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

IJARAI

International Journal of Advanced Research in Artificial Intelligence

Volume 5 Issue 4

www.ijarai.thesai.org

A Publication of
The Science and Information Organization

Editorial Preface

From the Desk of Managing Editor…

Artificial Intelligence is hardly a new idea. Human likenesses, with the ability to act as human, dates back to Geek mythology with Pygmalion's ivory statue or the bronze robot of Hephaestus. However, with innovations in the technological world, AI is undergoing a renaissance that is giving way to new channels of creativity.

The study and pursuit of creating artificial intelligence is more than designing a system that can beat grand masters at chess or win endless rounds of Jeopardy!. Instead, the journey of discovery has more real-life applications than could be expected. While it may seem like it is out of a science fiction novel, work in the field of AI can be used to perfect face recognition software or be used to design a fully functioning neural network.

At the International Journal of Advanced Research in Artificial Intelligence, we strive to disseminate proposals for new ways of looking at problems related to AI. This includes being able to provide demonstrations of effectiveness in this field. We also look for papers that have real-life applications complete with descriptions of scenarios, solutions, and in-depth evaluations of the techniques being utilized.

Our mission is to be one of the most respected publications in the field and engage in the ubiquitous spread of knowledge with effectiveness to a wide audience. It is why all of articles are open access and available view at any time.

IJARAI strives to include articles of both research and innovative applications of AI from all over the world. It is our goal to bring together researchers, professors, and students to share ideas, problems, and solution relating to artificial intelligence and application with its convergence strategies. We would like to express our gratitude to all authors, whose research results have been published in our journal, as well as our referees for their in-depth evaluations.

We hope that this journal will inspire and educate. For those who may be enticed to submit papers, thank you for sharing your wisdom.

Editor-in-Chief
IJARAI
Volume 5 Issue 4 April 2016
ISSN: 2165-4069(Online)
ISSN: 2165-4050(Print)

Editorial Board

Reviewer Board Members

CONTENTS

One of the Possible Causes for Diatom Appearance in Ariake Bay Area in Japan In the Winter from 2010 to 2015 (Clarified with AQUA/MODIS)

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

***Abstract*—One of the possible causes for diatom appearance in Ariake bay area I Japan in the winter seasons from 2010 to 2015 is clarified with AQUA/MODIS of remote sensing satellite. Two months (January and February) AQUA/MODIS derived chlorophyll-a concentration are used for analysis of diatom appearance. Match-up data of AQUA/MODIS with the evidence of the diatom appearance is extracted from the MODIS database. Through experiments, it is found that diatom appears after a long period time of relatively small size of red tide appearance. Also, it depends on the weather conditions and tidal effect as well as water current in the bay area in particular.**

Keywords—chlorophyl-a concentration; red tide; diatom; MODIS; satellite remote sensing

I. INTRODUCTION

The Ariake Sea is the largest productive area of Nori (Porphyra yezoensis[1]) in Japan. In winters in 2012, 2013, 2014 and 2015, a massive diatom bloom appeared in the Ariake Bay, Japan [1]. In case of above red tides, bloom causative was Eucampia zodiacus[2]. This bloom has being occurred several coastal areas in Japan and is well reported by Nishikawa et al. for Harimanada sea areas [2]-[10]. Diatom blooms have recurrently appeared from late autumn to early spring in the coastal waters of western Japan, such as the Ariake Bay [11] and the Seto Inland Sea [12], where large scale "Nori" aquaculture occurs. Diatom blooms have caused the exhaustion of nutrients in the water column during the "Nori" harvest season. The resultant lack of nutrients has suppressed the growth of "Nori" and lowered the quality of "Nori" products due to bleaching with the damage of the order of billions of yen [3].

In particular in winter since 2012, almost every year, relatively large size of diatoms of *Eucampia zodiacus* appears in Ariake Bay areas. That is one of the causes for damage of *Porphyra yezoensis*. There is, therefore, a strong demand to prevent the damage from Nori farmers. Since 2007, *Asteroplanus karianus* appears in the Ariake Bay almost every year. In addition, *Eucampia zodiacus* appears in Ariake Bay since 2012. Meanwhile, *Eucampia zodiacus* did not appeared in 2011, 2010. Therefore, there is a key for the diatom appearance. By comparing Ariake Bay situations in winter seasons in 2010 and 2011 and after 2012, it might be possible to find out possible causes for diatom appearance.

The chlorophyll-a concentration algorithm developed for MODIS[3] has been validated [13]. The algorithm is applied to MODIS data for a trend analysis of chlorophyll-a distribution in the Ariake Bay in the winter from 2010 to 2015 is made [14]. Also, locality of red tide appearance in Ariake Sea including Ariake Bay, Isahaya Bay and Kumamoto offshore is clarified by using MODIS data derived chlorophyll-a concentration [15]. On the other hand, red tide appearance (location, red tide species, the number of cells in unit water volume by using microscopy) are measured from the research vessel of the Saga Prefectural Fishery Promotion Center: SPFPC by once a 10 days. The location and size of the red tide appearance together with the red tide source would be clarified by using SPFPC data. Match-up data of MODIS derived chlorophyll-a concentration is used for investigation of relations between MODIS data and truth data of the red tide appearance. Through time series data analysis of MODIS derived chlorophyll-a concentration, one of the possible causes of diatom appearance is clarified with the evidence of research Bessel observations.

In the next section, the method and procedure of the experimental study is described followed by experimental data and estimated results. Then conclusion is described with some discussions.

II. EXPERIMENTAL METHOD AND RESULTS

A. Intensive Study Areas

Fig.1 shows the intensive study areas of Ariake Bay, Kyushu, Japan. Ariake Bay is a portion of Ariake Sea of which the width is around 20km (in direction of east to west) and the length is approximately 100km (in direction of north to south). It is almost closed sea area because the mouth of Ariake Sea is quite narrow. Sea water exchanges are, therefore, very small.

[1] http://en.wikipedia.org/wiki/Porphyra

[2] http://www.eos.ubc.ca/research/phytoplankton/diatoms/centric/eucampia/e_zodiacus.html

[3] http://modis.gsfc.nasa.gov/

Fig. 1. Intensive study areas

B. *MODIS Data Derived Chlorophyll-a Concentration and Truth Data and Truth Data of Red Tide in 2010 to 2015*

MODIS derived chlorophyll-a concentration which area acquired for the observation period of two months (in January and February) in 2010 to 2015 is used for the experiments. On the other hand, Fig.2 shows the example of the superimposed image with MODIS data derived chlorophyll-a concentration and truth data which is provided by Saga Prefectural Fishery Promotion Center on 21 January 2010. The number in the figure denotes the number of red tide cells / ml. Such the number of red tide is reported every 10 days.

It is found the following red tide at around the Shiota river mouth and its surrounding areas on January 21 2010,

Asterionella kariana; 3280 cells/ml

Skeletonema costatum: 1330 cells/ml

On January11 2011, it is found the following red tide along with the Shiroishi town offshore to the Shiota river mouth and its surrounding areas,

Asterionella kariana; 10150 cells/ml

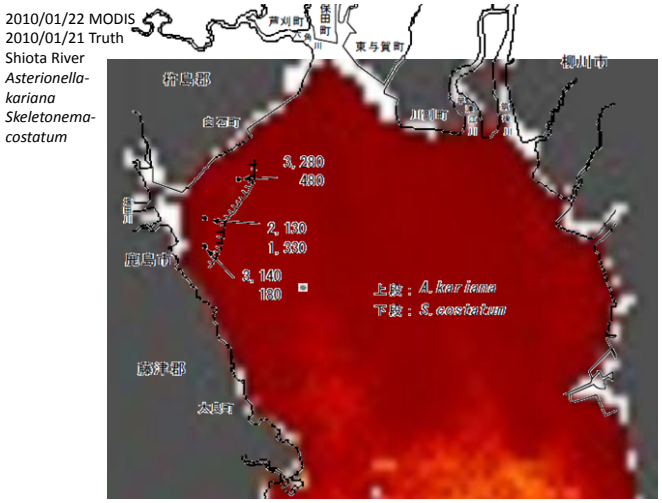


Fig. 2. Example of the superimposed image with MODIS data derived chlorophyll-a concentration and truth data which is provided by Saga Prefectural Fishery Promotion Center

Although the truth data say that the red tide is distributed at around Shiota river mouth and its surrounding areas as well as Shiroishi offshore, it cannot be seen due to the fact that it is covered with cloud in the MODIS data derived chlorophyll-a concentration. It is found the following red tide at around the Kashima offshore on February 25 2011,

Asterionella kariana; 4950 cells/ml

It is found the following red tide at around the Shiota River Mouth and its surrounding areas on December 30 2011,

Asterionella kariana; 5150 cells/ml

On January 23 2012, it is found the following red tide at the Shiroishi offshore,

Skeletonema spp .: 5150 cells/ml

The red tide is distributed at around Shiota river mouth and its surrounding areas as well as Shiroishi offshore.

The following red tide is found widely along with the Kawazoe offshore to the Tara offshore on February 22 2012,

Eucampia zodiacus: 1,090 cells/ml

Also it is found the following red tide along with the Shiota river mouth and its surrounding areas to the Kashima offshore on December 31 2012,

Skeletonema spp .: 6110 cells/ml

On January 7 2013, the following red tide are observed along with the Shiota river mouth and its surrounding areas to the Shiroishi offshore,

Asterionella kariana; 5630 cells/ml

Skeletonema costatum: 3390 cells/ml

The red tide distribution derived from MODIS data is almost coincident to the truth data.

It is found the following red tide at the Shiroishi offshore on January 6 2014,

Asterionella kariana; 4830 cells/ml

The following red tide is observed at the Shiroishi offshore on January 16 2014,

Skeletonema spp .: 6110 cells/ml

Thalassiosira spp.: 1510 cells/ml

On February 6 2014, the following red tide is observed almost whole Ariake bay area except the Shiroishi offshore,

Eucampia zodiacus: 568 cells/ml

It is observed the following red tide along with the Shiroishi offshore to the Tara offshore on December 30 2014,

Asterionella kariana; 3890 cells/ml

Skeletonema costatum: 8750 cells/ml

On March 6 2015, the following red tide is observed along with the Kashima offshore to the Tara offshore,

Eucampia zodiacus: 1310 cells/ml

It is clear that the diatom of *Eucampia zodiac* appeared in the winter in 2012, 2013, 2014 and 2015. The differences between the situations in the time period of 2010, 2011 and the other time period from 2012 to 2015 are (1) relatively small size of red tides, *Asterionella kariana* and *Skeletonema costatum* appeared at around Shiota river mouth and its surrounding areas for a long time period, (2).

C. Chlorophyll-a Concentration Trends in the Different Areas in theWInter of the Different Year

Chlorophyll-a concentration trends in the different areas, Isahaya Bay, Around the Shiota river mouth and its surrounding areas, and the middle of the Ariake Bay are investigated with the MODIS derived chlorophyll-a concentrations acquired on

(2010) January 1, 3, 9, 14, 16, 17, 18, 22, 24, 26, 27, 29, February 3, 4, 5, 6, 20, 21, 23, and 28 in 2010

(2011) January 1, 2, 7, 8, 14, 17, 22, 26, 27, February 1, 3, 4, 15, 21, 22, 24, and 26 in 2011

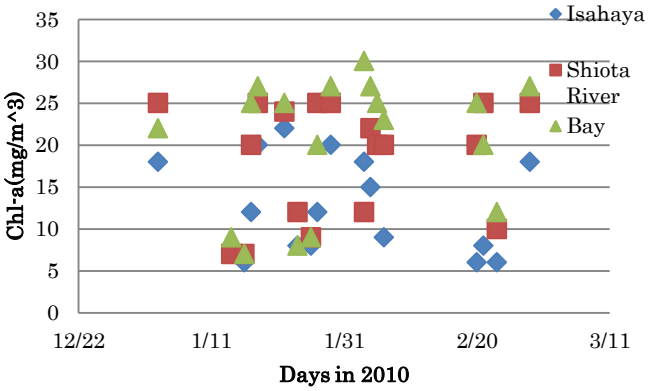
(2012) January 2, 6, 7, 12, 17, 20, 21, 23, 26, 29, 30, 31, February 4, 11, 12, 20, 24, and 29 in 2012

(2013) January 4, 6, 10, 11, 12, 15, 18, 25, 28, 30, 31, February 2, 3, 10, 13, 16, 20, 22, 23, 24, and 29 in 2013

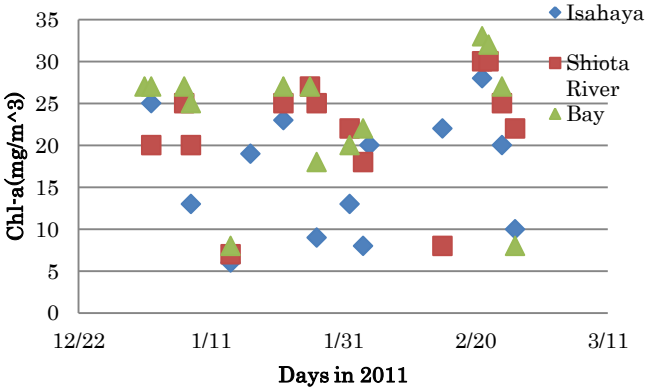
(2014) January 10, 13, 15, 16, 19, 23, 24, 26, 27, 29, 30, February 4, 8, 11, 12, 20, 21, 23, and 24 in 2014

(2015) January 4, 6, 7, 8, 9, 10, 12, 17, 18, 20, 23, February 1, 3, 6, 9, 13, 14, 20, and 27 in 2015

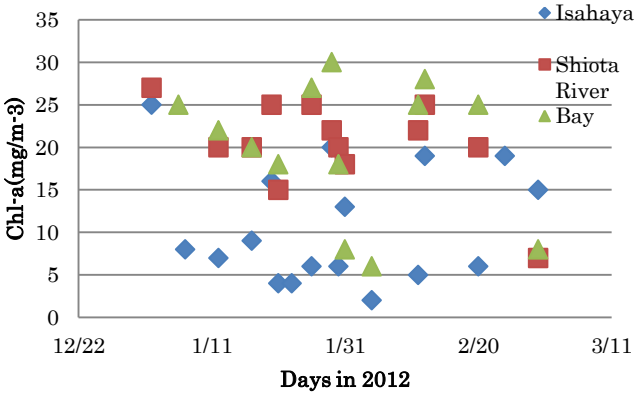
The results from the trend analysis are shown in Fig.3.



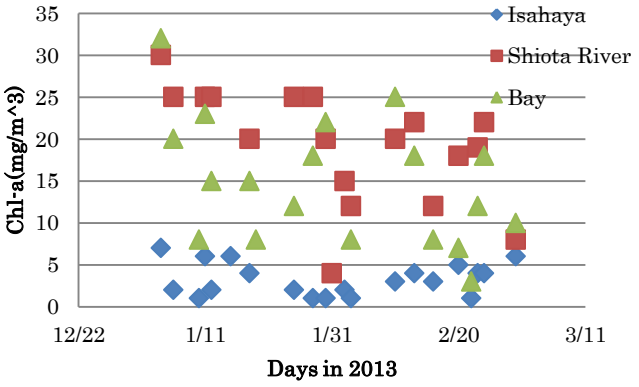
(a)2010



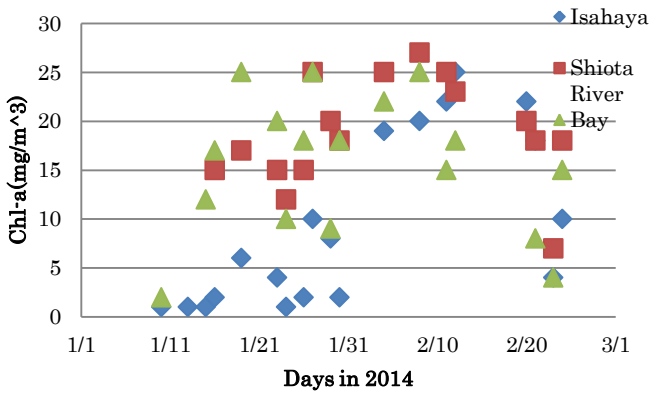
(b)2011



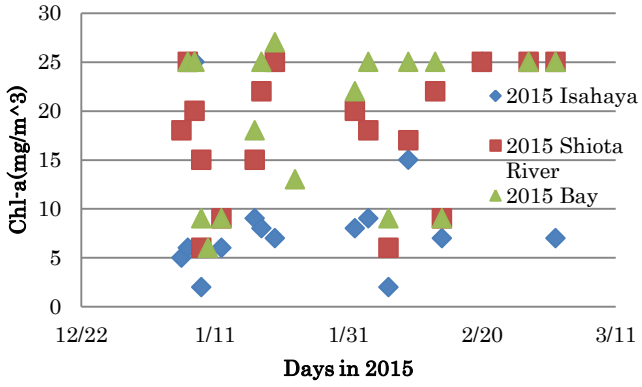
(c)2012



(d)2013



(e)2014



(f)2015

Fig. 3. Trends of chlorophyll-a concentration derived from MODIS data in the winter seasons in 2010-2015

The trends of chlorophyll-a concentrations of Shiota river mouth and its surrounding areas and the middle of Ariake Bay area are quite similar while that of Isahaya Bay area is not resemble to those of the trends of Shiota river mouth and its surrounding area and Ariake Bay area. Therefore, the origin of chlorophyll-a of Isahaya Bay area is different from those of Ariake Bay area as well as Shiota river mouth and its surrounding areas. Not only time series of trend of chlorophyll-a concentrations, but also spatial characteristics of chlorophyll-a concentration distributions between Isahaya Bay and the other Ariake Bay area as well as Shiota river mouth and its surrounding areas are different each other.

