilities are used.
- Li et al. [16] uses matrix-type encoding for chromosomes and a new crossover/mutation operator called "circle swap".
- Zhihua et al. [17] uses local information to get a better initial feasible solution by deducing the survival value matrix (heuristic information).
- Johansson and Falkman [10] uses enhanced MMR, GA, and PSO.

II. Problem Definition

The military domain requires operators to evaluate the tactical situation in a very short time interval, almost real-time, and then make decisions to protect the friendly assets against the enemy threats by allocating the available own weapons to the hostile units. Furthermore, when the severe stressful battlefield conditions added to this decision making process, operators become more overburdened and they need computerized decision support system tools to handle this complex situation [18]. This brings forth the evaluation on tracks (targets) that might have the hostility intent and also capabilities that might harm the friendly assets briefly called as Threat Evaluation (TE). TE process mostly requires decision support tools as the number of targets that need to be evaluated is crowded and rapid decisions are required to eliminate the enemy assets in time.

After the TE, operators need to assign the appropriate and efficient weapon or weapon-sensor pairs to the targets in order to get the maximum benefit which is called as Weapon Assignment or Weapon Assignment and Sensor Allocation (WA/ WASA). This assignment/allocation level may be at different force structures such as single asset, task group, and force. In this paper we will investigate the efficiency of genetic algorithm, tabu search, simulated annealing and variable neighborhood search algorithms which make the assignments of weapons to targets by employing a shoot-look-shoot (SLS) engagement policy.

As stated in [4, 8], the objective may be to minimize the total threat of all targets or maximize the total value of assets surviving through the whole defense process.

In this calculation, finding the probability of kill given hit of targets (threats) is a controversial issue as this information mostly confidential and hard to get. You can use this information in the algorithms by considering the availability of it. The WTA process is usually an ongoing process as it requires observing the status of targets and friendly weapons continuously to replan the engagements. The annihilated targets and unavailable friendly weapons are discarded for the next iterative plan phase. This process can be considered as the famous OODA (Observe, Orient, Decide and Act) loop.

The most of the work in the literature (asset-based, target-based, static or dynamic or combination of them) assume that the weapons are fired simultaneously, which is not the case in reality, and this one step engagement period is called as stage. After the each stage a damage assessment is performed, and based on this assessment available weapons are reassigned to the surviving targets iteratively.

III. Problem Formulation

In this paper, the series of static WTA (SWTA) is assumed and the iterative defense process is used. The problem formulation is a subset of [4, 2], every computing result will mimic the results of the stages by updating the status of the assets, weapons and targets. The status after the execution of found solution will be input for next iterative computing. Therefore the stages are omitted from the problem formulization and it is defined as the following;

T is the number of offensive targets with their attack aims exposed, K is the number of assets of the defender, W is the weapons available to intercept the targets,

$X = \left[x_{ij} \right]_{WxT}$ is the decision variable, $x_{ij} = 1$ if weapon i is assigned to target j , $x_{ij} = 0$ otherwise,

T_k is the index set of the targets that threaten asset k,

W_j is the index set of the weapons that are assigned to intercept target j,

v_k is the value of asset k,

q_{jk} is the lethality probability that target j destroys asset k,

p_{ij} is the lethality probability that weapon i destroys target j.

The expected total value of assets surviving through the whole defense process to be maximized,

$$J(X) = \sum_{k=1}^{K} v_k \prod_{j \in T_k} \left[1 - q_{jk} \prod_{i \in W_j} (1 - p_{ij})^{x_{ij}} \right] \quad (1)$$

The constraints involved in WTA primarily include capability constraints, strategy constraints, resource constraints, and engagement feasibility constraints as follows,

$$\sum_{j=1}^{T} x_{ij} \le n_i \quad \forall_i \in \{1, 2, ..., W\} \quad (2)$$

$$\sum_{i=1}^{W} x_{ij} \le m_j \quad \forall_j \in \{1, 2, ..., T\} \quad (3)$$

$$\sum_{j=1}^{T} x_{ij} \le N_i \quad \forall_i \in \{1, 2,....,W\} \tag{4}$$

$$x_{ij} \le f_{ij} \ \forall_i \in \{1, 2,....,W\} \quad \forall_j \in \{1,2,....,T\} \tag{5}$$

The constraint set (2) reflects the capability of weapons which is to fire at multiple targets at the same time. Even though modern weapon systems can shoot multiple targets at a time, this kind of weapon systems can be evaluated as multiple separate weapons. As a result of these facts, we set $n_i = 1$ for $\forall i \in \{1, 2, . . .,W\}$ [2].

The constraint set (3) restricts the weapon cost for each target. The setting of m_j (j = 1, 2,...,T) generally based on the combat performance of available weapons and desired lethal probability [2]. In this research, it is assumed that $m_j = 1$ for $\forall j \in \{1, 2,...,T\}$. This is a rational setting for missile-based defense systems and the "SLS" engagement policy [19]. For artillery-based defense systems, the value of m_j (j = 1, 2,...,T) may be substantially increased under the same demand in defensive strength. Therefore, the constraints in (3) can be considered as a strategy constraint [2].

The constraint set (4) basically reflects the amount of ammunition provided for weapons. N_i (i = 1, 2, . . .,W) is the maximum number of times that weapon i can be used due to its ammunition capacity [2].

In the constraint set (5), f_{ij} indicates the actual engagement feasibility for weapon i assigned to target j. $f_{ij} = 0$ if weapon i cannot engage to target j with any potential reason (weapon range, blind zone etc.); $f_{ij} = 1$ otherwise [2]. The time window of targets and weapons is the main factor that influences the engagement feasibility [20, 21]. Some scholars also use the term "cue" or "deadline" when referring to temporal and spatial issues [22, 23]. This will influence the time frame on the engagement feasibility of weapons. In addition, it increases the complexity of Dynamic WTA (DWTA) problems as well and the difficulty of creating feasible solutions. In this situation, it is difficult to implement an appropriate operator that can generate new solutions and guarantee their feasibility as well.

IV. Algorithms Used to Solve WTA

This paper will not discuss the details of the algorithms implemented here as you can find them in many text books and in many scholars' work. The work here is focused on the application of combinatorial optimization algorithms (genetic algorithm, tabu search, simulated annealing, and variable neighborhood search) to the weapon target assignment problem. Combinatorial Optimization Library framework of Skalicka [24] is modified and used in this work. This framework allows easy implementation of own algorithms and provides effective tools for bench-marking different algorithms on a set of various problems.

A. Genetic Algorithm

The theory of natural selection claims that the species living nowadays are evolved by complying with the harsh environmental conditions through millions of years of adaptation. The changes in the environment had forced the species to alter their genetic characteristics in order to survive. In ecosystem, recourses have always been scarce; therefore, a certain number of organisms have always competed for the same resources in the habitat. The capability to acquire the resources will determine the future of that organism. If an organism can acquire the resources, then it will successfully procreate and its descendants will have a tendency to be numerous in the future. These organisms are considered as fitted to the ecosystem. Organisms with less capability have a liability to have fewer or no descendants in the next days. Eventually, the new generations in the population will be more fitted to the ecosystem than previous ones and evolution will continue likewise [25]. Genetic Algorithm (GA) is based on the same idea of natural evolution.