Relatively large sized diatom of *Eucampia zodiacus*: appeared in entire Ariake Bay areas in 2012, 2013, 2014, and 2015 and it did not appeare in 2010 and 2011. Comparatively small sized diatom of *Asterionella kariana* and *Skeletonema costatum*, on the other hand, appeared almost every year though. In particular, large sized diatom appeared after small sized diatom appeared for a long time period in the winter seasons in year of 2012, 2013, 2014, and 2015. On the other hand, small sized diatom disappeared in a short time period in the winter seasons in years of 2010 and 2011 results in large sized diatom did not appeared.

D. Relations Between Chlorophyll-a Concentration and the Meteorological Data

It may say that red tide appears when the following conditions are situated, nutrition rich water and rich solar illumination and less wind. Therefore, the relations between chlorophyll-a concentration and rainfall in a day, averaged air-temperature in a day, and the averaged wind speed in a day are investigated together with MODIS data derived 8 day composite of the Photosynthetically Available Radiance: PAR (Einstein/m^2/day). The meteorological data are gotten from the Japanese Meteorological Agency: JMA of Shiroishi Automated Meteorological Data Acquisition System: AMEDAS data[4] which meteorological station is situated at the longitude and latitude of 33:11.0' North and 130:8.9' East. The elevation of the station is 2 m. The results from these analyses are shown in Fig.4. Meanwhile, MODIS data derived PAR is obtained from the NASA/GSFC site [5]. Solar illumination condition can be replaced to PAR.

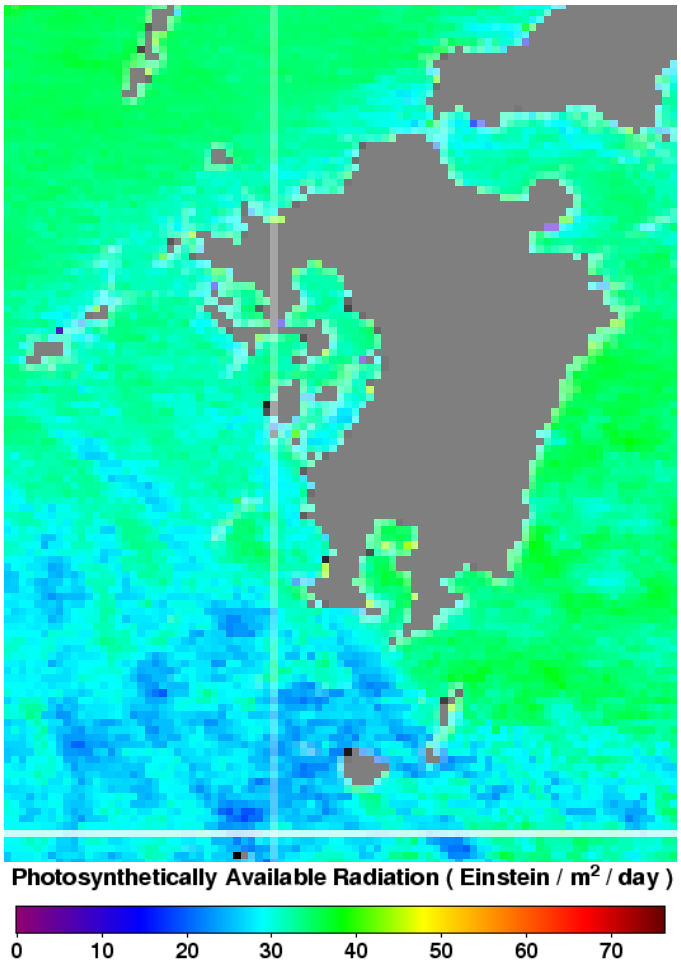


Fig. 5. Example of the 8 day composite of the PAR derived from MODIS data for the period from March 6 to March 14 in 2011

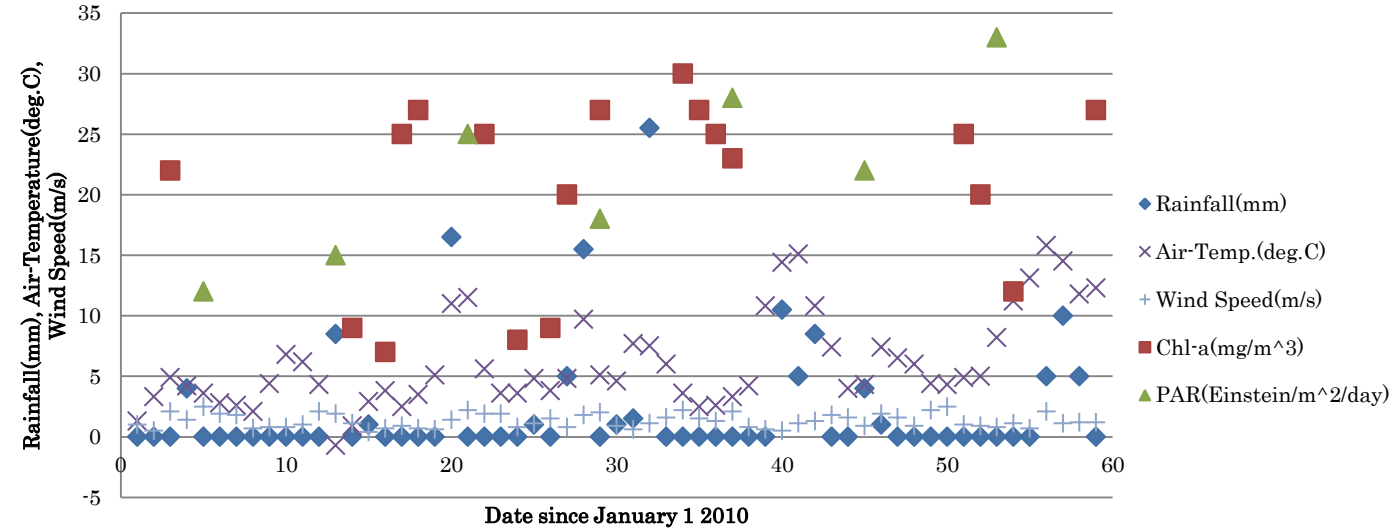
[4] http://www.data.jma.go.jp/obd/stats/etrn/index.php

[5] http://oceancolor.gsfc.nasa.gov/cgi/l3?per=8D&prd=PAR_par&sen=T&res=4km&num=24&ctg=Standard&date=1Jan2014

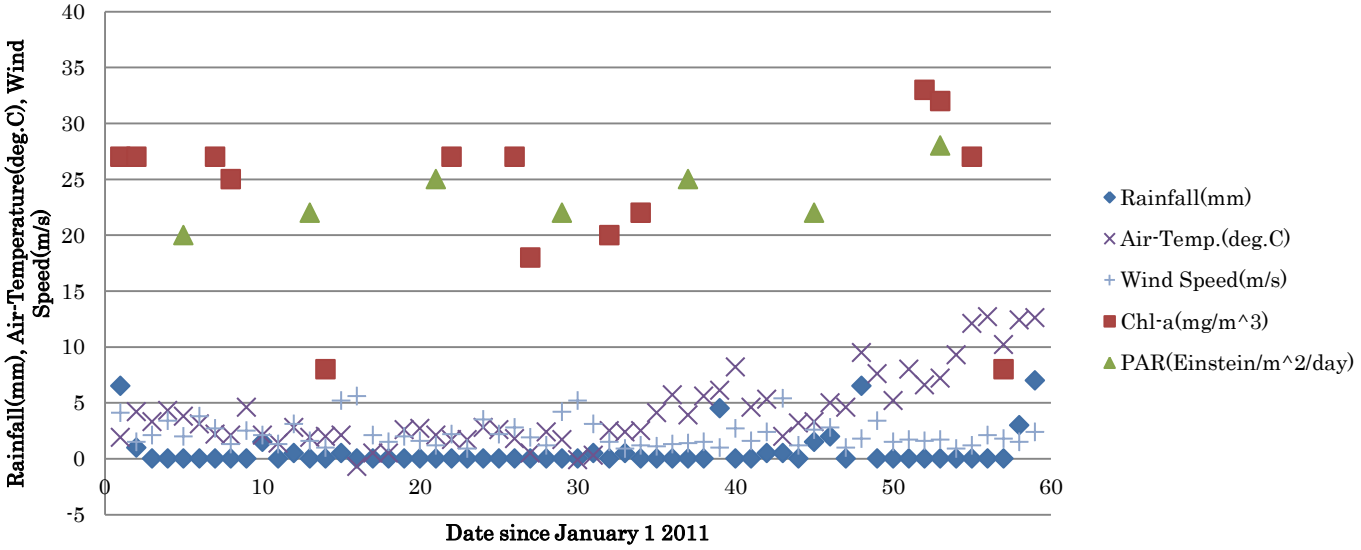
Eucampia zodiacus may grow a wide variety of water temperature and salinity conditions, the most preferable water temperature is around 25 ℃ tough. In order to grow, *Eucampia zodiacus* requires relatively strong solar illumination. It can maintain capability of consume nitrogen even if the water temperature is not so high. Therefore, the conditions of water temperature, solar illumination are key issues for *Eucampia zodiacus*

One the other hand, rainfall provide nutrient to the Bays from the rivers. Therefore, *Eucampia zodiacus* did not appear in 2010 and 2011 because of less of the rainfall while *Eucampia zodiacus* appeared in 2012 and after because of rich of the rainfall.

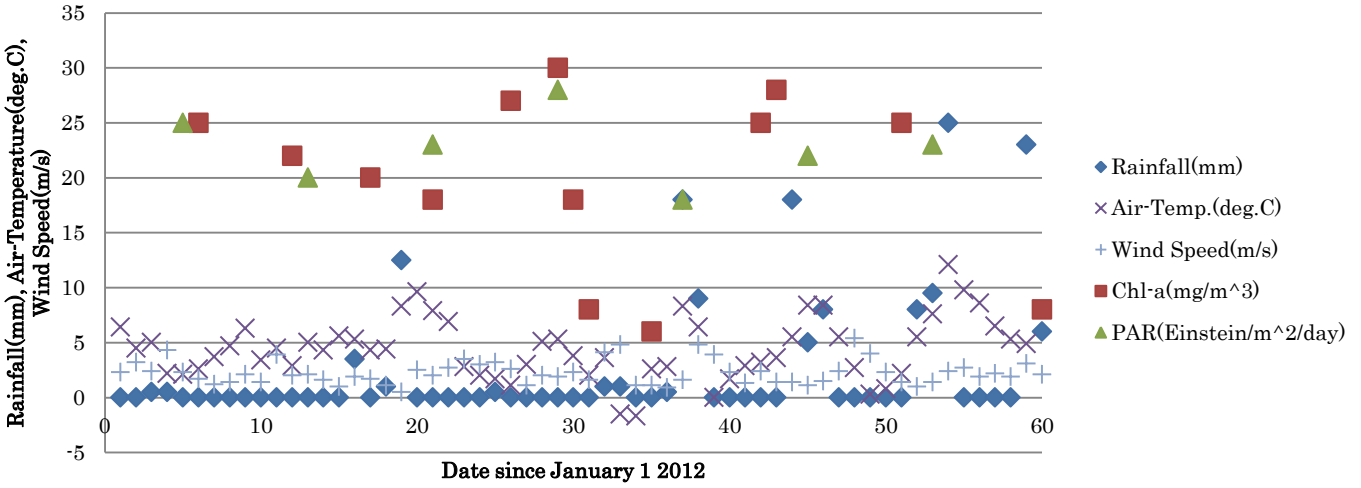
Meanwhile, PAR in the first half of January in both 2010 and 2011 is relatively small in comparison to the PAR in that period in 2012 and after. Therefore, diatom did not appear in both 2010 and 2011.



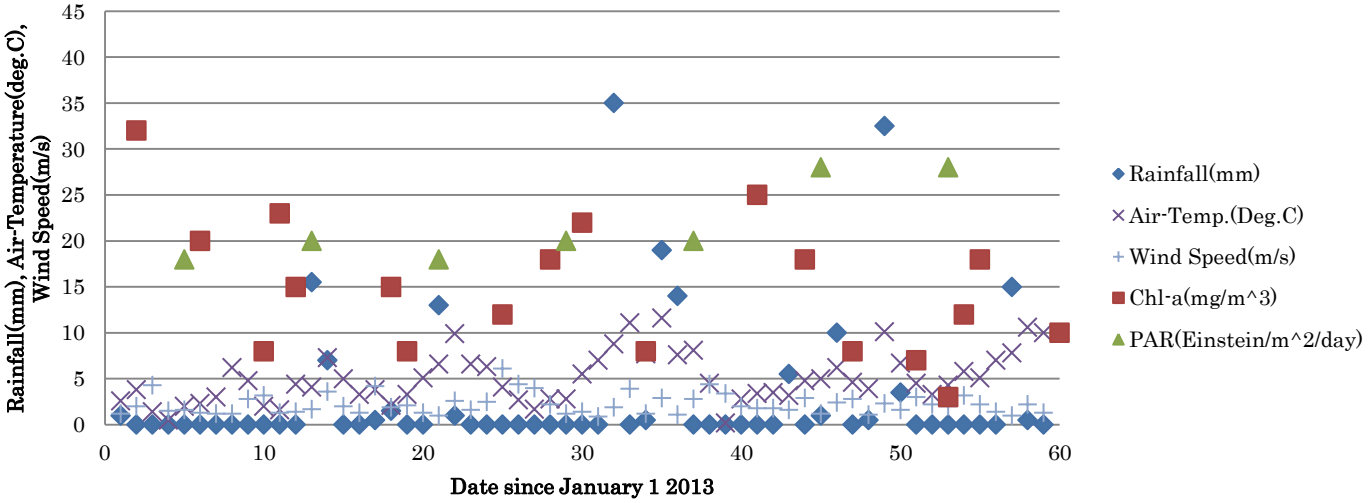
(a)2010



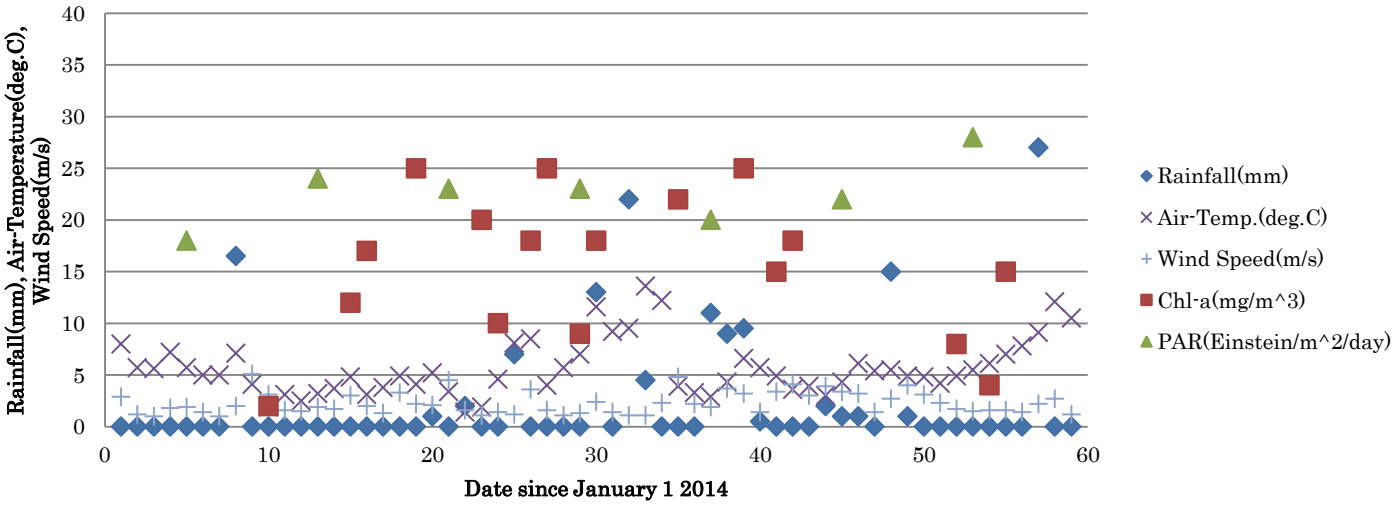
(b)2011



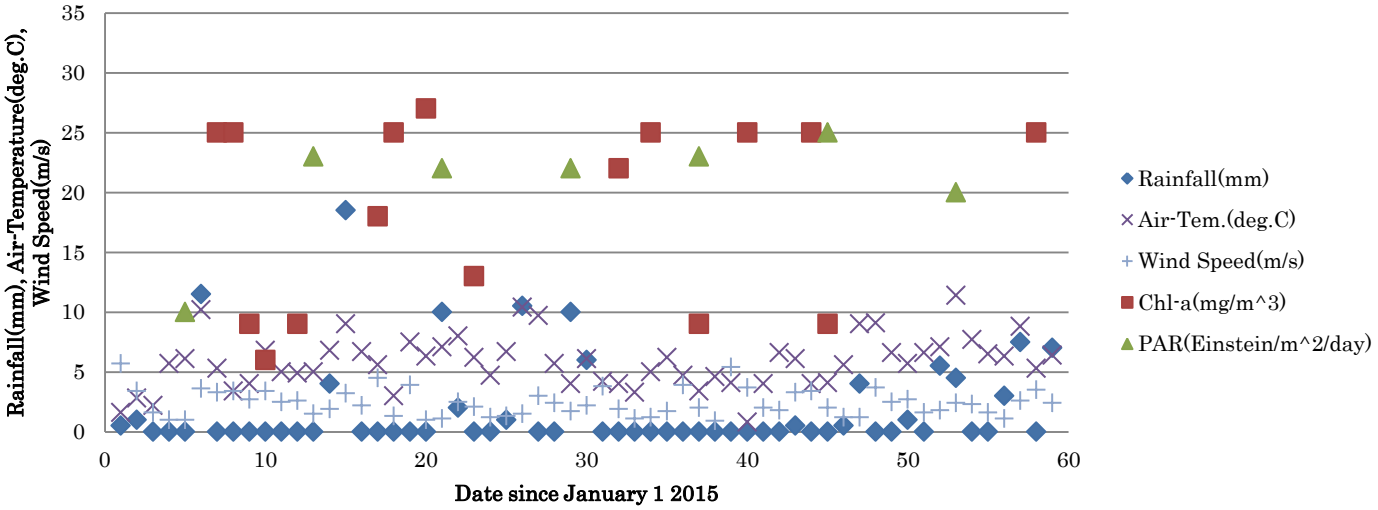
(c)2012



(d)2013



(e)2014



(f)2015

Fig. 4. Trends of meteorological data and chlorophyll-a concentration in the winter seasons of 2010 to 2015

III. Conclusion

Asterionella kariana and *Skeletonema costatum* are used to be appeared in the Ariake Bay area in the winter seasons followed by *Eucampia zodiacus* appearance in the early spring almost every year after 2012 in particular, on February 22 2012, February 26 2013, February 6 2014 and March 6 2015.

Through the trend analysis with the superimposed images of the truth data and the MODIS data derived chlorophyll-a concentration which are acquired in the period starting from February 27 to March 5 2015, it is found that chlorophyll-a is distributed densely in the Ariake bay area and Isahaya bay area on February 27. Then the densely distributed chlorophyll-a is flown to the south direction along with the sea water current in the Ariake bay while the densely distributed chlorophyll-a is flown from the Isahaya bay to the Tara-machi and far beyond the Shimabara offshore. Therefore, it may say that the sources of red tide are different between Ariake bay and Isahaya bay.

Further investigations are required to clarify the mechanism of red tide appearance with the consideration three dimensional of cross section analysis the red tide source movement.

Acknowledgment

The authors would like to thank Dr. Toshiya Katano of Tokyo University of Marine Science and Technology, Dr. Yuichi Hayami, Dr. Kei Kimura, Dr. Kenji Yoshino, Dr. Naoki Fujii and Dr. Takaharu Hamada of Institute of Lowland and Marine Research, Saga University for their great supports through the experiments.

References

[1] Yuji Ito, Toshiya Katano, Naoki Fujii, Masumi Koriyama, Kenji Yoshino, and Yuichi Hayami, Decreases in turbidity neap tides initiate late winter large diatom blooms in a macrotidal embayment, Journal of Oceanography,69: 467-479. 2013.

[2] Nishikawa T (2002) Effects of temperature, salinity and irradiance on the growth of the diatom *Eucampia zodiacus* caused bleaching seaweed *Porphyra* isolated from Harima-Nada, Seto Inland Sea, Japan. Nippon Suisan Gakk 68: 356–361. (in Japanese with English abstract)

[3] Nishikawa T (2007) Occurrence of diatom blooms and damage tocultured *Porphyra* thalli by bleaching. Aquabiology 172: 405–410. (in Japanese with English abstract)

[4] Nishikawa T, Hori Y (2004) Effects of nitrogen, phosphorus and silicon on the growth of the diatom *Eucampia zodiacus* caused bleaching of seaweed *Porphyra* isolated from Harima-Nada, Seto Inland Sea, Japan. Nippon Suisan Gakk 70: 31–38. (in Japanese with English abstract)

[5] Nishikawa T, Hori Y, Nagai S, Miyahara K, Nakamura Y, Harada K, Tanda M, Manabe T, Tada K (2010) Nutrient and phytoplankton dynamics in Harima-Nada, eastern Seto Inland Sea, Japan a 35-year period from 1973 to 2007. Estuaries Coasts 33: 417–427.

[6] Nishikawa T, Hori Y, Tanida K, Imai I (2007) Population dynamics of the harmful diatom *Eucampia zodiacus* Ehrenberg causing bleachings of *Porphyra* thalli in aquaculture in Harima- Nada, the Seto Inland Sea, Japan. Harmful algae 6: 763–773.

[7] Nishikawa T, Miyahara K, Nagai S (2000) Effects of temperature and salinity on the growth of the giant diatom *Coscinodiscus wailesii* isolated from Harima-Nada, Seto Inland Sea, Japan. Nippon Suisan Gakk 66: 993–998. (in Japanese with English abstract)

[8] Nishikawa T, Tarutani K, Yamamoto T (2009) Nitrate and phosphate uptake kinetics of the harmful diatom *Eucampia zodiacus* Ehrenberg, a causative organism in the bleaching of aquacultured *Porphyra* thalii. Harmful algae 8: 513–517.

[9] Nishikawa T, Yamaguchi M (2006) Effect of temperature on lightlimited growth of the harmful diatom *Eucampia zodiacus* Ehrenberg, a causative organism in the discoloration of *Porphyra* thalli. Harmful Algae 5: 141–147.

[10] Nishikawa T, Yamaguchi M (2008) Effect of temperature on lightlimited growth of the harmful diatom *Coscinodiscus wailesii*, a causative organism in the bleaching of aquacultured *Porphyra* thalli. Harmful Algae 7: 561–566.

[11] Syutou T, Matsubara T, Kuno K (2009) Nutrient state and nori aquaculture in Ariake Bay. Aquabiology 181: 168–170. (in Japanese with English abstract)

[12] Harada K, Hori Y, Nishikawa T, Fujiwara T (2009) Relationship between cultured *Porphyra* and nutrients in Harima-Nada, eastern part of the Seto Inland Sea. Aquabiology 181: 146–149. (in Japanese with English abstract)

[13] Arai K., T. Katano, Trend analysis of relatively large diatoms which appear in the intensive study area of the ARIAKE Sea, Japan, in winter (2011-2015) based on remote sensing satellite data, Internationa Journal

of Advanced Research in Artificial Intelligence (IJARAI), 4, 7, 15-20, 2015.

[14] Arai, K., Locality of Chlorophyll-a Concentration in the Intensive Study Area of the Ariake Sea, Japan in Winter Seasons Based on Remote Sensing Satellite Data, Internationa Journal of Advanced Research in Artificial Intelligence (IJARAI), 4, 8, 18-25, 2015.

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 April 1974 to December 1978 also was with National Space Development Agency of Japan from January, 1979 to March, 1990. 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 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 33 books and published 510 journal papers. He is now Editor-in-Chief of IJACSA and IJISA. http://teagis.ip.is.saga-u.ac.jp/index.html

Wildlife Damage Estimation and Prediction Using Blog and Tweet Information

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

Shohei Fujise
Graduate School of Science and Engineering
Saga University
Saga City, Japan

***Abstract*—Wildlife damage estimation and prediction using blog and tweet information is conducted. Through a regressive analysis with the truth data about wildlife damage which is acquired by the federal and provincial governments and the blog and the tweet information about wildlife damage which are acquired in the same year, it is found that some possibility for estimation and prediction of wildlife damage. Through experiments, it is found that R^2 value of the relations between the federal and provincial government gathered truth data of wildlife damages and the blog and the tweet information derived wildlife damages is more than 0.75. Also, it is possible to predict wildlife damage by using past truth data and the estimated wildlife damages. Therefore, it is concluded that the proposed method is applicable to estimate and predict wildlife damages.**

Keywords*—*Wildlife damage; Blog; Tweet; Big data analysis; Natural language recognition

I. Introduction

Wildlife damage in Japan is around 23 Billion Japanese Yen a year in accordance with the report from the Ministry of Agriculture, Japan. In particular, wildlife damages by deer and wild pigs are dominant (10 times much greater than the others) in comparison to the damage due to monkeys, bulbuls (birds), rats. Therefore, there are strong demands to mitigate the wildlife damage as much as we could. It, however, is not so easy to find and capture the wildlife due to lack of information about behavior. For instance, their routes, lurk locations are unknown and not easy to find. Therefore, it is difficult to determine the appropriate location of launch a trap. In Kyushu, Japan, wildlife damage is getting large and is one of severe problems for farmers as well as residents in the districts near the mountainous areas. The federal and provincial agricultural management organizations in the districts are surveying the wildlife damages every year. It is time consuming task and requires large budget. Also, it takes almost two years. Therefore, it is hard to make a plan for wildlife damage controls. It would be helpful to estimate and predict wildlife damages with some other methods. Meanwhile, blog and tweet information can gather with some software tools. Furthermore, it would be possible to extract some valuable information relating to wildlife damages. The method proposed here is to estimate and predict wildlife damages by using blog and tweet information. It can be done immediately after the end of the Japanese fiscal year. Therefore, wildlife damage prevention plan can be created by the end of the Japanese fiscal year.

The following section describes the proposed method for wildlife damage estimation and predictions followed by experimental data. Then, concluding remarks and some discussions are followed.

II. Literature and Related Work

According to the West, B. C., A. L. Cooper, and J. B. Armstrong, 2009, "Managing wild pigs: A technical guide. Human-Wildlife Interactions Monograph"[1], 1–551, there are the following wild pig damages, Ecological Impacts to ecosystems can take the form of decreased water quality, increased propagation of exotic plant species, increased soil erosion, modification of nutrient cycles, and damage to native plant species [1]-[5]. Agricultural Crops Wild pigs can damage timber, pastures, and, especially, agricultural crops [6]-[9]. Forest Restoration Seedlings of both hardwoods and pines, especially longleaf pines, are very susceptible to pig damage through direct consumption, rooting, and trampling [10]-[12]. Disease Threats to Humans and Livestock Wild pigs carry numerous parasites and diseases that potentially threaten the health of humans, livestock, and wildlife [13]-[15]. Humans can be infected by several of these, including diseases such as brucellosis, leptospirosis, salmonellosis, toxoplasmosis, sarcoptic mange, and trichinosis. Diseases of significance to livestock and other animals include pseudorabies, swine brucellosis, tuberculosis, vesicular stomatis, and classical swine fever [14], [16]-[18]. There also are some lethal techniques for damage managements. One of these is trapping. It is reported that an intense trapping program can reduce populations by 80 to 90% [19]. Some individuals, however, are resistant to trapping; thus, trapping alone is unlikely to be successful in entirely eradicating populations. In general, cage traps, including both large corral traps and portable drop-gate traps, are most popular and effective, but success varies seasonally with the availability of natural food sources [20]. Cage or pen traps are based on a holding container with some type of a gate or door [21]. The method and system for monitoring the total number of wild pigs in the certain district in concern is proposed [22]. All the aforementioned system is not so cheap. It requires huge resources of human-ware, hardware and software as well. Also, it is totally time consumable task. Usually, it takes two years to finalize the total number of wild animals and wildlife damages. Therefore, it is hard to plan the countermeasures for the wildlife damages.

[1] www.berrymaninstitute.org/publications,

III. PROPOSED METHOD

A. Methods for Acquisition of Blog and Tweet Information Relating to Wildlife Damages

There are some sites which allow acquisition of tweet and blog information. Fig.1 (a) shows one of the tweet information acquisition sites while Fig.1 (b) shows one of the blog information acquisition sites. For the tweet information acquisition site (https://dev.twitter.com/rest/public/search), the Search API is not complete index of all Tweets, but instead an index of recent Tweets. At the moment that index includes between 6-9 days of Tweets. Therefore, tweet information has to be acquired within 6-9 days after the event of wildlife appearance. It required some information collection robots. These examples are http://blog.ritlweb.com/ for blog information collection while http://twitter.com/ is for tweet information collections.

(a)Tweet

(b)Blog

Fig. 1. Examples of the tweet and the blog information acquisition sites

B. Methods for Extraction of Wildlife Damage Information from the Acquired Blog and Tweet Information

It has to be done to extract wildlife damage related information from the acquired blog and tweet information. The following set of three parameters have to be extracted, (1) the area name, (2) the types of wildlife damages, (3) the date of the wildlife damage reported. In order to extract sets of information, "Chasen" of sentence structure and words analysis software tool is used. It is morphological analysis tool. The extracted words and sentences acquired from the twitter and blog data collection sites are input to the "Chasen". Then noun and the other part of speech can be extracted as shown in Fig.2.

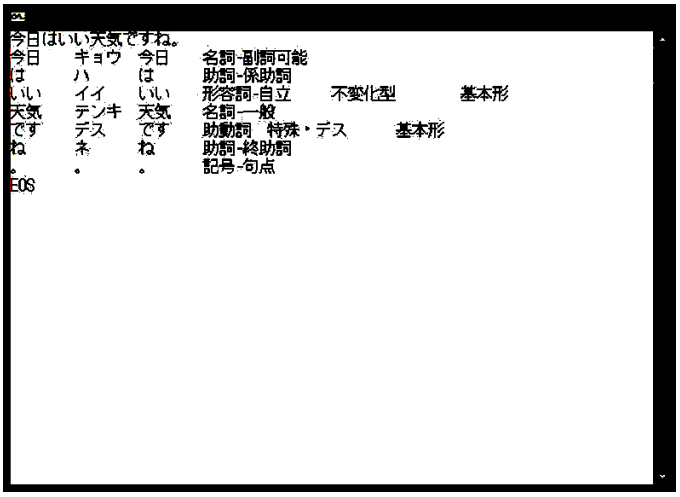


Fig. 2. Example of the screen shot of the Chasen analysis

The acquired sentence is "It is fine today" in Japanese and is appeared at the first line of the example. The first column of the second to the eighth lines "Today", "is" "Fine", "Weather", "it", and "is not it" show the words extracted from the acquired sentence. The second column shows their sounds while the forth column shows their part of speech. Thus, the words can be divided and extracted from the sentence together with their part of speech. Therefore, nouns can be extracted from the sentences. After that full text search is conducted to the extracted words.

Firstly, area names are extracted from the extracted words. In this regards, City name, Town name, and Village name in Kyushu provided by the federal and provincial governments are used in order to extract the area names. After that the names of the wildlife which is provided by the federal government of Agriculture, Forestry and Fishery ministry are extracted from the words. In this regards, combined words such as "prevention of bird damage" is recognized as the words of wildlife damage. The date of the tweet and blog information is easily extracted from the tweet and blob information because the information is dated information. Thus when, where, which wildlife can be extracted from the tweet and blog information.