GA algorithm imitates the natural selection process during the search phase of the problem domain to reach better solutions. It is inspired by natural evolution, like inheritance, mutation, selection, and crossover and commonly used in optimization problems to generate efficient solutions. The pseudo code is given in Table I.

TABLE I. Pseudo Code For GA

```
Initialize Algorithm (population size,
mutation rate, and problem)
Create Initial Random Population ()
Find fitness of each and store the best
Check for end conditions
Do
   Optimize by selection, mutation and
   disaster
   Find fitness of new individuals and
   store the best
   (Kill the worst 2 individuals)
Until end conditions meet
```

1) Fitness

The fitness function for this problem is the objective function.

2) Individuals

An individual is a single solution and a group of the individuals forms the population. The structure of the individual is composed of various formatted data elements and it is called as chromosome. In this work, permutation encoding (real number coding) is used to represent the weapons and their order mimics the assignments of them to the targets sequentially.

Targets	1	2	3	4	5	6	7	8
Chromosome 1	3	2	1	5	4	8	6	7
Chromosome 2	5	7	8	2	1	3	4	6

Fig. 1. Sample chromosome structure

As you can see from the Figure 1, in chromosome 1, weapons 3, 2, 1, 5, 4, 8, 6, and 7 are assigned to targets 1, 2, 3, 4, 5, 6, 7, and 8 respectively. If weapon and target size is different then, dummy weapons/targets are used in order to make them equivalent. This number is also the population size of the genetic algorithm used in this work.

3) Selection

After creation of the initial population, first step is selecting the individuals for reproduction. This selection is random with a probability depending on the relative fitness of individuals, so that often the best ones have more chance to be selected then poor ones. [25]. Roulette Wheel Selection is used in this work.

4) Reproduction

In the second step, offspring are bred by the selected individuals. For generating new chromosomes, the algorithm can use recombination, mutation and disaster. In this paper a disaster is introduced with a probability that hoped to improve the solution by exchanging a small percentage of population with the new randomly generated individuals.

Crossover (Recombination) is the process of taking two parent solutions and producing two children from them. After selection (reproduction) the population is enriched with better people. Crossover operator executed on the appropriate couple with the hope that it generates a better offspring [25].

Crossover is a three step recombination operator [25]: (1) the reproduction operator selects at random a couple strings for mating. (2) A cross site is chosen at random along the string length. (3) Finally, the position values are exchanged between the two strings according to the cross site [25].

In this work Partially Mapped/Matched Crossover method (PMX) is used. For instance, 3 and 6 positioned genes are selected and the genes between these two points are swapped for the chromosomes of son and daughter. For son, the rest of the genes are copied from father to the same place if they do not exist at the son. Daughter is created analogously like son.

Father	3	2	1	5	4	8	6	7
Mother	5	7	8	2	1	3	4	6

Fig. 2. Chromosome structure of the parents

Son			8	2	1	3	6	7
Daughter		7	1	5	4	8		6

Fig. 3. Incomplete chromosome structure of the children

After that the holes filled as follows: (1) Take the gene from father check if it exists at the son. (2) if not, put the same gene to the same position at the son. (3) If that gene exists at the son, then find the same gene at the mother and return that gene position. Look at the gene at the father with position returned from the mother. If found new gene not exist at son, put that gene to the original hole position. Repeat this cycle until all the holes are filled.

Daughter is created analogously like son.

For 3 at the father at position 1; 4 is placed at position 1 at son after total of 12 checks as follows:

Father	3(1)	2	1(7)	5	4(10)	8(4)	6	7
Mother	5	7	8(6)	2	1(9)	3(3)	4	6

Fig. 4. Iterations at the chromosome of the parents

Son	4(11)		8(5)	2	1(8)	3(2)	6	7
Daughter		7	1	5	4	8		6

Fig. 5. Iteartions at the incomplete chromosome of the children

One check after the 11th is to control if 4 does exist at the son, seeing not there, and then positioned it to 1 at son. The final new two offspring are as follows:

Son	4	5	8	2	1	3	6	7
Daughter	2	7	1	5	4	8	3	6

Fig. 6. Completed chromosome structure of the children

Following the crossover, the chromosomes are exposed to the mutation which will recover the algorithm from local minimum trap. Mutation has the possibility to introduce the lost genetic materials as it randomly alters the genetic structure. It ensures against losing the irreversible genetic material and also has been considered as a simple search operator.

While crossover is improving the current solution with better ones, mutation is expected to explore the whole search space and maintains the genetic diversity in the population. It incorporates new genetic structures in the population by randomly changing some of their components [25].

In this work, swap mutation is used with the probability 0.1. A uniform random number is generated and compared with the mutation probability for each offspring. If this number is smaller than this probability, then individual is subject to mutation. If we assume that son is subject to mutation with the positions 2 and 6 then, we simply exchange the genes at those positions.

Son	4	5	8	2	1	3	6	7
Daughter	2	7	1	5	4	8	3	6

Fig. 7. Chromosome structure of the children after the crossover

After mutation, the new individuals are as follows;

Son	4	3	8	2	1	5	6	7
Daughter	2	7	1	5	4	8	3	6

Fig. 8. Chromosome structure of the children after the mutation

With a very small probability a natural disaster is introduced in that a small portion of the population is replaced with randomly created new individuals. It is useful for diversification.

5) Evaluation

After the reproduction, the fitness of the new chromosomes is evaluated.

6) Replacement

During the last step, the worst fitted two individuals are killed from the population.

B. Tabu Search

Tabu Search (TS) is unquestionably distinguishable from the local search technique, as it incorporates intelligence. It keeps track of the history of iterative search or, equivalently, to enable the search with memory [26]. Thus, it will avoid both applying the same operations repeatedly and revisiting local optima. A tabu list is incorporated into the algorithm to serve as the memory function and it is an important mechanism to recover from local optima [2].

Even though random components are introduced to overcome the technical difficulties, in fact, tabu search is not

based on the chance. The basic idea of tabu search is to use the memories (tabu list) to explore the search space of the problem and move from one solution to a neighboring solution continually [26]. Yet it may be seen as local search to some people, some methods are used to carry out the jumps in the solution space, in this manner, TS is not a pure local search.

In this research, the candidate list is constructed from the possible *n(n-1)/2* moves except the tabu ones where n is the problem size. Sequentially all moves are tried and the first candidate operation that makes the fitness better is executed. In order to get rid of the local optima, an aspiration criteria, which is to accept the current iteration best move among the all possible moves with a probability of 0.8 even though it is not better than the fitness found so far, is applied. For tabu list a short term memory is used with random duration. The structure of it as follows [26]:

		Site (Targets)				
		1	***2***	***3***	***4***	***5***
Objects (Weapons)	***1***	0	0	0	0	0
	2	0	0	0	0	0
	3	0	0	0	0	0
	4	0	0	0	0	0
	5	0	0	0	0	0

Fig. 9. Tabu list

		Site (Targets)				
		1	***2***	***3***	***4***	***5***
Objects (Weapons)	***1***	0	0	10	0	0
	2	10	0	0	0	0
	3	0	0	0	0	0
	4	0	0	0	0	0
	5	0	0	0	0	0

Fig. 10. Filled sample tabu list

Let's assume our current solution is (2, 4, 1, 5, 3) that is the assignment of weapons 2, 4, 1, 5, 3 to targets 1, 2, 3, 4, 5 respectively and we found a better move as the exchange of weapons (2 and 1) at sites 1 and 3 then we can fill the place of them with randomly generated duration as in Figure 10.