C. Methods for Estimation of Wildlife Damage from the Acquired Tweet and Blob Information

The number of wildlife damage reports which are extracted from the acquired tweet and blog information in the year in concern must be proportional to the wildlife damages in that year. Therefore, linear regression would work for estimation of wildlife damage with the acquired tweet and blob information.

D. Methods for Prediction of Wildlife Damage Information from the Acquired Blog and Tweet Information

Based on the well known time series analysis method, it is possible to predict using the past wildlife damage. If the estimated wildlife damage with tweet and blob information is used for the wildlife damage in year in concern together with the past wildlife damage, then it is possible to predict future wildlife damage. In this regards, the following linear prediction is used for this,

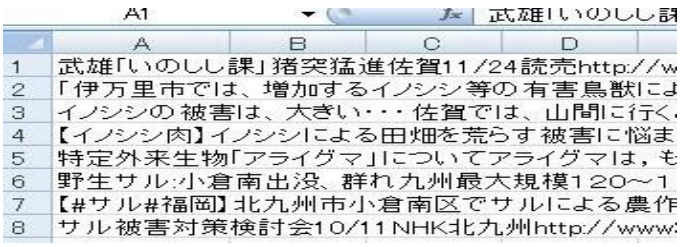
$$y - \bar{y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}(x - \bar{x}) \quad (1)$$

where x and y denote the past wildlife damage and the current wildlife damage, respectively. xbar and ybar denote mean of the past and the current wildlife damage, respectively.

IV. EXPERIMENTS

A. Examplessof the Acquied Blog and Tweet Information Relating to Wildlife Damages

One of the examples of the tweet and blog information relating to wildlife damage is shown in Fig.3 (a). Meanwhile, the extracted words of area names and the types of wildlife are shown in Fig.3 (b) while the results from the wildlife damage estimated from the acquired tweet and blog information is shown in Fig.3 (c), respectively. The summarized results of the number of wildlife damage which are reported by twitter and blog at every province, Fukuoka, Saga, Nagasaki, Ohita, Kumamoto, Miyazaki, and Kagoshima prefectures in Kyushu in 2013.



(a)Tweet and blob

(b)Area name and types of wildlife name

県名	鳥獣計	鳥類計		獣類計			
			カラス		イノシシ	サル	シカ
福岡	118	3	2	115	45	22	27
佐賀	82	8	1	74	45	5	19
長崎	103	2	2	101	58	12	17
熊本	18	0	0	18	3	0	4
大分	97	2	2	95	38	4	22
宮崎	50	2	2	48	18	6	24
鹿児島	0	0	0	0	0	0	0

(c)Example of the results of the number of wildlife damages (for every Provinces)

Fig. 3. Examples of the acquired tweet and blog information, the area name and the types of wildlife name as well as the summarized results from the wildlife damage in Kyushu in 2013

B. True Wildlife Damage Reported by the Regional Govermental Insititude of Kyushu Agricultural Management

True wildlife damage reported by the regional governmental institute of Agricultural Management in 2013 is shown in Table 1.

TABLE I. TRUE WILDLIFE DAMAGE REPORTED BY THE REGIONAL GOVERNMENTAL INSTITUTE OF AGRICULTURAL MANAGEMENT IN 2013

	Wildlife	Birds	Crow	Animals	Wild pig	Monkey	Deer
Fukuoka	91671	36306	19551	55366	35867	2578	7986
Saga	20486	6040	4559	14446	11023	1130	0
Nagasaki	28724	3171	2194	25553	23930	1	470
Kumamoto	45531	10398	8745	35133	28031	1311	4030
Ohita	21550	1078	749	20472	14474	1355	3229
Miyazaki	72978	4242	3363	68736	33396	7287	26066
Kagoshima	43950	8848	3219	35102	17070	2183	12878

The prefecture which shows the largest wildlife damage is Fukuoka followed by Miyazaki, Kumamoto, Kagoshima. Nagasaki, Ohita and Saga. The number of reports of wildlife damage, on the other hand, is shown in Table 2. The correlation coefficient between the total numbers of the reports and the total wildlife damage is just 0.013 as shown in Table 2.

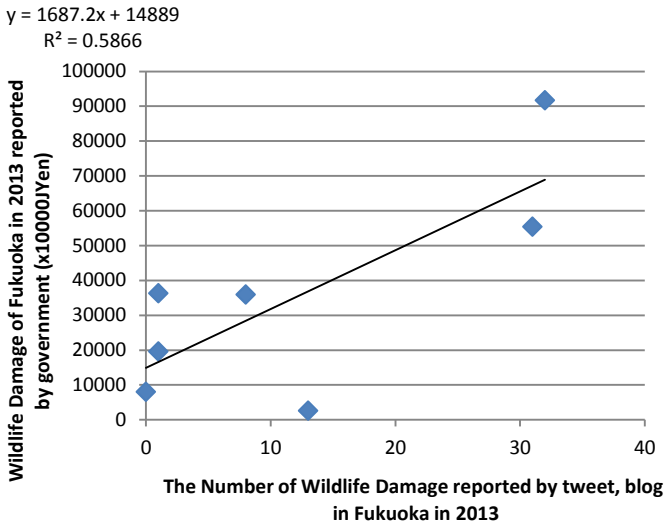
TABLE II. NUMBER OF REPORTS OF WILDLIFE DAMAGE AND TOTAL WILDLIFE DAMAGE IN KYUSHU IN 2013

	Fukuoka	Saga	Nagasaki	Kumamoto	Oita	Miyazaki	Kagoshima
Wildlife	32	7	42	12	29	8	1
Birds	1	0	0	1	0	1	0
Crow	1	0	0	0	0	0	0
Animals	31	7	42	11	29	7	1
Wild pig	8	6	35	2	3	0	0
Monkey	13	0	1	0	0	5	0
Deer	0	0	6	3	11	2	0
Wildlife	91671	20486	28724	45531	21550	72978	43950
No.of report	86	14	126	23	72	23	2

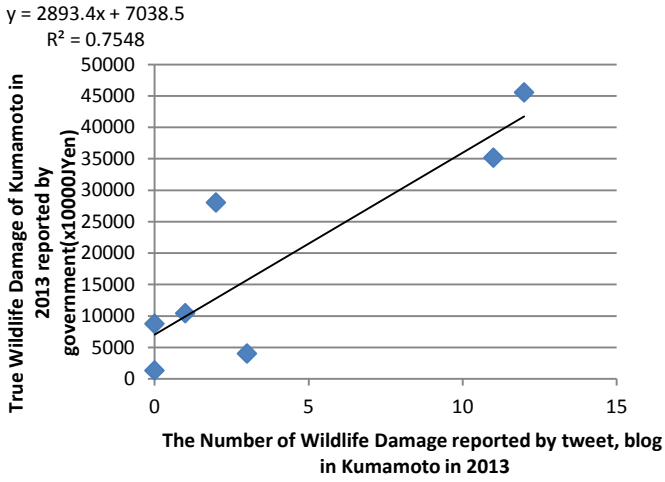
Although correlation coefficient is so poor, R=0.013, if the number of reports of wildlife damages of crow and birds, as well as monkey is deleted together with the number of report of Saga, Kumamoto and Kagoshima due to the fact that the number of reports are so small then the correlation coefficient between the total wildlife damage and the total number of the reports of wildlife damage through blog and tweet is increased R=0.538. Therefore, the relation between both is not so poor.

C. Estimation of Wildlife Damage from the Number of Reported Tweet and Blog for Every Province

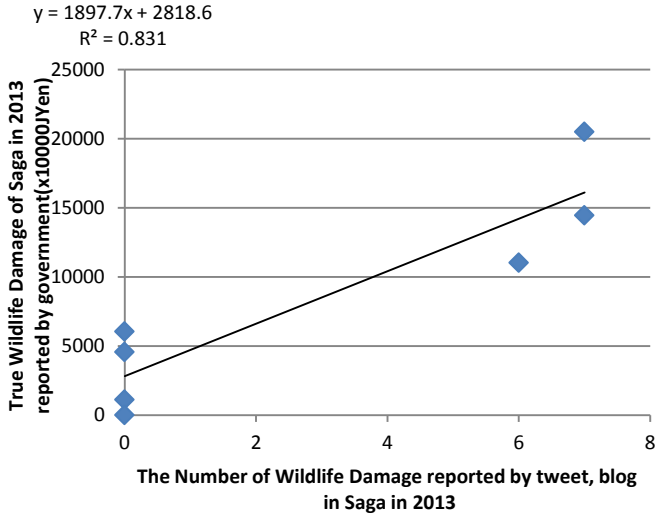
Through the linear regressive analysis, it can be done to estimate wildlife damages using the reported tweet and blog information. The results from the regressive analysis are shown in Fig.5. At the top left corners of the figures in Fig.4, there are regressive equations and the R^2 values. The R^2 values range from 0.5657 to 0.9693 while slope (gain) coefficients range from 607.17 to 30686. On the other hand, the number of reports of tweet and blog (Horizontal axis of the graphs in Fig.4) range from 1 to 42. The uncertainty of the regressive analysis is totally dependent to the number of reports. Therefore, the regressive analysis results of Saga, Kagoshima, Miyazaki are not so reliable. Then the ranges of the R2 values and gain coefficients are (0.5866 – 0.9693), and (607.17 – 2893.4), respectively.



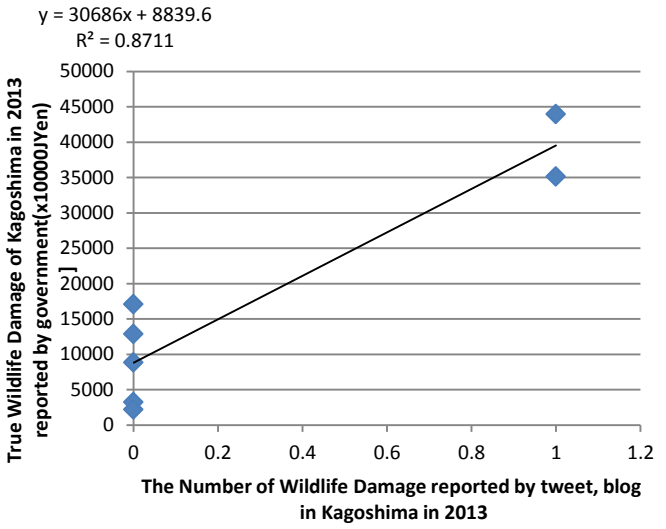
(a)Fukuoka



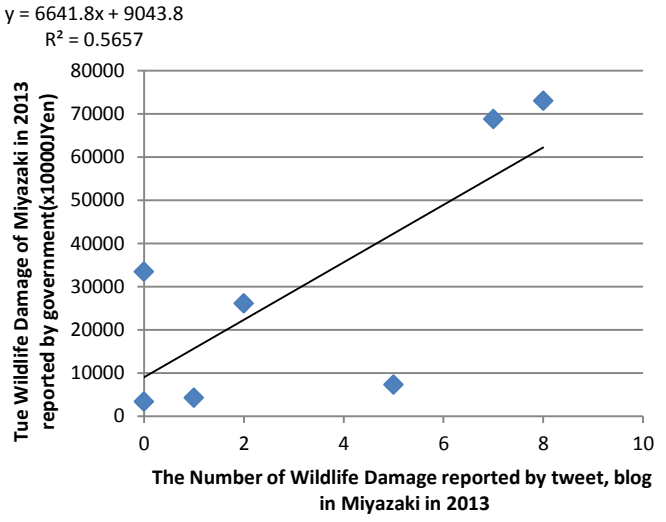
(b)Kumamoto



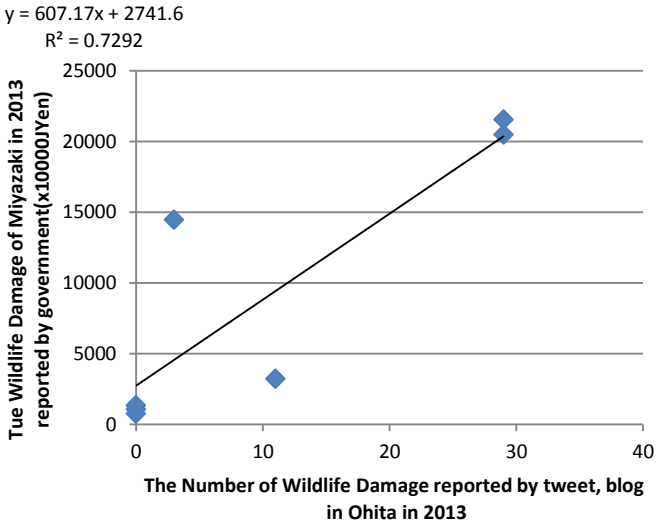
(c)Saga



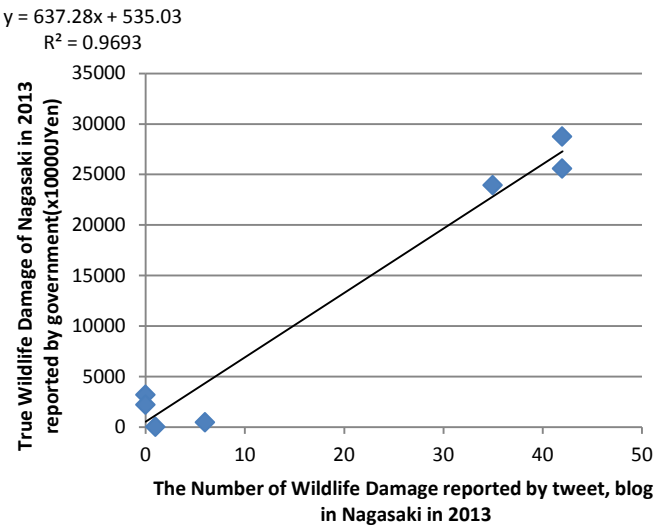
(d)Kagoshima



(e)Miyazaki



(f)Ohita



(g)Nagasaki

Fig. 4. Estimate wildlife damages for every province using the reported tweet and blog information

D. *Predictions of Wildlife Damage from the Number of Reported Tweet and Blog for Every Province*

The newest true wildlife damage data is 2014 which is provided by Kumamoto prefecture. There is no other prefecture of which true wildlife damage of 2014 is reported. Therefore, the wildlife damage of 2014 is predicted by using the past data of wildlife damage (2008 to 2013) based on the linear prediction which is expressed in equation (1). Table 3 shows the results from the predicted wildlife damage (in the second row of Table 3). The correlation between the wildlife damage from the true report of prefecture Kumamoto and predicted wildlife damage from the report of blog and tweet information is 0.996. By taking into account the compensation of mean and standard deviation of the predicted wildlife damage (adjusted), the difference between true wildlife damage and the predicted wildlife damage from the acquired blog and tweet information ranges from -1158 to 2944 in unit of 10,000 Japanese Yen.

TABLE III. PREDICTED WILDLIFE DAMAGE BY USING THE PAST DATA FOR 6 YEARS, 2008 TO 2013

Kumamoto	Wildlife	Birds	Crow	Animals	Wild pig	Monkey	Deer
True report	45531	10398	8745	35133	28031	1311	4030
Predicted	65270.8	14856	9520.9	52210.4	37813.5	4761.7	7181.1
Adjusted	44689.56	9399.2	5664.63	35547.28	25469.45	2333.19	4026.77
Difference	841.44	998.8	3080.37	-414.28	2561.55	-1022.19	3.23

From the relation between year and wildlife damage in Kumamoto in unit of 10,000 Japanese Yen, the wildlife damage can be calculated with the number of the tweet and the blog. Red colored number in Table 4 shows the calculated wildlife damage and the blue colored number indicates the predicted wildlife damage derived from the linear prediction with the true wildlife damage for five years (2008 – 2012) and the estimated wildlife damage in 2013. Through a comparison between true wildlife damage and the predicted one is approximately 6.0 %. Therefore, it is capable to predict wildlife damage in the next year with the past true wildlife damage reported by the local prefectural government and the relation between wildlife damage and the number of report by twitter and blog.

TABLE IV. COMPARISON OF THE WILDLIFE DAMAGES BETWEEN TRUE AND THE PREDICTION

Year	2008	2009	2010	2011	2012	2013	2014
True report	61468	70013	84516	54495	51975	45531	47235
True+Estimated	61468	70013	84516	54495	51975	58509.25	51000

Fig.5 shows the true and the predicted wildlife damages as a function of year. Therefore, it may say that wildlife damage in the next year can be predicted with the past true data of wildlife damage and the relation between the number of reports by twitter and blog.

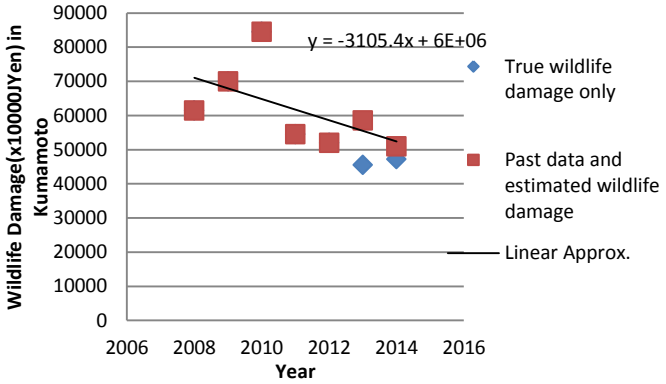


Fig. 5. True and the predicted wildlife damages as a function of year

V. CONCLUSION

Method for wildlife damage estimation and prediction using blog and tweet information relating to wildlife appearances is proposed in this paper. Through regressive analysis with the truth data about wildlife damage which is acquired by the federal and provincial governments and the blog and tweet information about wildlife damage which are acquired in the same year, it is found that some possibility for estimation and prediction of wildlife damage. Through experiments, it is found that R^2 value of the relations between the federal and provincial government gathered truth data of wildlife damages and blog tweet information derived wildlife damages is more than 0.75. Also, it is possible to predict wildlife damage by using past truth data and the estimated wildlife damages. Therefore, it is concluded that the proposed method is applicable to estimate and predict wildlife damages.

It is also found that the correlation between the wildlife damage from the true report of prefecture Kumamoto and predicted wildlife damage from the report of blog and tweet information is 0.996. By taking into account the compensation of mean and standard deviation of the predicted wildlife damage (adjusted), the difference between true wildlife damage and the predicted wildlife damage from the acquired blog and tweet information ranges from -1158 to 2944 in unit of 10,000 Japanese Yen. Therefore, future wildlife damage can be predicted by using the reports from blog and tweet information in some extent.

Further investigations are required for increasing the cases of wildlife damages for improving prediction accuracy.

ACKNOWLEDGMENT

The author would like to thank Prof. Dr. Hiroshi Okumura of Saga University for his valuable suggestions and comments for this study.

REFERENCES

[1] Patten, D. C. 1974. Feral hogs — boon or burden. Proceedings of the Sixth Vertebrate Pest Conference 6:210–234.

[2] Singer, F. J., W. T. Swank, and E. E. C. Clebsch. 1984. Effects of wild pig rooting in a deciduous forest. Journal of Wildlife Management. 48:464–473.

[3] Stone, C. P., and J. O. Keith. 1987. Control of feral ungulates and small mammals in Hawaii's national parks: research and management strategies. Pages 277–287 in C. G. J. Richards and T. Y. Ku, editors. Control of mammal pests. Taylor and Francis, London, England, and New York and Philadelphia, USA.

[4] Cushman, J. H., T. A. Tierney, and J. M. Hinds. 2004. Variable effects of feral pig disturbances on native and exotic plants in a California grassland. Ecological Applications 14:1746–1756.

[5] Kaller, M. D., and W. E. Kelso. 2006. Swine activity alters invertebrate and microbial communities in a coastal plain watershed. American Midland Naturalist 156:163–177.

[6] Bratton, S. P. 1977. The effect of European wild boar on the flora of the Great Smoky Mountains National Park. Pages 47–52 in G. W. Wood, editor. Research and management of wild hog populations. Belle W. Baruch Forest Science Institute, Clemson University, Georgetown, South Carolina, USA.

[7] Lucas, E. G. 1977. Feral hogs — problems and control on National Forest lands. Pages 17–22 in G. W. Wood, editor. Research and management of wild hog populations. Belle Baruch Forest Science Institute, Clemson University, Georgetown, South Carolina, USA.

[8] Thompson, R. L. 1977. Feral hogs on National Wildlife Refuges. Pages 11–15 in G. W. Wood, editor. Research and management of wild hog populations. Belle W. Baruch Forest Science Institute, Clemson University, Georgetown, South Carolina, USA. Kohei Arai, Preliminary Assessment of Radiometric Accuracy for MOS-1 Sensors, International Journal of Remote Sensing, Vol.9, No.1, pp.5-12, Apr.1988.

[9] Schley, L, and T. J. Roper. 2003. Diet of wild boar Sus scrofa in Western Europe, with particular reference to consumption of agricultural crops. Mammal Review 33:43–56.

[10] Whitehouse, D. B. 1999. Impacts of feral hogs on corporate timberlands in the southeastern United States. Pages 108–110 in Proceedings of the Feral Swine Symposium, June 2–3, 1999, Ft. Worth, Texas, USA.

[11] Mayer, J. J., E. A. Nelson, and L. D. Wike. 2000. Selective depredation of planted hardwood seedlings by wild pigs in a wetland restoration area. Ecological Engineering, 15(Supplement 1): S79–S85.

[12] Campbell, T. A., and D. B. Long. 2009. Feral swine damage and damage management in forested ecosystems. Forest Ecology and Management 257:2319–2326

[13] Forrester, D. J. 1991. Parasites and diseases of wild mammals in Florida. University of Florida Press, Gainesville, Florida, USA.

[14] Williams, E. S., and I. K. Barker. 2001. Infectious diseases of wild mammals. Iowa State University Press, Ames, Iowa, USA.

[15] Sweeney, J. R., J. M. Sweeney, and S. W. Sweeney. 2003. Feral hog. Pages 1164–1179 in G. A. Feldhamer, B. C. Thompson, and J. A. Chapman, editors. Wild mammals of North America. Johns Hopkins University Press, Baltimore, Maryland, USA.

[16] Nettles, V.F., J. L. Corn, G. A. Erickson, and D. A. Jessup. 1989. A survey of wild swine in the United States for evidence of hog cholera. Journal of Wildlife Diseases 25:61–65.

[17] Davidson, W. R., and V. F. Nettles, editors. 1997. Wild swine. Pages 104–133 in Field manual of wildlife diseases in the southeastern United States. Second edition. Southeastern Cooperative Wildlife Disease Study, Athens, Georgia, USA.

[18] Davidson, W. R., editor. 2006. Wild swine. Pages 105–134 in Field manual of wildlife diseases in the southeastern United States. Third . Southeastern Cooperative Wildlife Disease Study, Athens, Georgia, USA.

[19] Choquenot, D. J., R. J. Kilgour, and B. S. Lukins. 1993. An evaluation of feral pig trapping. Wildlife Research, 20:15– 22.

[20] Barrett, R. H., and G. H. Birmingham. 1994. Wild pigs. Pages D65–D70 in S. Hyngstrom, R. Timm, and G. Larsen, editors. Prevention and control of wildlife damage. Cooperative Extension Service, University of Nebraska, Lincoln, Nebraska, USA.

[21] Mapston, M. E. 1999. Feral hog control methods. Pages 117–120 in Proceedings of the Feral Swine Symposium, June 2–3, 1999, Fort Worth, Texas, USA

[22] Kohei Arai, Indra Nugraha Abdullah, Kensuke Kubo, Katsumi Sugaw, Methods for Wild Pig Identifications from Moving Pictures and Discrimination of Female Wild Pigs based on Feature Matching Method, (IJARAI) International Journal of Advanced Research in Artificial Intelligence, Vol. 4, No.7, 41-46, 2015

AUTHORS PROFILE

Kohei Aarai 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 33 books and published 510 journal papers. He is Editor-in-Chief of International Journal of Advanced Computer Science and Applications as well as I ternational Journal of Intelligent Systsems and Applications. http://teagis.ip.is.saga-u.ac.jp/index.html

Location Monitoring System with GPS, Zigbee and Wifi Beacon for Rescuing Disable Persons

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

Taka Eguchi
[1]Graduate School of Science and Engineering
Saga University
Saga City, Japan

***Abstract*—Location monitoring system for rescue disable persons by switching the location estimation methods with GPS, ZigBee and WiFi beacon is proposed. Rescue system with triage using health condition monitoring together with location and attitude monitoring as well as the other data acquired with mobile devices is evaluated with the proposed location monitoring system. Through simulation study, influence due to location estimation error on rescue time is evaluated together with effect of the proposed location monitoring system. Also, it is found that the effect of triage on rescue time is clarified.**

Keywords—Rescue system; Location estimation; Attitude estimation; Health monitoring; Mobile applications; Triage; Rescue planning

I. INTRODUCTION

Most computer based simulation evacuation models are based on flow model, cellular automata model, and multi-agent-based model. Flow based model lacks interaction between evacuees and human behavior in crisis. Cellular automata model is arranged on a rigid grid, and interact with one another by certain rules [1]. A multi agent-based model is composed of individual units, situated in an explicit space, and provided with their own attributes and rules [2]. This model is particularly suitable for modeling human behaviors, as human characteristics can be presented as agent behaviors. Therefore, the multi agent-based model is widely used for evacuation simulation [1]-[4].

Recently, Geographic Information Systems: GIS is also integrated with multi-agent-based model for emergency simulation. GIS can be used to solve complex planning and decision making problems [5]-[7]. In this study, GIS is used to present road network with attributes to indicate the road conditions.

We develop a task allocation model for search and rescue persons with disabilities and simulate the rescue process to capture the phenomena and complexities during evacuations. The task allocation problem is presented by decision of volunteers to choose which victims should be helped in order to give first-aid and transportation with the least delay to the shelter. The decision making is based on several criteria such as health condition of the victims, location of the victims and location of volunteers [8]-[18].

A rescue model for people with disabilities in large scale environment is proposed. The proposed rescue model provides some specific functions to help disabled people effectively when emergency situation occurs. Important components of an evacuation plan are the ability to receive critical information about an emergency, how to respond to an emergency, and where to go to receive assistance. Triage is a key for rescue procedure. Triage can be done with the gathered physical and psychological data which are measured with a sensor network for vital sign monitoring. Through a comparison between with and without consideration of triage, it may be possible to find that the time required for evacuation from disaster areas with consideration triage is less than that without triage [19]-[20].

These studies do not taken into account location estimation accuracy. GPS utilized location accuracy is assumed to be 100% accurate (no error). There, actually, are location estimation errors. Other methods for location estimation are available, for instance, ZigBee, WiFi beacon utilizing methods. All the methods have errors which depend on the conditions, location of radio wave absorbance, weather condition, multi-path condition, etc. Some of these error sources are controllable except weather condition. A prior to location estimation, location of radio wave absorbance, and multi-path condition can be assessed. It would be possible to use the best accuracy of location estimation method can be used alternatively from among GPS, WiFi beacon, and ZigBee. In particular, ZigBee transmitter can be layout arbitrary in accordance with the required location estimation accuracy. Thus the best location estimation accuracy could be achieved.

The next section describes influence due to location estimation error on the rescue time in concern followed by the proposed location estimation method by using GPS, WiFi beacon and ZigBee alternatively. Then experimental results from rescue simulation studies which is based on triage with health condition of victims. Finally, conclusion is describes together with some discussions.

II. INFLUENCE DUE TO LOCATION ESTIMATION ERROR ON RESCUE TIME

A. Comparison of Location Estimation Methods Among GIS, WiFi Beacon and ZigBee Based Methods

The location estimation with GPS is accurate for outdoor situation with the condition without radio wave absorbance as shown in Fig.1. It, on the other hand, is poor accuracy for indoor situation with the condition with radio wave absorbance. Meanwhile, WiFi beacon based location estimation is available for both indoor and outdoor situations. Accuracy is dependent

on the number of available routers and the existing radio wave absorbance.

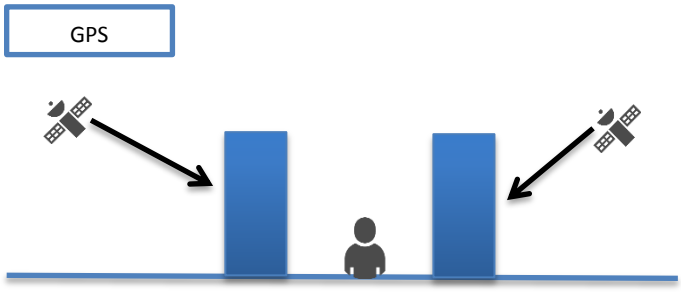


Fig. 1. GPS based location estimation system

On the other hand, ZigBee based location estimation needs very poor power consumption as shown in Fig.2. The size of ZigBee transmitter and receiver as well as repeater is very compact so that it can be set up anywhere. Furthermore, ZigBee transmitter and receiver is relatively cheap in comparison to the WiFi router. The transmitters of WiFi beacon router and ZigBee transmitter can be set-up arbitrary depending on the required accuracy.

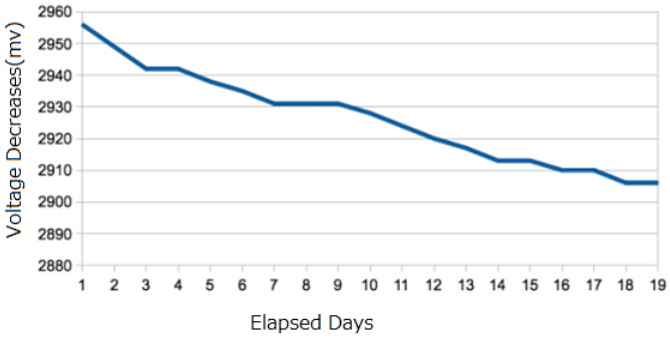


Fig. 2. Power consumption of ZigBee

Measured and theoretical signal strengths of WiFi beacon based location estimation method are shown in Fig.3. The theoretical receiving signal strength is expressed in equation (1).

$$RSSI = -(10\log_{10} d + A) \quad (1)$$

where RSSI denotes receiving signal strength while d and A denotes distance between transmitter (WiFi router) and receiver as well as signal strength for the case of the distance is 1m, respectively. Measured signal strength shows a good coincidence to the theoretical strength.

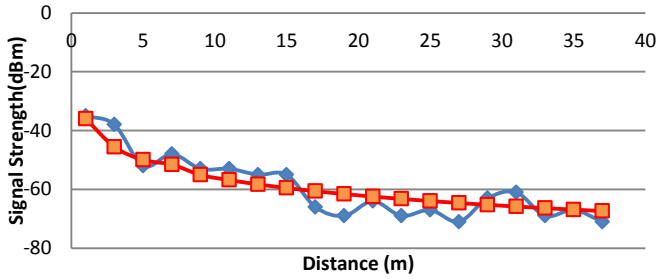
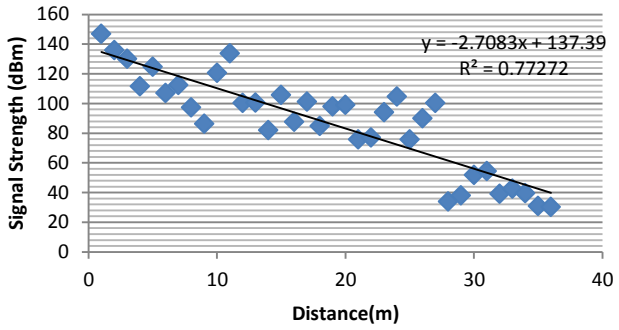
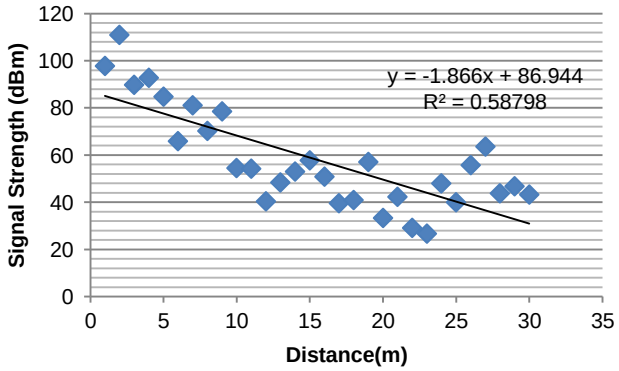


Fig. 3. Measured and theoretical signal strength as a function of distance

On the other hand, measured signal strength of ZigBee receiver as a function of distance between transmitter and receiver is shown in Fig.4 (a) for the indoor situation while that for the outdoor situation is shown in Fig.4 (b), respectively.