As it can be seen at Figure 10, 10 is randomly generated and assigned to the cells (2,1) and (1,3) of the tabu list. This assignment will prevent the allocation of weapon 1 to target 3 and weapon 2 to target 1 for the next 10 iteration. The random duration is a uniform iteration count between 1 and *lifespan + 1*. Lifespan is calculated as follows:

TABLE II. LIFE SPAN CALCULATION FOR TABU LIST

```
if ( problem.Dimension() < 10 ){
  lifespan=5*log10(problem.Dimension());
}else{
  lifespan= 8*log10(problem.Dimension());
}
lifespan++;
```

If the problem size is too large, then using the problem dimension as uniform upper bound will tend to prevent the efficient moves that might help to get better solutions. Thus, taking the logarithm of problem dimension will smooth the durations in the tabu list.

At initialization, a random configuration is created for the initial solution. After that TS move operations and aspiration criteria are used together to get better solutions.

C. Simulated Annealing

The annealing technique is to heat a material earlier to impart high energy to it and then cool it down slowly by keeping at each stage a temperature of sufficient duration. The controlled decrease strategy of the temperature ends up with a crystallized solid state. This is the stable state that corresponds to an absolute minimum of energy. The temperature must be decreased gradually and systematically to avoid from the defects. If so, this can lead to an amorphous structure, a metastable condition that equivalent to a local minimum of energy [26].

The behavior of the temperature in annealing seems as the same with the control parameter in optimization. The temperature has a role to guide the algorithm towards the better solutions. This can be done by lowering the temperature gradually in a controlled manner. If the temperature is lowered suddenly, then the algorithm stops with a local minimum [26].

How the thermodynamic balance of a physical system at a given temperature successfully achieved can be simulated by the Metropolis algorithm [27] on the basis of a given configuration. The system is subjected to an elementary proposed modification and this modification is accepted if it improves the objective function of the system more than the current value, on the other hand, if it improves the objective function ΔE, but below the current value, it is also accepted with a probability of $\exp(-\Delta E/T)$. In practice, this condition is realized by drawing a random real number ranging between 0 and 1, and if this random number is lower than or equal to $\exp(-\Delta E/T)$, then, even though the configuration causes a ΔE degradation in the objective function, it is still accepted. When this Metropolis rule of acceptance is observed subsequently, a chain of events is generated whose configuration depends on the one that immediately precedes it. Under these assumptions, when the chain is limitless, it can be shown that the system can reach its thermodynamic balance at the measured temperature which leads us to a Boltzmann distribution of the energy states at this temperature [26].

From this point of view, the function given to the temperature by the Metropolis rule is well understood. When the temperature is high, $\exp(-\Delta E/T)$ approximates to 1. Thus, most of the moves are accepted and the algorithm turns into a simple random walk in the configuration space. On the contrary, when the temperature is too low, $\exp(-\Delta E/T)$ approximates to 0. In this case, the moves that increase the energy are wrongly rejected.

Thus, the algorithm seems like the classical iterative improvement. At an intermediate temperature, the algorithm occasionally approves the transformations that degrade the objective function; hence a chance is given to the system to escape from a local minimum [26]. Upon reaching the thermodynamic equilibrium at a certain temperature, the temperature is reduced "slightly" and a new Markov chain is executed for this new temperature level [26].

In our case following simulated annealing approach [26] is used:

1) Initial Configuration

A random configuration is used as a starting solution.

2) Initial Temperature

The algorithm can be initialized with user configurable initial temperature and geometric low coefficient (α) or it can be calculated as a preliminary step using the following algorithm if these parameters are unknown:

(1) 100 disturbances created randomly and the average ΔE variations evaluated

(2) Initial rate of acceptance τ_0 of degrading perturbations of quality τ_0 = 50% is selected (starting at high temperature) and T_0 is calculated as $T_0 = -\Delta E/ln(\tau_0)$ deduced from $\tau_0 = exp(-\Delta E/T_0)$

3) Acceptance Rule of Metropolis

An elementary proposed modification is accepted if this transformation causes an improvement in the objective function of the system; on the contrary (ΔE > 0), a number r in [0, 1] is drawn randomly, and accept the disturbance if r < exp (–ΔE/T), where T indicates the current temperature.

4) Change in Temperature Stage

It can take place as soon as a better solution is found or if no improvement seen up to 100 trials.

5) Decrease of the Temperature

It can be carried out according to the geometrical law: $T_{k+1} = \alpha * T_k$, where α is usually between 0.8 and 0.9999.

6) Program Termination

It can be done by the stetted start parameters such as time and iteration count or if the current temperature approaches zero. The operations used in the study are random swaps up to three.

D. Variable Neighborhood Search

Variable Neighborhood Search (VNS) is a simple and efficient metaheuristic for combinatorial and global optimization that systematically change the neighborhood within a possibly randomized local search algorithm. Local search methods starts usually by performing a sequence of local changes to the initial solution which gradually improve the value of the objective function each time until no further improvements are found, that is the local optimum. The VNS does not follow a trajectory; rather it explores the distant neighborhoods of the current solution and jumps from there to a new solution if and only if an improvement was made. After reaching the local optimum of the first algorithm used in the neighborhood structure set, the next algorithm or subroutine tries to improve the current solution. This iterative cycle continues until certain stopping criteria are satisfied. In this way, most of the time, favorable characteristics of the current solution will be kept and used to get the next promising solutions [28, 29].

VNS uses the pre-selected neighborhoods structures of size k_{max} and this structure set is donated as N_k where k is from 1 to k_{max}. Hansen and Mladenovic [28] have studied the structure of the VNS and proposed a couple of VNS algorithms one of which presented at Table III. If a deterministic rule is applied to improve the search in VNS neighborhoods, there is possibility to be trapped in cycle. To avoid this drawback, shaking (in step 2a) is introduced and point x' is selected randomly in the algorithm [28].

TABLE III. STEPS OF BASIC VNS [28]

```
Initialization:
     Select the set of neighborhood structures N_k, k=1,…,k_max,
     Find an initial solution x,
     Choose the stopping conditions.
Repeat the following until stopping conditions are met:
      (1)   Set k=1
      (2)   Repeat the following steps until k = k_max
            (a)  Shaking: Generate a point x' at random from the kth
                          neighborhoodof x ( x'∈ N_k(x))
            (b)  Move or not:
                          If this point is better than the incumbent, move there
                          (x = x'), and continue the search with N_1(k=1)
                          otherwise, k = k + 1;
```

The selection of neighborhood structures, how many of them will be used, their order and changing search strategy from one to the other depends on the problem characteristics and it's not very easy to identify. Jarbouia et al. [30] give 5 good examples of neighborhood structures for the location routing problem with multiple capacitated depots and one incapacitated vehicle per depot. These neighborhood structures are namely *insertion*, *swap(interchange)*, *extended Or-opt*, *inverse Or-opt* and *add/drop depot* and they are explained in detail at the article.