(a)Indoor

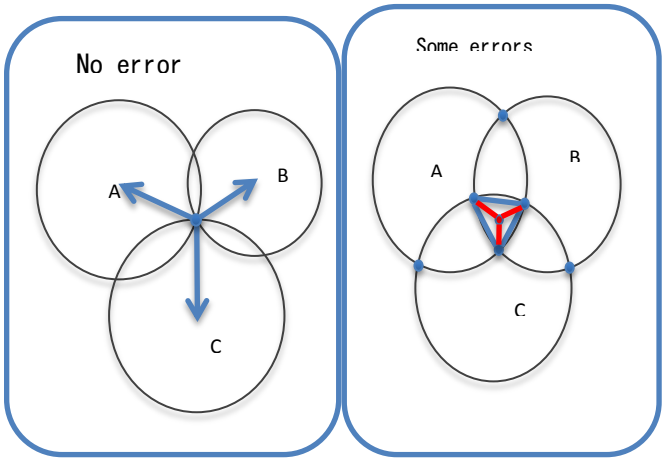


(b)Outdoor

Fig. 4. Measured signal strength of ZigBee receiver as a function of distance between transmitter and receiver

Signal strength in the case of outdoor situation is rather weak in comparison to that for indoor situation. Therefore, correlation between signal strength and distance is not so high for outdoor situation which results in relatively poor location estimation accuracy for ZigBee based location estimation method.

There is no error when the position is situated at the cross point among three circles of which the locations of WiFi routers and ZigBee transmitters are situated as shown in Fig.5 (a). On the other hand, some errors would occur in the situations which is shown in Fig.5 (b) for both WiFi router based and ZigBee based location estimation methods, Location estimate has to be made at the location as gravity center of the triangle which is formed with three circles



(a)No error (b)Some errors

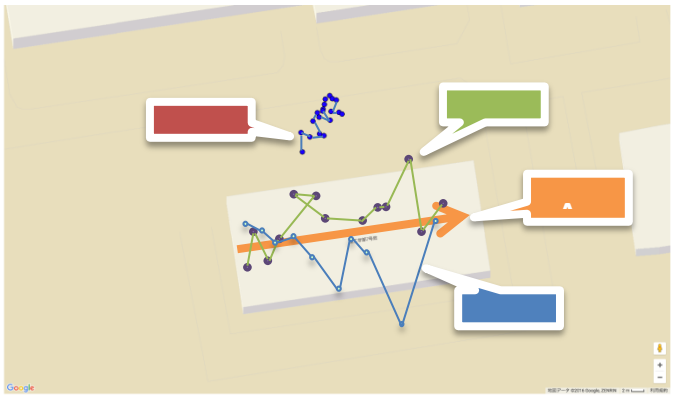
Fig. 5. Situation when some errors occurrence

B. Location Estimation Accuracy Assessment

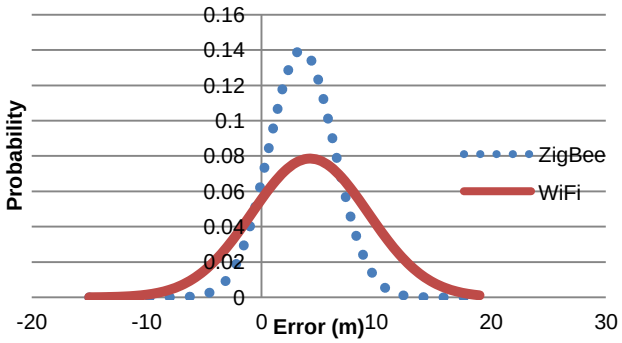
Location estimation accuracy of three location estimation methods with GPS, WiFi beacon and ZigBee is assessed for both indoor and outdoor situations. The actual location is the 7th building of the Science and Engineering Faculty of Saga University for indoor situation (Collider on the fourth floor). Meanwhile, outdoor is situated at the southern portion of the 7th building (Parking lot). Fig.6 (a) shows exact measured locations with GPS, WiFi beacon and ZigBee based methods while Fig.6 (b) shows probability density function of the measured location errors for WiFi beacon and ZigBee based methods.

The measured locations with GPS based method have obvious and significant bias error due to the fact that radio wave from GPS satellites comes from the left windows nearby the receiver. Even for the measured locations with WiFi beacon and ZigBee based methods have 4 to 5 meters of bias errors. In terms of mean and standard deviation, ZigBee based method is superior to WiFi beacon based method.

Meanwhile, Fig.7 (a) shows the measured locations in outdoor situation for GPS, WiFi beacon and ZigBee based methods. On the other hand, Fig.7 (b) shows probability density function of the measured location errors.

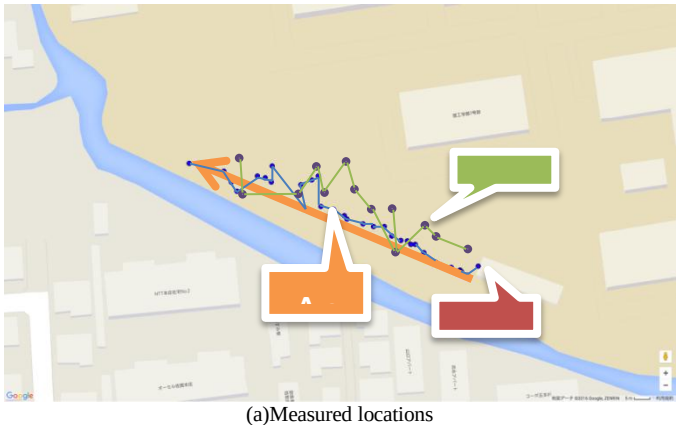


(a)Measured locations

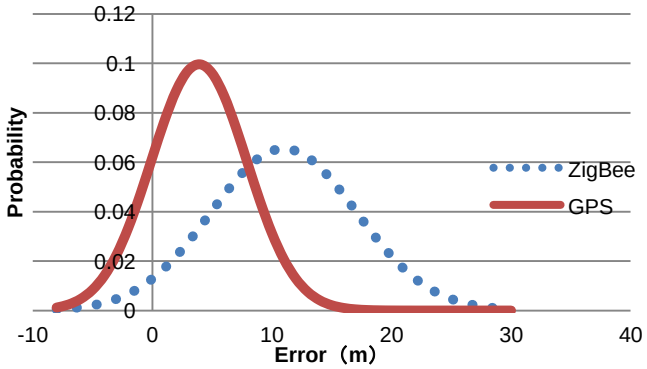


(b)Probability Density Function

Fig. 6. Location estimation accuracy in indoor and outdoor situations



(a)Measured locations



(b)Probability Density Function

Fig. 7. Location estimation accuracy in indoor and outdoor situations

In the case of ZigBee based location estimation, the locations and layouts of transmitters are key for location estimation accuracy. For both cases of location estimations with ZigBee based method between indoor and outdoor situations, just three transmitters are used for location estimations. In the case of outdoor situation, GPS based method is superior to ZigBee based method. GPS based method shows 3m of bias error while ZigBee based method shows 11m of bias error.

Mean and standard deviation of location estimation error for both ZigBee and WiFi based location estimation methods in indoor and outdoor situations are shown in Table 1 and 2, respectively.

TABLE I. MEAN AND STANDARD DEVIATION OF LOCATION ESTIMATION ERROR IN THE CASE OF INDOOR SITUATION

	ZigBee	WiFi
Mean: μ	3.51	4.24
Standard Deviation: σ	2.85	5.08

TABLE II. MEAN AND STANDARD DEVIATION OF LOCATION ESTIMATION ERROR IN THE CASE OF OUTDOOR SITUATION

	ZigBee	GPS
Mean: μ	11	3.90
Standard Deviation: σ	6.09	4.00

C. Relation Between Location Estimation Error and Rescue Time

By using GAMA simulation platform, relation between location estimation error and rescue time is clarified. Fig.8 shows road network and the initial positions of victim who needs a help for evacuation (with wheel chair), shelter (green colored house), rescue people (blue colored person), and the measured location (Orange colored circle).

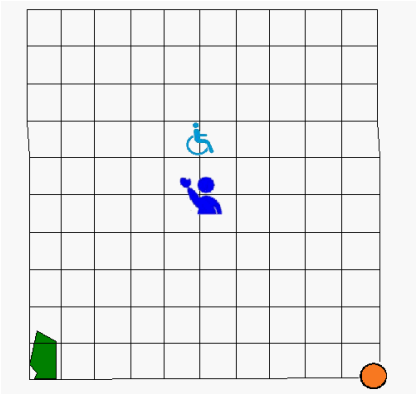


Fig. 8. Initial locations of victim, shelter, recue people and the measured location

The simulation result is shown in Fig.9.

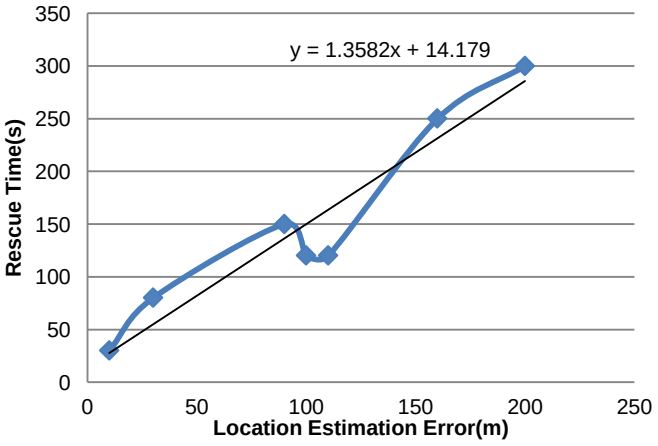


Fig. 9. Relation between location estimation error and rescue time

Even if there is no location estimation error, 14.179 of rescue time is required. In accordance with the location estimation error, rescue time is increased. Linear regressive analysis shows the following relation,

$$y = 1.3582x + 14.179 \quad (2)$$

where x and y denotes location estimation error and rescue time, respectively.

III. RESCUE SIMULATION

A. Rescue Simulation Model

The centralized rescue model is presented which has three types of agent: volunteers, disabled people and route network. The route network is also considered as an agent because the condition of traffic in certain route can be changed when disaster occurs. The general rescue model is shown in Fig.10.

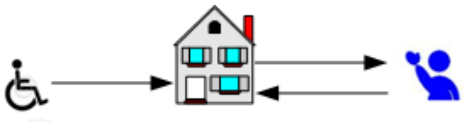


Fig. 10. Centralized Rescue Model

As shown in Fig.10, the concepts of the proposed rescue system. There are three major components, persons who need a help for evacuation, Information Collection Center: ICC for health, traffic, and the other conditions together with the location and attitude information of the persons who need a help and the rescue peoples. Body attached sensors allow measurements of health conditions and the location and attitude of the persons who need a help. The measured data can be transmitted to the ICC through smart-phone, or i-phone, or tablet terminals of which the persons who need a help are carrying. By using the collected health condition and the location/attitude as well as traffic condition information, most appropriate rescue peoples are determined by the person by the person.

B. Road Network and Initial Positions of Victims, Rescue Peoples, Shelters

The sample GIS map consists of 5 layers: road, building, rescue peoples (volunteer agents), victims (disabled persons) and shelter. Fig.11 shows the map.

Fig. 11. Road network and the initial locations of victims, rescue peoples, shelters

The red points and green points indicate the locations of disabled persons and locations of volunteers respectively. These locations are generated randomly along the roads. Blue buildings are shelters. The initial health level of disabled persons is generated randomly between 100 and 500. Every time step of simulation, these health levels decrease by 0.5. If the health level is equal to zero, the corresponding agent is considered as dead. The movements of volunteer agents are controlled by Morimoto Traffic Simulator.

C. Task Allocation Model

The decision making of volunteers to help disabled persons can be treated as a task allocation problem [10]-[14]. The task allocation for rescue scenario is carried out by the central agents. The task of volunteers is to help disabled persons; this task has to be allocated as to which volunteers should help which disabled persons in order to maximize the number of survivals.

We utilize the combinatorial auction mechanism to solve this task allocation problem. At this model, the volunteers are the bidders; the disabled persons are the items; and the emergency center is the auctioneer. The distance and health level of disabled person are used as the cost for the bid. When the rescue process starts, emergency center creates a list of victims, sets the initial distance for victims, and broadcasts the information to all the volunteer agents. Only the volunteer agents whose distance to victims is less than the initial distance will help these victims. It means that each volunteer agent just help the victims within the initial distance instead of helping all the victims. The initial distance will help volunteers to reduce the number of task so that the decision making will be faster.

D. Rescue Simulation Results for the Case Without Any Error

With a fixed number of disabled persons and the number of volunteers increase, the correlation between number of volunteers and rescue time is shown in Fig.12.

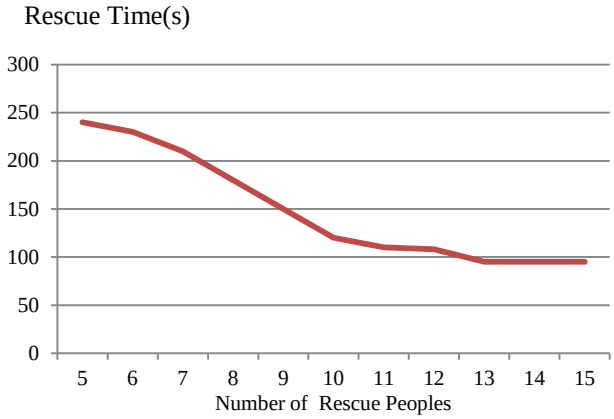


Fig. 12. Correlation between Number of Volunteers and Rescue Time

With a fixed number of volunteers and the number of disabled persons increase, the correlation between number of disabled persons and rescue time is shown in Fig.13.

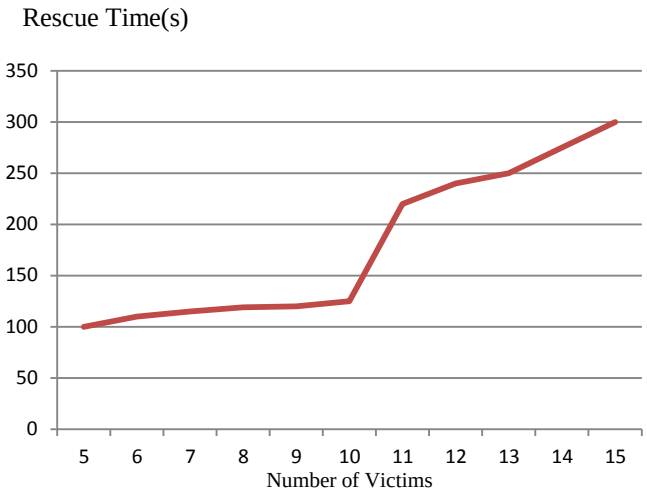


Fig. 13. Correlation between Number of Disabled Persons and Rescue Time

The number of volunteers and the number of disabled persons are fixed, whereas the number of vehicle increases. We test with the total length of road of 500 meters. The increasing number of vehicles will make traffic density higher. The correlation between number of vehicle and rescue time is shown in Fig.14.