At WTA problem, we have used three neighborhood structures sequentially (1) a number of swap or interchange moves, (2) insertion neighborhood and (3) triple swap or interchange moves. While running the experiments, it was seen that the most efficient neighborhood structure that improved the solution was the first one. As mentioned earlier, selection of

the neighborhood structures and their order essentially problem dependent.

V. Experiments and Results

In the literature, there are benchmark problems for well-defined problems such as travelling salesman problem (TSP). The aim of the TSP is the same for everyone: given a list of cities and the distances between each pair of cities, find the shortest possible route by visiting each city exactly once and return to the origin city. For benchmarking purpose of TSP problem, researchers have proposed certain well-defined scenarios *publicly* and every researcher can use this data set to test his/her algorithm. The goal is not changed from scenario to scenario and from researcher to researcher, always find the shortest possible tour. When it comes to WTA domain, there are no benchmark problems in WTA domain as in TSP. This may be considered as normal because the studies in literature are not formulated exactly the same and various formulation assumptions are proposed that come up with different objective function formulations. That is why, every study uses different unpublished scenarios (mostly random) and different objective function formulations. Therefore, almost every study has different solutions. For these reasons, we couldn't compare the results of this study with the aforementioned studies.

As we couldn't find the benchmark scenarios in the literature, we also created random scenarios like our colleagues. In order to compare the effectiveness of the algorithms herein applied to WTA, a benchmark reference solution is needed and for that purpose, the problem is also formulated in mathematics and solved with the commercial General Algebraic Modeling System (GAMS) version 23.5.1. In this work, to the best of our knowledge, for the first time GAMS solutions are used as reference to compare the effectiveness and quality of heuristic algorithms applied to WTA. We claim that this approach can be used by the community for benchmarking. The test is conducted with a Notebook PC that has an Intel Core i3 CPU at 1.13GHz and a 4 GB RAM.

The Combinatorial Optimization Library [24] used here is developed for education of basic algorithms and problems. Some performance issues like speed of solution are sacrificed for the sake of education. But the algorithms and problems can be implemented and compared with each other.

TABLE IV. Sample Wta Solution Summary

Scenario 6	GA	TS	SA	VNS
CPU Time (msec)	1560	530	1529	1388
Optimization Counter	10809	20	37991	33
Depth	N/A	19	3591	29
Fitness	80.7	80.7	80.7	80.7

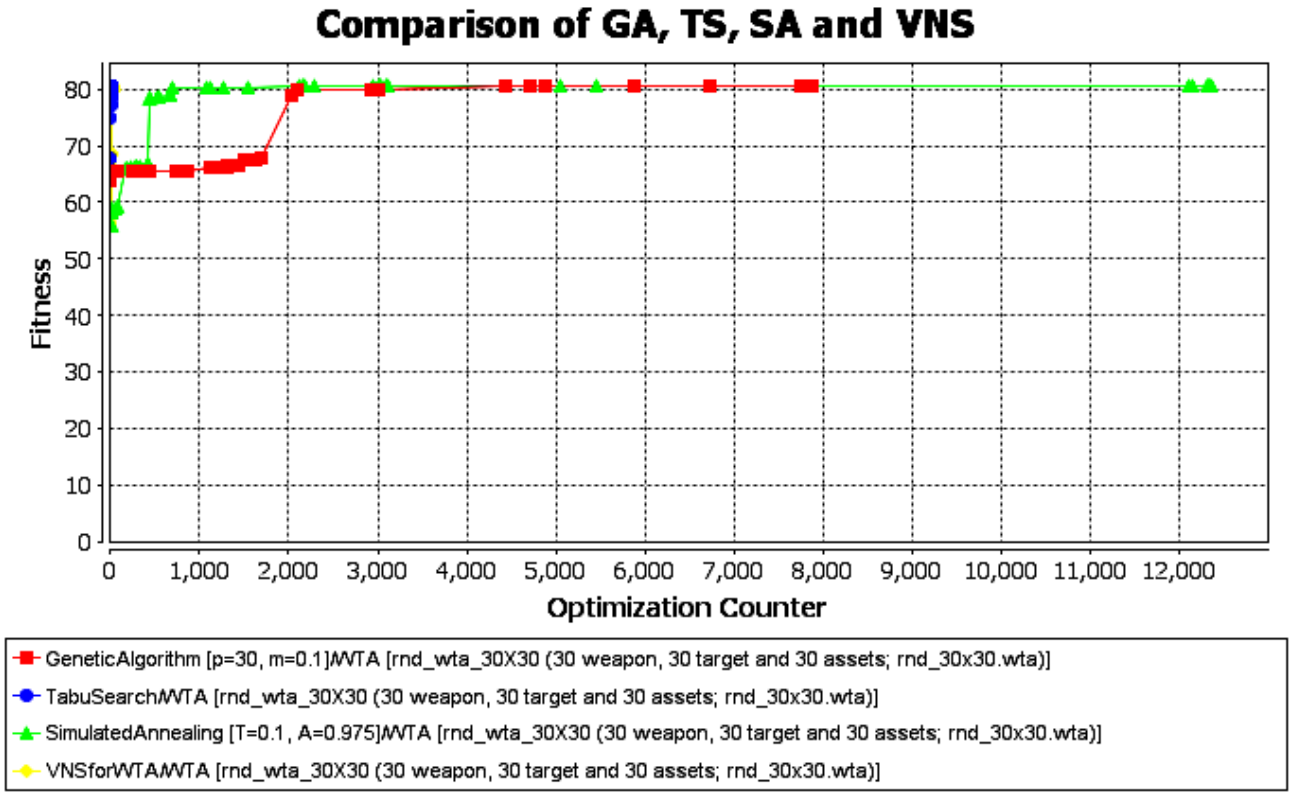


Fig. 11. A Sample (Scenario 6) WTA Solution

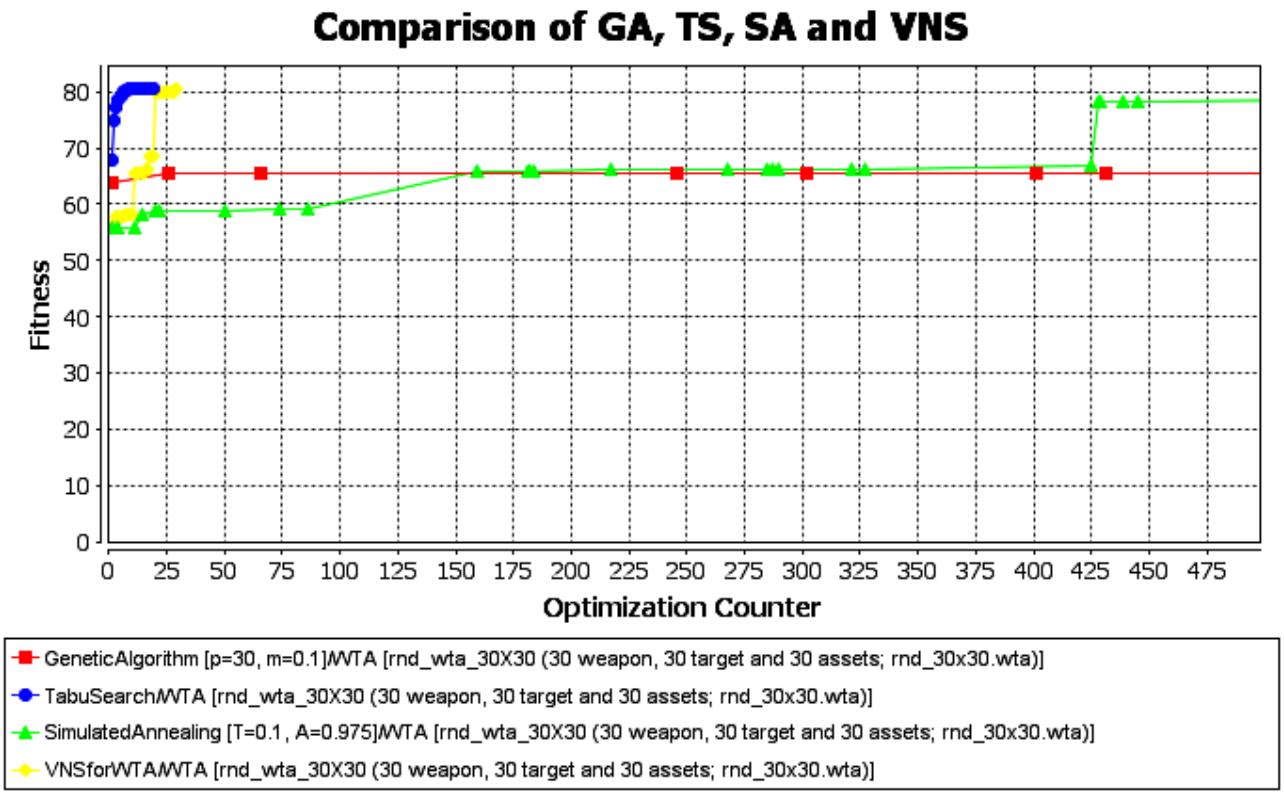


Fig. 12. A Sample (Scenario 6) WTA Solution (zoomed in up to 475 optimization counter)

The algorithms have variable iteration complexity, some of them are very simple and takes less time while others may need more arithmetic operations and exhaust much more time. So the *efficiency* is measured to compare the exhausted time for the same feasible solution, and the *quality* is to compare the feasible solution reached in same time [13].

As it can be seen at the above sample solution (Table IV, Figure 11 and Figure 12), even though the GA, VNS and SA exhausted almost the same CPU time they have quite different optimization/iteration counters. TS has both less optimization/iteration counter and CPU time than all the other three algorithms namely GA, VNS and SA. Even though Figure 11 and Figure 12 show the solution results of the same scenario, Figure 12 is added to discriminate the behavior of the TS and VNS for better perception at lower optimization counter level.

For experiment, 19 random scenarios are created starting from dimension of 5 to 95 with 5 step increments as in the Table IV. Thus these scenarios cover both simple and large-scale instances of the WTA. A moderate WTA combat situation is usually not greater than 30.

TABLE V. WEAPON-TARGET SIZE TO SCENARIO MAPPING

Scenario	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Weapon	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95
Target	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95

These scenarios are solved with GAMES Couenne and Bonmin MINLP solvers and these solutions are used as reference for the solutions of implemented algorithms. GAMS solvers executed the scenarios once while the GA, TS, SA and VNS algorithms tried ten times.

As it can be seen from the minimum and maximum values, the gap is not so large between them. That is why we assumed that ten run is enough for significance. CPU time is recorded as solution time for the algorithms. The Table VI, Figure 13 and Figure 14 summarize the observed time values.