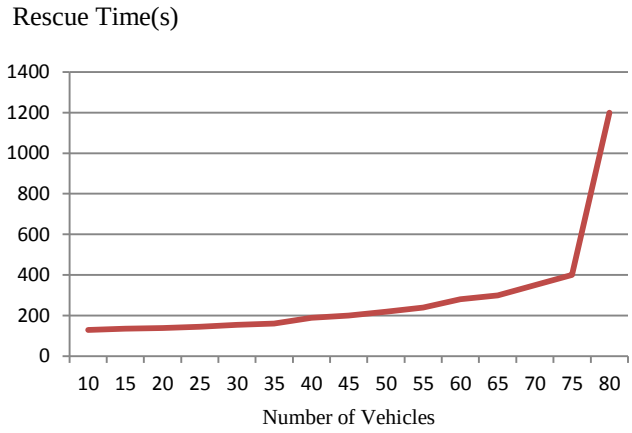
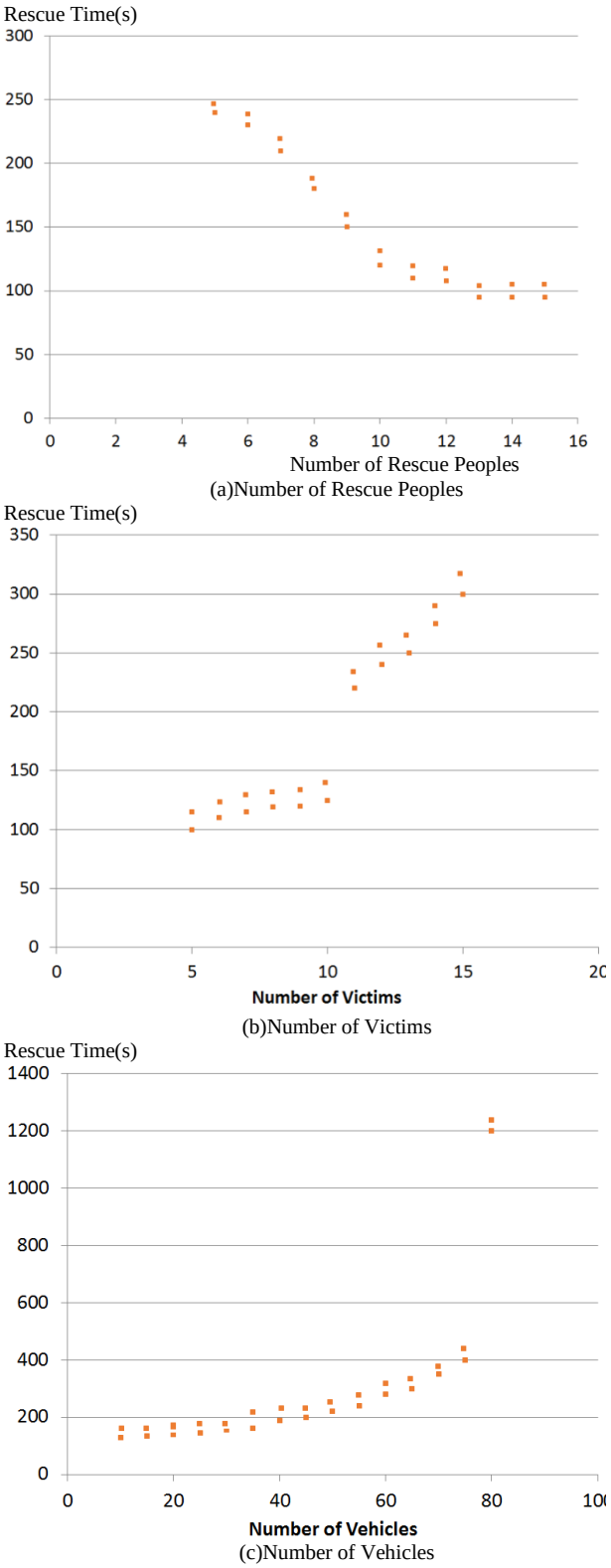


Fig. 14. Correlation between Number of Vehicles and Rescue Time

E. Rescue Simulation Results of the Proposed Location Estimation Method

The results of Fig.12 to 14 are for the case of no location estimation error. Meanwhile, the simulation results taking into account location estimation errors are shown in Fig.15. The basic idea of the proposed location estimation method is that three different location estimation methods are switched depending on the circumstances of the locations of radio wave absorbance, the number of access points (WiFi routers), the number of acquired GPS satellites, and the number of ZigBee transmitters. As shown in Fig.7, rescue time is increased when location estimation error is taken into account even if the best accuracy of location estimation method is selected. The rescue

time for the case of which location estimation error is taken into account is evaluated. Around 10 to 20 seconds are required additionally in comparison to the rescue time without consideration of location estimation error.



(a)Number of Rescue Peoples

(b)Number of Victims

(c)Number of Vehicles

Fig. 15. Rescue time for the case of which location estimation error is taken into account

IV. Conclusion

Location monitoring system for rescue disable persons by switching the location estimation methods with GPS, ZigBee and WiFi beacon is proposed. Rescue system with triage using health condition monitoring together with location and attitude monitoring as well as the other data acquired with mobile devices is evaluated with the proposed location monitoring system. Through simulation study, influence due to location estimation error on rescue time is evaluated together with effect of the proposed location monitoring system. Also, it is found that the effect of triage on rescue time is clarified.

Further study is required for clarifying effects of triage on reducing rescue time in actual situations.

Acknowledgment

The author would like to thank Dr. Trang Xuang Sang of Vinh University in Vietnam for his effort to conduct simulation studies.

References

[1] C. Ren, C. Yang, and S. Jin, “Agent-Based Modeling and Simulation on emergency”, Complex 2009, Part II, LNICST 5, 1451 – 1461, 2009.

[2] M. H. Zaharia , F. Leon, C. Pal, and G. Pagu, “Agent-Based Simulation of Crowd Evacuation Behavior”, International Conference on Automatic Control, Modeling and Simulation, 529-533, 2011.

[3] C. T. Quang, and A. Drogoul, “Agent-based simulation: definition, applications and perspectives”, Invited Talk for the biannual Conference of the Faculty of Computer Science, Mathematics and Mechanics, 2008.

[4] Z. Bo, and V. Satish, “Agent-based modeling for household level hurricane evacuation”, Winter Simulation Conference, 2009.

[5] J. W. Cole, C. E. Sabel, E. Blumenthal,K. Finnis, A. Dantas,S. Barnard, and D. M. Johnston, “GIS-based emergency and evacuation planning for volcanic hazards in New Zealand”, Bulletin of the New Zealand society for earthquake engineering, vol. 38, no. 3, 2005.

[6] M. Batty, “Agent-Based Technologies and GIS: simulating crowding, panic, and disaster management”, Frontiers of geographic information technology, chapter 4, 81-101, 2005

[7] T. Patrick, and A. Drogoul, “From GIS Data to GIS Agents Modeling with the GAMA simulation platform”, TF SIM 2010.

[8] J. Kaprzy Edt., Kohei Arai, Rescue System for Elderly and Disabled Persons Using Wearable Physical and Psychological Monitoring System, Studies in Computer Intelligence, 542, 45-64, Springer Publishing Co. Ltd., 2014.

[9] K.Arai, T.X.Sang, Emergency rescue simulation for disabled persons with help from volunteers, International Journal of Research and Review on Computer Science, 3, 2, 1543-1547, 2012.

[10] K.Arai, Wearable healthy monitoring sensor network and its application to evacuation and rescue information server system for disabled and elderly person, International Journal of Research and Review on Computer Science, 3, 3, 1633-1639, 2012.

[11] K.Arai, T.X.Sang, N.T.Uyen, Task allocation model for rescue disable persons in disaster area with help of volunteers, International Journal of Advanced Computer Science and Applications, 3, 7, 96-101, 2012.

[12] Kohei Arai, Tran Xuan Sang, Decision making and emergency communication system in rescue simulation for people with disabilities, International Journal of Advanced Research in Artificial Intelligence, 2, 3, 77-85, 2013.

[13] Kohei Arai, Frequent physical health monitoring as vital sign with psychological status monitoring for search and rescue of handicapped, disabled and elderly persons, International Journal of Advanced Research in Artificial Intelligence, 2, 11, 25-31, 2013.

[14] Kohei Arai, Vital sign and location/attitude monitoring with sensor networks for the proposed rescue system for disabled and elderly persons who need a help in evacuation from disaster areas, International

Journal of Advanced Research in Artificial Intelligence, 3, 1, 24-33, 2014.

[15] Kohei Arai, Method and system for human action detection with acceleration sensors for the proposed rescue system for disabled and elderly persons who need a help in evacuation from disaster areas, International Journal of Advanced Research in Artificial Intelligence, 3, 1, 34-40, 2014.

[16] Kohei Arai, Method and system for human action detection with acceleration sensors for the proposed rescue system for disabled and elderly persons who need a help in evacuation from disaster areas, International Journal of Advanced Research in Artificial Intelligence, 3, 1, 34-40, 2014.

[17] Kohei Arai, Taka Eguchi, Realistic rescue simulation method with consideration of roadnetwork ristrictions, International Journal of Advanced Research on Artificial Intelligence, 4, 7, 21-28, 2015.

[18] K.Arai, T.X. Sang, Multi agent based rescue simulation for disabled persons with the help of volunteers in emergency situation, Proceedings of the 260th conference in Saga of Image and Electronics Engineering Society of Japan, 81-85, 2012.

[19] Kohei Arai, Rescue system with sensor network for physical and psychological health monitoring, Proceedings of the International Seminar on Intelligent Technology and Its Applications: ISITIA 2015, (Keynote Speech), 2015.

[20] Kohei Arai, Rescue system with sensor network for physical and psychological health monitoring, Proceedings of the SAI Confeernce 2015

AUTHORS PROFILE

Kohei Aarai 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 33 books and published 510 journal papers. He is Editor-in-Chief of International Journal of Advanced Computer Science: IJACSA and Applications as well as International Journal of Intelligent Systsems and Applications: IJISA. http://teagis.ip.is.saga-u.ac.jp/index.html

Multiple-Language Translation System Focusing on Long-distance Medical and Outpatient Services

Rena Aierken and Li Xiao
The Xinjiang Technical Institute of Physics & Chemistry Academy of Sciences, Urumqi, China
University of Chinese Academy of Science,

Su Sha and Dawa Yidemucao
Key Laboratory of Xinjiang Multi-Language Technology, Xinjiang University, Urumqi, China
Department of information science and engineering, Xinjiang University, Urumqi, China

***Abstract*—For people living in the countryside, an effective long-distance medical and health service is very important. People living in western China, especially, require convenient communication in their native language with doctors working in a modern city. To address this problem, a multiple-language translation system for long-distance medical and outpatient services is discussed. This system initially provides a table containing basic information including disease names and symptoms for different medical classifications, and then translates the sentences selected from the table automatically using a machine translation system. Finally, a PDF file is created for the doctor and the patient. In this paper, the system construction and evaluation of the machine translation are introduced.**

Keywords—questionnaire for outpatient cases; Chinese Uyghur language; medical Chinese-Minority language parallel corpus; statistical machine translation.

I. INTRODUCTION

Long-distance medical and health services via internet is extremely beneficial for people living in the countryside. A large percentage of people living in the countryside of Xinjiang cannot speak Mandarin or other languages, while most medical practitioners or doctors do not know the local language. Thus, there is a significant communication-gap between the doctors and the patients. Although a manual translation system could be a possible solution, it costs time and money, besides increasing the burden on patients [1].

A study is going for overcoming the communication problems existing in the medical and health field by building a translation system. The system was built in three steps: first, we built a high-quality Chinese and Uyghur language corpora focusing on the various medical terms in collaboration with medical and health institutions. The corpora were carefully collected by hospital clinics and professionals from medical universities. Next, a set of parallel sentences (PS) including a larger volume of words used in the medical and health field was created. Finally, a translation system that searches the PS and a statistical machine translation (SMT) method were implemented.

When using the translation system, a patient first answers questions in his/her native language. The results of the translation obtained by searching through the PS followed by an SMT are then displayed on the table. Next, the patient selects the most suitable result from the translation result by clicking on the table buttons. Finally, a PDF file that will be used as an electronic medical record (EMR) is generated by the system automatically.

The remainder of this paper is organized as follows. The related works are presented in section 2, and discuss the overall translation system in section 3. Section 4 shows the experimental setup and analyzes the experimental results. The conclusions and future work will be presented in section 5.

II. RELATED WORKS

Until now, there have been only limited reports on domestic multilingual translation systems for health care research and development. In our previous study, we have reported an electronic medical record translation system that used the strategy of statistical and example-based machine translation for the Chinese and Uyghur languages [2]. The research group of the Ministry of Japan and Ikeda, reported a multi-language support system using a practical text-set for outpatient services [3,4]. They also confirmed that the system performance deteriorates with a larger number of unknown words. Collection of all the relevant words and dialogs in a native language is considerably difficult. In this study, we report a high performance machine translation system focused on the outpatient service. Our system is built based on bilingual parallel sentences in Chinese and Uyghur, relevant to the medical field and the statistical machine translation method.

III. PREPARE YOUR PAPER BEFORE STYLING

A. Questionnaire for the Outpatient Service

An outpatient doctor may provide a patient's condition survey table, as shown in Figure 1, to the patient. The patient answers each question in his/her native language. For demonstration illustration purposes, a Uyghur Latin alphabet-based questionnaire is presented.

To use the system, a patient first answers the question as prompted, entry, depicted in Figure 1 (2) and Figure 1 (3). Next, when the patient clicks on the appropriate buttons, depicted in Figure 1 (2) or Figure 1 (4), the system will display an interface, as shown in Figure 2, automatically, according to patient's input, in two languages (the patient's language and the doctor's language). Column (3) of Figure 2 shows the PS search results and column (4) of Figure 2, shows

Sponsor: National Natural Science foundation of China (No. 61163030)

the MT result for an entry in column (2) of Figure 2. The PS search results are shown in column (3) of Figure 2, obtained by an N-gram (2 gram or 4 gram) statistical language model and a maximum likelihood criterion. In this case, the system can provide three matching results. If the instance or the input string does not exist in the PS, the system does not display the results. Column (4) of Figure 2 shows the result of the MT. the most suitable result from the translation result by clicking

Common machine translation systems may generate mistranslation results. Therefore, column (5) of Figure 2 is highlighted with a red string display, reminding the user that the machine translation results may have a translation error.

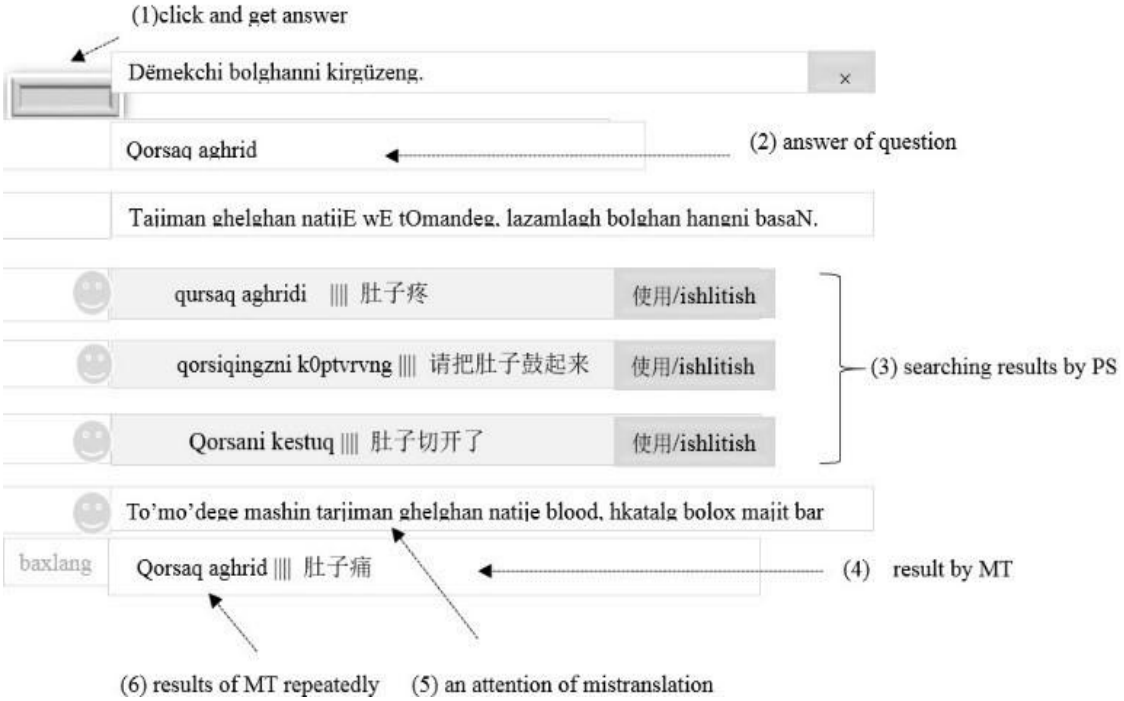


Fig. 1. The complete interface

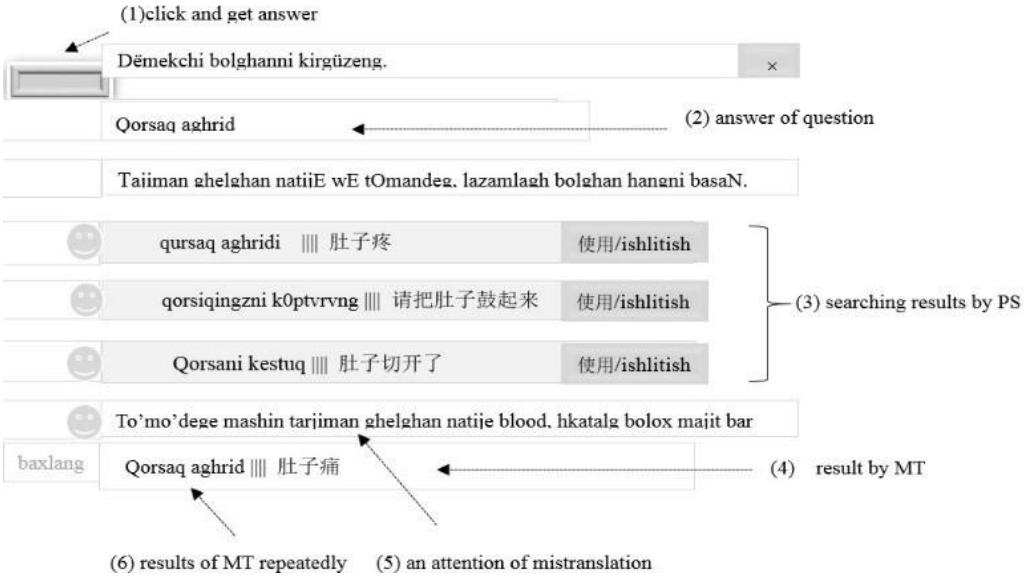


Fig. 2. Translation system

In addition, column (4) of Figure 2 shows the results of the original input strings that are repeatedly translated. Thus, the patient can repeatedly modify the original input, namely column (2) of Figure 2, and choose the most satisfactory translation result to improve the accuracy of the machine translation. The MT provides only one result and the PS gives three, hence, the translation system in Figure 2 can display four different translation results. Finally, the patient can select and click column (3) or column (4) of Figure 2 according to the actual situation, using the right button "use." Then the system prints a PDF file in two languages. Finally, both the doctor and the patient can save the PDF file for electronic medical records.