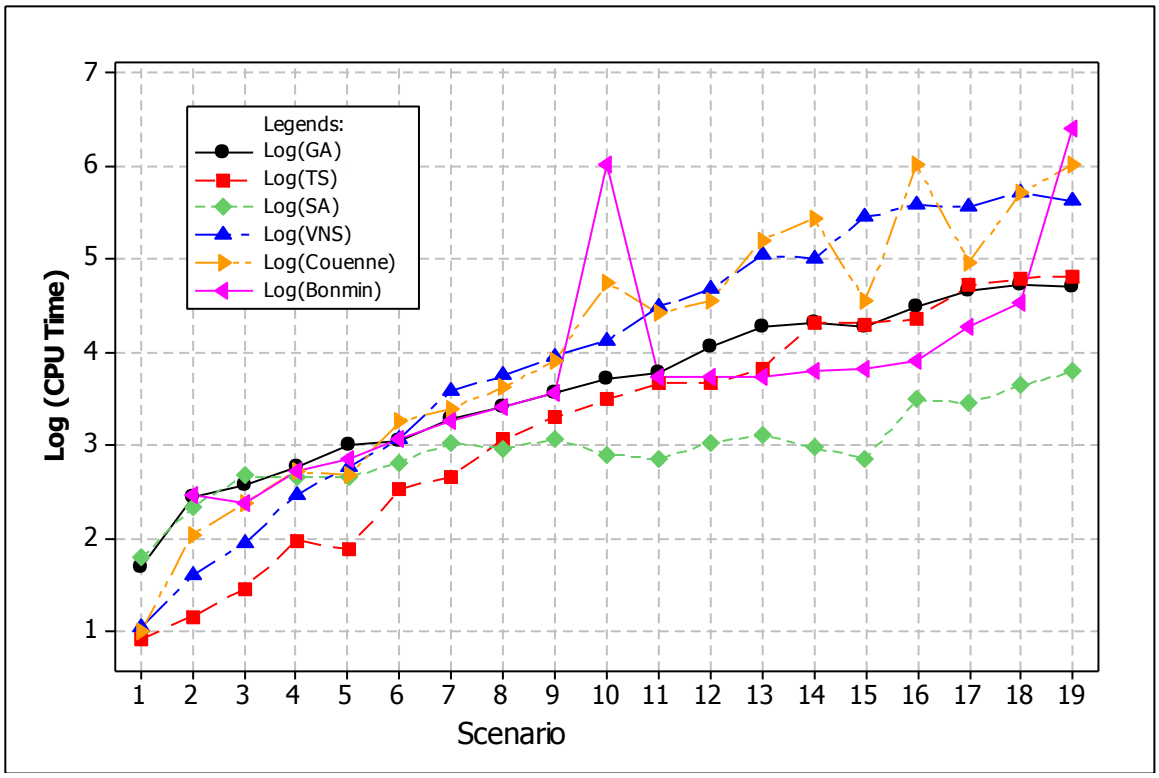


Fig. 13. Time Comparison of Algorithms and GAMS Solvers

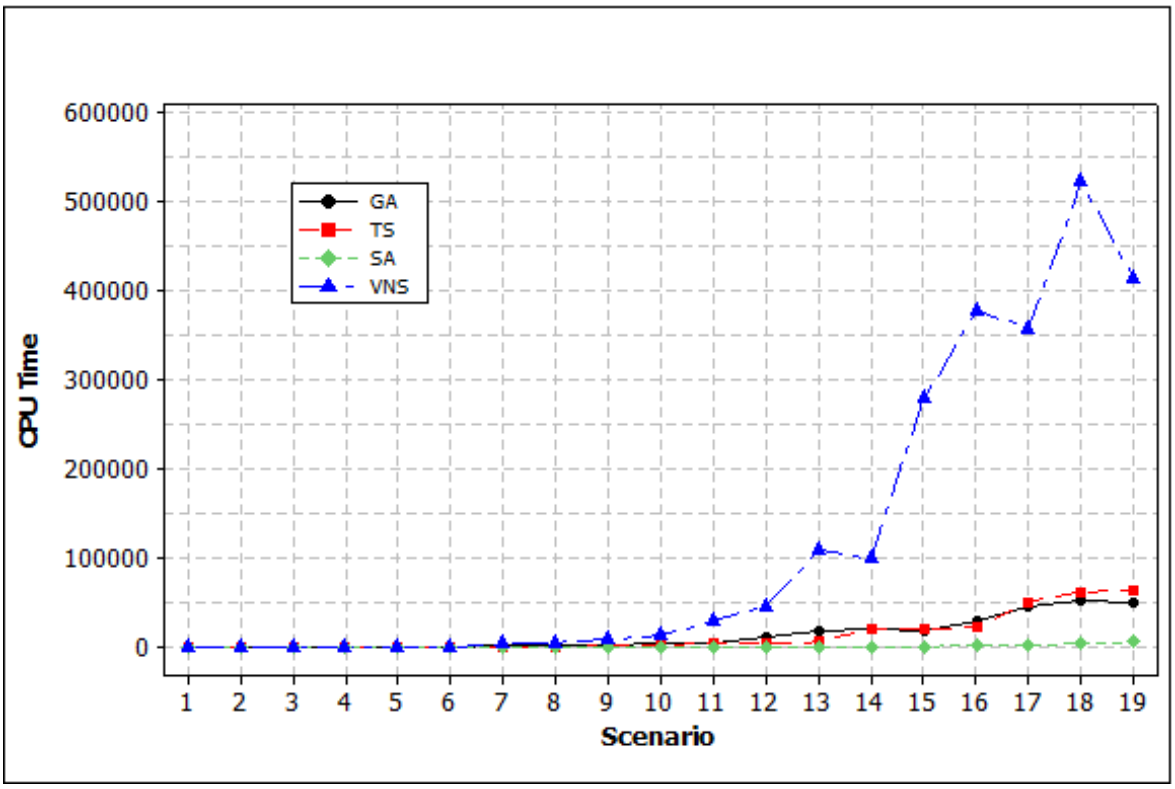


Fig. 14. CPU Time Comparison of Algorithms

TABLE VI. TIME COMPARISON OF SOLUTIONS

S.N.	Average				GAMS	
	GA	*TS*	*SA*	*VNS*	*Couenne*	*Bonmin*
1	48.3	48.3	61.1	11.1000	10	0
2	279.3	279.3	216.8	39	109	281
3	373	373	463.4	88.7000	234	234
4	563.3	563.3	444.7	290	515	515
5	971.9	971.9	441.6	575.7	464.86	717
6	1106	1106	647.1	1176.2	1794	1154
7	1923.3	1923.3	1057.8	3870.2	2481	1779
8	2572.6	2572.6	915.6	5534.6	4087	2605
9	3524.3	3524.3	1148.2	8839.1	8159	3604
10	4998.3	4998.3	792.6	13026	55271	1012670
11	5948	5948	695.8	30170.8	26333	5366
12	11261.6	11261.6	1067	46541.2	34866	5367
12	18561.1	18561.1	1266.7	109367.3	157171	5382
14	20127.4	20127.4	932.7	99990.6	269320	6271
15	18292.5	18292.5	713	279313.4	34825	6474
16	29904	29904	3071.9	379383.6	1004070	7800
17	44904.7	44904.7	2812.7	358432.5	92087	18517
18	53274.1	53274.1	4311.5	524971.6	506473	34133
19	50031.2	50031.2	6084.1	414502.5	1022700	2446890

TABLE VII. FITNESS OF GAMS SOLUTIONS

S.N.	GAMS Found Fitness		GAMS Best Possible Fitness	
	Couenne	*Bonmin*	*Couenne*	*Bonmin*
1	204.0261	204.0261	204.0261	204.0261
2	289.7098	289.7098	289.7098	289.7098
3	128.5998	128.5998	128.5998	128.5998
4	119.6860	120.1146	120.4492	120.1146
5	58.5471	58.9760	59.0120	58.9760
6	80.6784	80.1635	80.8405	80.1635
7	33.7837	33.6791	33.8938	33.6791
8	13.0186	13.0687	13.0856	13.0687
9	21.3787	21.3802	21.6859	21.3802
10	5.6773	5.4292	5.9119	6.3052
11	1.7663	1.7831	1.9030	1.7831
12	6.9788	6.9673	7.2542	6.9673
12	2.6835	2.5334	2.7313	2.5334
14	0.6800	0.6809	0.6915	0.6809
15	0.2573	0.2608	0.2783	0.2608
16	0.2798	0.2805	0.3171	0.2805
17	0.3315	0.3281	0.3424	0.3281
18	0.3523	0.3539	0.3870	0.3539
19	0.0422	0.0424	0.0538	0.0424

When Table VI is examined, GAMS Bonmin has 2 spikes at the scenarios 10 and 19 while GAMS Couenne has 3 spikes at the scenarios 16, 18 and 19. Because of these spikes, when the CPU times are plotted, the distinction among the solutions can't be discriminated easily. Therefore, in order to get a better visualization, we plotted the logarithm of the CPU times at Figure 13. The CPU times of the solutions of the algorithms, but GAMS solutions, are also plotted as in Figure 14.

When we compare the GA, TS, SA and VNS solution times, the CPU time values are close to each other up to scenario 10 and after that, there are steep slopes for the GA, TS and especially for VNS. SA solution time goes on a very low slope and it can be easily seen that SA outperforms these algorithms when the problem dimension gets bigger.

On the other hand, SA cooling down needs special care especially when the problem dimension gets bigger in order to converge to the global optimum. For cooling down in the trials up to scenario 15, $\alpha = 0.95$ is used while at scenarios 16, 17 and 18, $\alpha = 0.975$ and at scenario 19, $\alpha = 0.985$ are used.

TABLE VIII. GENETIC ALGORITHM FITNESS COMPARISON

S.N.	Fitness			Best Found GAMS Fitness
	Minimum	*Average*	*Maximum*	*Approximation %*
1	204.0261	204.0261	204.0261	100.000
2	289.7098	289.7098	289.7098	100.000
3	128.5998	128.5998	128.5998	100.000
4	119.8600	119.6903	120.1146	99.647
5	58.9760	58.9760	58.9760	100.000
6	80.6788	80.6788	80.6788	100.001
7	33.7987	33.8114	33.8146	100.082
8	13.0041	13.0235	13.0687	99.654
9	21.3830	21.3830	21.3830	100.013
10	5.7034	5.7074	5.7078	100.530
11	1.7754	1.7808	1.7832	99.871
12	6.9693	6.9785	6.9800	99.996
12	2.4192	2.6690	2.6968	99.460
14	0.6560	0.6791	0.6819	99.744
15	0.2627	0.2627	0.2630	100.708
16	0.2798	0.2804	0.2805	99.970
17	0.3334	0.3334	0.3334	100.568
18	0.3553	0.3557	0.3559	100.521
19	0.0427	0.0427	0.0427	100.798

TABLE IX. TABU SEARCH FITNESS COMPARISON

S.N.	Fitness			Best Found GAMS Fitness
	Minimum	*Average*	*Maximum*	*Approximation %*
1	204.0261	204.0261	204.0261	100.000
2	289.7098	289.7098	289.7098	100.000
3	128.5998	128.5998	128.5998	100.000
4	120.1146	120.1146	120.1146	100.000
5	58.9760	58.9760	58.9760	100.000
6	80.6788	80.6788	80.6788	100.001
7	33.7987	33.7987	33.7987	100.044
8	13.0041	13.0041	13.0041	99.505
9	21.3830	21.3830	21.3830	100.013
10	5.7078	5.7078	5.7078	100.537
11	1.7832	1.7832	1.7832	100.008
12	6.9800	6.9800	6.9800	100.017
12	2.6968	2.6968	2.6968	100.497
14	0.6818	0.6818	0.6818	100.134
15	0.2630	0.2630	0.2630	100.825
16	0.2805	0.2805	0.2805	100.008
17	0.3334	0.3334	0.3334	100.568
18	0.3558	0.3558	0.3558	100.531
19	0.0427	0.0427	0.0427	100.812

When the GA, TS and SA algorithms compared with GAMS solutions, they had reasonably better temporal solutions as it can be seen at Figure 13. On the other hand, VNS has almost the same time values with the GAMS solutions as it tries three local search techniques sequentially which takes more time. Even with a single neighborhood structure for VNS, we had good fitness results with less time. Thus for this kind of WTA problem, less neighborhood structure can be applied that will end with good results.

The optimization/fitness values of the GAMS models are given at Table VII. These values are used as reference values for the solutions of the algorithms. GA, TS, SA, and VNS solution fitness values are compared with the GAMS models and found that they converge to the global optimum at most trials and the approximation results are promising as can be seen at Table VIII, IX, X and XI. The VNS algorithm produced better results than the other three algorithms and GAMS models. So for the non-real time constrained problems, VNS may give promising solution results. But as mentioned before, construction of the VNS topology requires special care.

TABLE X. SIMULATED ANNEALING FITNESS COMPARISON

S.N.	Fitness			Best Found GAMS Fitness
	Minimum	*Average*	*Maximum*	*Approximation %*
1	204.0261	204.0261	204.0261	100.000
2	289.7098	289.7098	289.7098	100.000
3	128.5998	128.5998	128.5998	100.000
4	120.1146	120.1146	120.1146	100.000
5	58.9760	58.9760	58.9760	100.000
6	80.6788	80.6788	80.6788	100.001
7	33.8146	33.8146	33.8146	100.092
8	13.0041	13.0429	13.0687	99.802
9	21.3830	21.3830	21.3830	100.013
10	5.7074	5.7077	5.7078	100.536
11	1.7794	1.7822	1.7831	99.954
12	6.9492	6.9742	6.9800	99.934
12	2.4163	2.6674	2.6967	99.400
14	0.6783	0.6805	0.6817	99.945
15	0.2532	0.2575	0.2615	98.716
16	0.2791	0.2801	0.2804	99.874
17	0.3291	0.3323	0.3333	100.248
18	0.3536	0.3554	0.3557	100.429
19	0.0426	0.0426	0.0427	100.636

TABLE XI. VNS FITNESS COMPARISON

S.N.	Fitness			Best Found GAMS Fitness
	Minimum	*Average*	*Maximum*	*Approximation %*
1	204.0261	204.0261	204.0261	100.000
2	289.7098	289.7098	289.7098	100.000
3	128.5998	128.5998	128.5998	100.000
4	120.1146	120.1146	120.1146	100.000
5	58.9760	58.9760	58.9760	100.000
6	80.6788	80.6788	80.6788	100.001
7	33.8146	33.8146	33.8146	100.092
8	13.0687	13.0687	13.0687	100.000
9	21.3830	21.3830	21.3830	100.013
10	5.7078	5.7078	5.7078	100.537
11	1.7832	1.7832	1.7832	100.009
12	6.9800	6.9800	6.9800	100.018
12	2.6968	2.6968	2.6968	100.497
14	0.6819	0.6819	0.6819	100.150
15	0.2630	0.2630	0.2630	100.834
16	0.2805	0.2805	0.2805	100.008
17	0.3334	0.3334	0.3334	100.568
18	0.3559	0.3559	0.3559	100.554
19	0.0427	0.0427	0.0427	100.812

VI. CONCLUSIONS

For the reasons explained in the experiments and results section, we couldn't compare our results with the studies in the literature.

That is why we created random scenarios from simple to hard ones and solved them with the proposed algorithms. In order to verify and validate the correctness of the solutions, we also solved the same scenarios with GAMS software. Our main goal is to investigate the most efficient algorithm that solves WTA problem in a reasonable time interval. The exact solvers can also solve the problems optimally but usually it takes too much time to use them in the military domain.

The NP-Complete WTA problem [3] is hard to solve and different heuristics were used to solve them in the past. After the advances in the computing power at computer technology, the algorithms that need huge computer power now can be applied to solve WTA problems. This study is made to measure the efficiency and quality of GA, TS, SA, and VNS algorithms that applied to WTA problem.

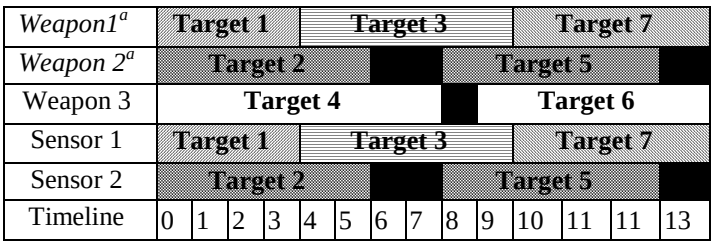
The quality of algorithms applied to WTA problem seems very significant while the efficiency may need to be improved considering the problem dimension. For small size WTA problems, all the four algorithms are promising, but for the large size WTA problems, only SA is good. The VNS algorithm applied here with three neighborhood structure took significant time to solve the large sized WTA problem. A hybrid algorithm that start with SA and continue with the TS and VNS will probably produce good results that meet the temporal and objective constraints, which are worth to try.

The efficiency, speed of solution, is ignored during the implementation of the algorithms. A better coding of the algorithms might produce much better results that meet the temporal constraints also.

VII. FUTURE WORK

WTA is usually considered as part of weapon assignment and sensor allocation scheduling that assessed together in military fields and it is a challenging problem. In this schedule, the factors such as weapon and sensor ranges, weapon and sensor blind zones, kill probabilities, weapon and sensor availability times, ammunition supply have to be considered and after that an engagement must be scheduled to annihilate the enemy targets and defend friendly assets. A simple example plan can be seen at Figure 15.

The static or dynamic multi-staged defense strategies usually don't produce realistic engagement schedules as they assume that all the engagements are simultaneous or weapons are fired all together which is not common in reality. In future, realistic WASA schedules will be studied.



[a] Semi-active weapon system which is guided by a sensor.

Fig. 15. A Sample WASA engagement schedule

REFERENCES

[1] Toet Alexander and Waard Huub de (1995). The Weapon Target Assignment Problem. CALMA Report CALMA.TNO.WP31.AT.95c.

[2] Xin Bin, Chen Jie, Zhang Juan, Dou Lihua and Peng Zhihong (2010). "Efficient Decision Makings for Dynamic Weapon-Target Assignment by Virtual Permutation and Tabu Search Heuristics, IEEE Transactions On Systems, Man, And Cybernetics—Part C: Applications and Reviews, 40(6):649-662.

[3] Lloyd S.P. and Witsenhausen H.S (1986). Weapon allocation is NP-complete. IEEE Summer Simulation Conference, Reno; NV (USA); 28-30 July 1986. pp. 1054-1058.

[4] Hosein, Patrick A and Athans, Michael (1989). Dynamic Weapon-Target Assignment Problem. Symposium on C2 Research, Washington, DC. [online]. Available at: http://www.dtic.mil/dtic/tr/fulltext/u2/a210442.pdf.

[5] Hosein, Patrick A and Athans, Michael (1990). Some Analytical Results for the Dynamic Weapon Target Allocation Problem. February 1990. [online]. Available at: https://dspace.mit.edu/bitstream/1721.1/3179/1/P-1944-21258906.pdf.

[6] Cheng CH and Mon DL, (1994). Evaluating Weapon System by Analytical Hierarchy Process Based on Fuzzy Scales. Journal of Fuzzy Sets and Systems, 63:1-10.

[7] Ahuja, R. K., A. Kumar, K. Jha and J. B. Orlin (2003). Exact and Heuristic Methods for the Weapon Target Assignment Problem. MIT Sloan School of Management, Working Paper No 4464-03.

[8] Owechko, Y and Shams, S,(1994). Comparison of Neural Network and Genetic Algorithms for a resource allocation problem. Neural Networks. IEEE World Congress on Computational Intelligence, 7:4655-4660.

[9] Naidoo, Shahen (2008). Assignment to Resource Allocation for Emergency Response – A Literature Survey. [online]. Available at: http://www.orssa.org.za/wiki/uploads/Johannesburg/Naidoo2008.pdf.

[10] Johansson, Fredrik and Falkman, Göran (2011). Real-time Allocation of Firing Units To Hostile Targets. Journal of Advances in Information Fusion, 6(2):187-199.

[11] [11] Lee Zne-Jung, Su Shun-Feng, and Lee Chou-Yuan (2002). A Genetic Algorithm with Domain Knowledge for Weapon-Target Assignment Problems. Journal of the Chinese Institute of Engineers, 25(3):287-295.

[12] Lee Zne-Jung, Su Shun-Feng, and Lee Chou-Yuan (2003). Efficiently Solving General Weapon-Target Assignment Problem by Genetic Algorithms with Greedy Eugenics. IEEE Transactions On Systems, Man, and Cybernetics—Part B: Cybernetics, 33(1):113-121.

[13] Lu Houqing, Zhang Hongjun, Zhang Xiaojuan and Han Ruixin (2006). An Improved Genetic Algorithm for Target Assignment, Optimization of Naval Fleet Air Defense. 6th World Congress on Intelligent Control and Automation, Dalian, China, pp. 3401-3405.

[14] Shang Gao, Zaiyue Zhang, Xiaoru Zhang and Cungen Cao (2007). Immune Genetic Algorithm for Weapon-Target Assignment Problem. Workshop on Intelligent Information Technology Application, pp.145-148.

[15] [15] Dou Jihua, Yang Xingbao and Lu Yonghong (2009). Improved Genetic Algorithm For Multichannel Ship-To-Air Missile Weapon System WTA Problem. 2nd IEEE International Conference on Computer Science and Information Technology, pp.210-214.

[16] Li Peng, Wu Ling and Lu Faxing (2009). A Mutation-Based GA for Weapon-Target Allocation Problem Subject to Spatial Constraints. International Workshop on Intelligent Systems and Applications, pp.1-4.

[17] Zhihua Song, Fashun Zhu and Duolin Zhang (2009). A heuristic genetic algorithm for solving constrained Weapon-Target Assignment problem. Intelligent Computing and Intelligent Systems, 1:336-341.

[18] Roux JN and Vuuren JH van (2007). Threat evaluation and weapon assignment decision support: A review of the state of the art. ORION, 23(2):151–187.

[19] Karasakal, Orhan (2008). Air defense missile-target allocation models for a naval task group. Computers & Operations Research, 35(6):1759–1770.

[20] Cai H., Liu J., Chen Y., and Wang H. (2006). Survey of the research on dynamic weapon-target assignment problem. Journal of Systems Engineering and Electronics, 17(3):559–565.

[21] Li J., Cong R., and Xiong J. (2006). Dynamic WTA optimization model of air defense operation of warships' formation. Journal of Systems Engineering and Electronics, 7(1):126–131.

[22] Blodgett D., Gendreau M., Guertin F., and Potvin J. Y. (2003). A tabu search heuristic for resource management in naval warfare. Journal of Heuristics, 9:145–169.

[23] Wu L., Xing C., Lu F., and Jia P. (2008). An anytime algorithm applied to dynamic weapon-target allocation problem with decreasing weapons and targets. IEEE Congress on Evolutionary Computation, Hong Kong, China, pp. 3755–3759.

[24] Skalicka Ondrej (2010). Combinatorial Optimization Library, Master's Thesis, Czech Technical University in Prague.

[25] Sivanandam S.N. and Deepa S.N.(2008). Introduction to Genetic Algorithms. Springer. ISBN9783540731900.

[26] Dréo Johann, Siarry Patrick, Pétrowski Alain and Taillard Eric (2006). Metaheuristics for Hard Optimization. Springer. ISBN-139783540230229.

[27] Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., and Teller E.(1953). Equation of state calculation by fast computing machines. Journal of Chemical Physics, 21(6):1087–1092.

[28] Hansen, Pierce and Mladenovic, Nenad (2001). Variable Neighborhood Search: Principles and applications. European Journal of Operational Research, 130:449-467.

[29] Mladenovic, Nenad and Hansen, Pierce (1997). Variable Neighborhood Search. Computers & Operations Research, 24(11):1097-1100.

[30] Jarbouia B, Derbela H, Hanafic S and Mladenović N (2013). Variable neighborhood search for location routing. Computers & Operations Research, 40(1):47–57.