B. Parallel Sentence and Statistical Machine Translation

This section describes the PS system and the decoder of the machine translation. 240K Chinese sentences including medical, medicine, and drug-names were developed in our previous research. Then, sentences in the Uyghur and Kazak languages corresponding to these Chinese sentences were added manually. Finally, a multi-lingual, aligned medical corpus was created for the system test [5]. We applied a machine translation software based on the general Moses software, developed by the Key Laboratory of multiple language information technology of the Xinjiang University [6,7,8].

IV. EXPEXPERIMENTAL RESULTS

In order to confirm the effectiveness of the questionnaire, we investigated the test results of the PS and the SMT. Then, we tested the performance using an integrated system (PS and SMT).

A. Test Conditions

Ten students, interested in science and engineering, and proficient in Chinese and Uyghur, participated in the experiment. They were asked to enter their own experience of the pathology and relevant medical information. Six students, who specialized in medicine, checked the results of the translation.

To increase the usability of the system, the source language is set to Uyghur and target language is set to Chinese. Additionally,

1) We guaranteed confidentiality in this study. Private information was not revealed.

2) Each participant filled both the investigation form and the system-provided form. These two forms were filled with the same content and in the same order.

3) Taking into account the user's familiarity with the system operation, the two forms were repeated twice. Table 1 shows examples of the questions. Answers were filled in both a paper-form and a system-provided form. The experiment was carried out with native speakers of the local language (Uyghur). The system assumes that 258 pairs of sentences can be used in the experiment.

TABLE I. QUESTIONNAIRE CONTENTS

No.	So'al sorash mazmi (Question content)	o'al sorash shekli (Question form)
1	qandaq Alamet körulwatidu?	ixtiyari kirgüzüsh
2	qachandin bashlap?	
3	iqki takxurux ainiki bilen tegxarganma?	qemni yaki kunupkini chëkip kirgüzüng
4	Qandaq dora yedingiz?	ixtiyari kirgüzüsh
5	dorini hazirmu yemsiz?	

B. Test by Searching the PS only

In this experiment, we investigate two kinds of test results. Figure 3 shows a PS set search result and the MT for question No. 1 in Table Ⅰ. It is clear that the three results shown in (3) of Figures 2, 3, and 4, by searching the PS alone, are not better than the result of the MT shown in (4) of Figures 2, 3, and 4, for the answer string /*axqazan mijaz yaqxiemes* (meaning: my stomach is upset) /. This is because the answer string contains commonly used words and does not contain medical terms. We can observe anther translation result for question No. 3 in Table Ⅰ, shown in figure 4. Here, an answer for question No. 3, /iqki takxurux ainiki tegxarixni qlghan/, contains medical terms, hence, the PS search results, (3) of Figures 4 is considerably better than that of the MT, (4) of Figures 3 and 4.

C. Tests using MT only

In this experiment, we assume that the machine translation is an independent implementation. Each sentence that appears in MT is compared with the results. Table Ⅱ shows the results of the 10 sentences using the PS search. The column, results of the PS (Chinese), includes results of translation by the PS for the source language (Uyghur).

The column, results of the MT, includes results by the MT for the source language (Uyghur). The column, indicating correctness, shows a manual evaluation of the MT for the source language (Uyghur). In this experiment, a total of 20 sentences were selected for the test and 10 sentences were evaluated after excluding repetitive sentences. From Table Ⅱ, it can be seen that there are four mistranslations among the 10 sentences or four of the translation results do not match the source language. In addition, for a sentence with ID:1 in Table Ⅱ, although the result using MT and the result of the PS are close in meaning, the result by MT does not answer the question "What are the symptoms?", hence, it is judged as an incorrect result. From the above discussions, it is obvious that

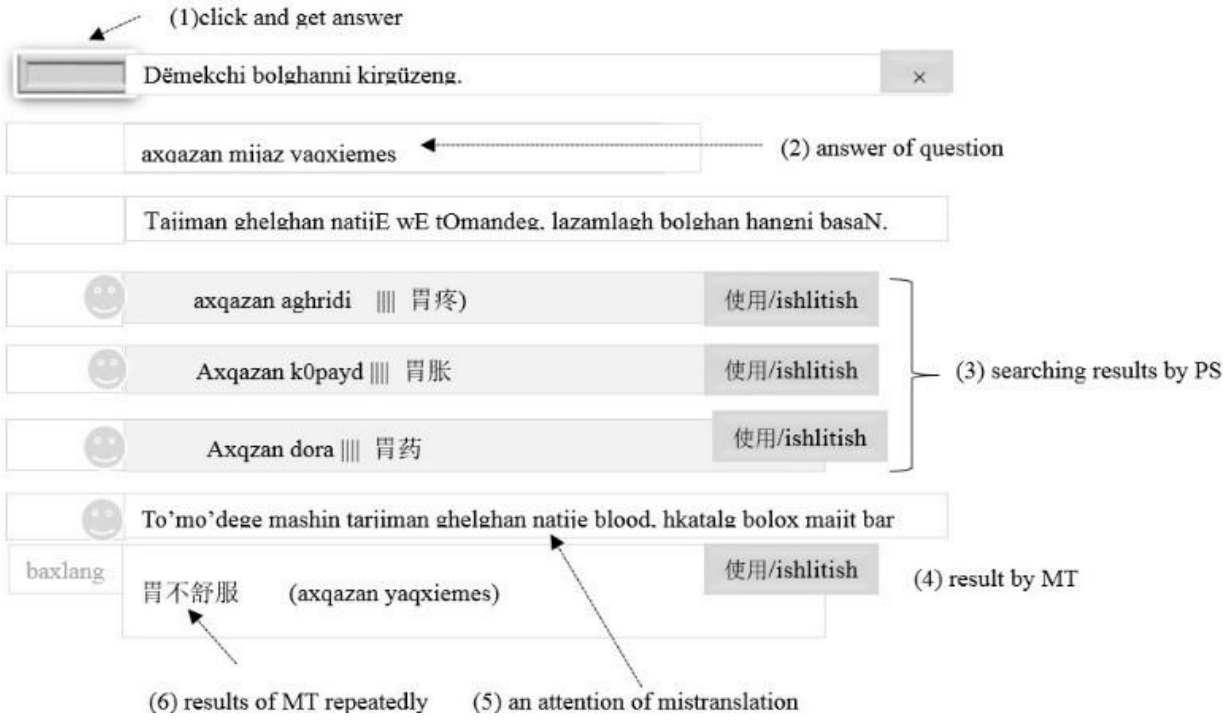


Fig. 3. A translation result for question No. 1

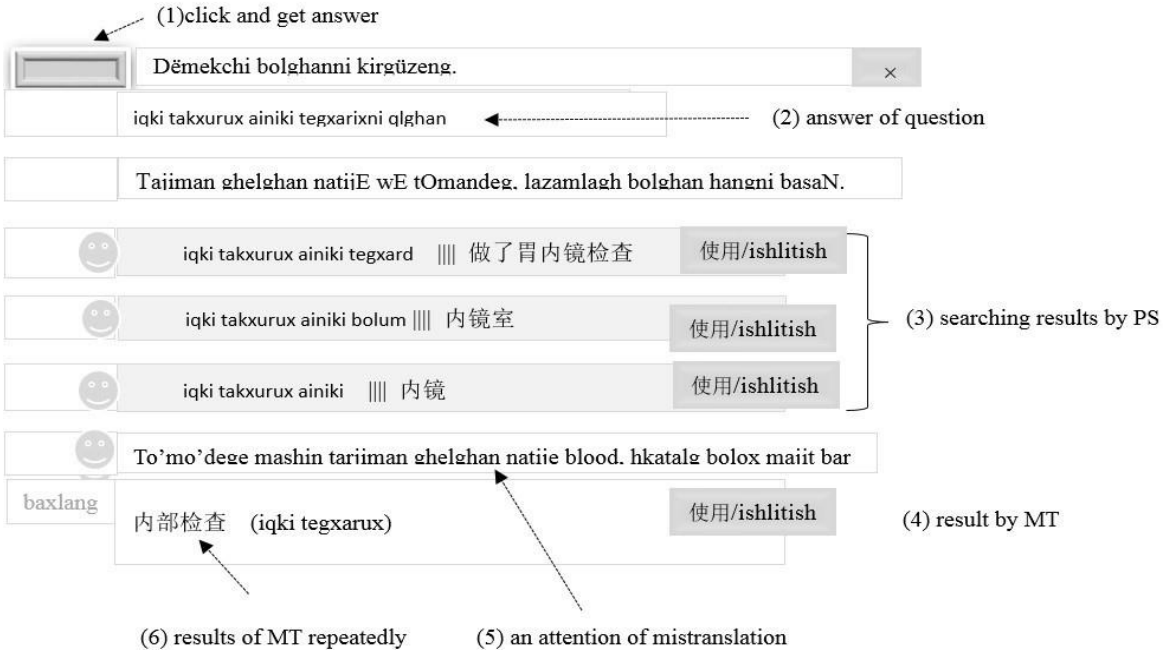


Fig. 4. Translation result for question No. 3

it is difficult to collect all the native language information used in the medical field and the machine translation technique cannot ensure the translation accuracy for the special terms used in the medical and health fields. If the mistranslation by machine translation is to be corrected using the PS search, then, the proposed method of merging both the PS and the SMT is feasible and complete. A real test EMR is shown in Appendix A.

TABLE II. EXAMPLES OF THE SENTENCES USED IN THE EXPERIMENT

ID	source language (Uyghur)	results of PS (Chinese)	results of MT (Chinese)	correctness
1	Qizitma örlesh	发烧 (fever)	有发热 (have hot)	x
2	Bash aghrish	头疼 (headache)	头疼 (headache)	○
4	Müre sirqirap aghrish	肩膀酸疼 (shoulder ache)	肩膀很疼(shoulder is very painful)	x
8	Zukam	感冒 (have a cold)	感冒 (have a cold)	○
21	Qorsaq aghrish	肚子疼(collywobbles)	肚子疼 (collywobbles)	○
47	Dora rë'aksiye qilish	食物过敏(food allergy)	食品有过敏(food allergies)	x
50	Bash qëyish	头昏(dizzy)	头昏(dizzy)	○
60	Ishtiha yoq	没有食(loss of appetite)	没有食欲 (loss of appetite)	○
72	Nepeste qëlinish	呼吸困难(dyspnea)	呼吸困难 (dyspnea)	○
83	Ashqazan qattiq aghris	胃疼的厉害(have a stomachache)	有时胃疼(sometimes a stomachache)	

V. CONCLUSIONS

In this paper, we have presented a multiple-language translation system that focuses on long-distance medical and outpatient services. The system merges the PS and the SMT approaches and translates the dialogues of the doctors and patients automatically, to create an electronic medical record (EMR). Through experiments, we confirmed that the proposed method is feasible and practical.

In the future, we will further improve the system accuracy and extend the system to service more languages, such as languages along Silk Road.

REFERENCES

[1] People government of Xinjiang Uyghur Autonomous Region, China. Study and publicize the article implement the second central work conference.http://www.xj.xinhuanet.com/

[2] Dawa Idomucao. A Study on Chinese Minority Medical Document Translation Based on Hybrid-strategy [J]. Journal of Xinjiang University (Natural Science Edition). 2015. Vol. 32(1). 123-128.

[3] G. Zhenyi. Establishment multi-language Medical Outpatient registered system for foreign patients [J], Jornal of the Japan Institute of electronic information and communication. 2009, Vol.192-D No.6 708-718.

[4] Ikeda, T, Ando, S, Satoh, K Okumura, A. and Watanabe., T, Aoutomatic interpretation system integrating Free-style Sentence translation and Parallel text based Translation[C], Prec.Workshop on speech-to-speech translation 2002,85-92.

[5] Dawa Idomucao. Research based on multi-language machine translation system with electronic medical record [M], 2012, National Natural Science and Technology Fund: Number:61163030.

[6] Hieu Hoang and Philipp Koehn. Design of the moses decoder for statistical machine translation. n Proceedings of ACL Workshop on Software engineering, testing, and quality assurance for NLP, pp. 58–65, 2008.

[7] Philipp, Koehn, Statistical Machine Translation[M], Uk, Cambridge University,20

[8] YANG Pan, LI Miao,ZHANG Jian. Chinese-Uyghur ranslation system for phrase-based statical translation [J], Journal of Computer Application 2009，29(7),2022-2025.

Appendix A:

医院/：دوختۇرخانا　　病类科/：بۆلۈم

<<电子病历 EMR>>　　2015 年 10 月 07 日

==========

患者姓名/: بىمار ئىسمى

出生年月日/:تۇغۇلغان يىل ئاي كۈن

民族/:مىللىتى

住址/：ئادىرىس

联系电话/:تېلفون

病情调查表/：كېسەللىك ئەھۋالى تەكشۈرۈش جەدۋىلى

医生提问? دوختۇر سورىغان سوئال؟	患者选择回答 بىمار تاللىغان جاۋاب	机器翻译结果/ ماشىنا تەرجىمە نەتىجىسى	医生确认结果/ دوختۇر تەكشۈرۈش نەتىجىسى
1. 你有什么症状？/ سىزدە قانداق ئەھۋال كۆرۈلىدۇ؟	بېشىم ئاغرىيدۇ	头疼|||باش ئاغرىش	√
2.什么时候开始/ قاچان باشلانغان؟	2-3 كۈن	2-3 天|||2-3 كۈن؟	√
3.你有高血压吗/ سىزدە يۇقىرى قان بېسىم بارمۇ؟	بار	有/بار	√
4.你现在还服降压药物吗？/ سىز ھازىر قان بېسىم چۈشۈرۈش دورىسى ئىچەمسىز؟	بېقىندىن بېرى دورا ئىچمىگەن	近期没有服药|||دورا ئىچمىگەن	×
5.其他还有什么症状？/ يەنە قانداق ئالاھىتلەر بار	دىئابىت كېسىلىم بار	患有糖尿病|||دىئابىت كېسىلگە گىرىپتار بولغان	√
6.对于糖尿病服用了什么药物？ تەتۇر سىيىش كېسىلگە نېمە دورا ئىشلىتۋاتىسىز؟	ئىنسۇلىن	胰岛素|||ئىنسۇلىن	√
7.你睡眠好吗？/ ئۇيقىڭىز ياخشىمۇ؟	ئانچە ياخشى ئەمەس	不怎么好|||ياخشى ئەمەس	√
8.现在有降压药物吗?/ قان بېسىم چۈشۈرۈش دورىسى بارمۇ؟	ھازىر يوق	目前没有|||ھازىرچە يوق	√
9.服用几天的降压药物好吗/ بىرنەچچە كۈنلۈك قان بېسىم چۈشۈرۈش دورىسى ئىشلەتسىگىز بولامدۇ؟	بولىدۇ	吃吧|||دورا يەي	√
10.吃完降压药物再来检查可以吗?/ قان بېسىم چۈشۈرۈش دورىسى ئىچىپ بولۇپ يەنە تەكشۈرۈپ باقسىگىز قانداق؟	بولىدۇ	可以|||بولىدۇ	√

医生门诊结果填写/：دوختۇر ئامبۇلاتورىيە نەتىجىسى تولدۇرۇلىدۇ

1） 患者症状/：كېسەل ئەھۋال

2） 门诊结果/：دىئاگنوز نەتىجىسى

3） 用药情况/：دورا ئىشلىتىش ئەھۋالى

医生签字（盖章）

دوختۇر ئىمزاسى(تامغا)/

Hybrid Intelligent Approach for Predicting Product Compositions of a Distillation Column

Yousif Al-Dunainawi
Electronic and Computer Engineering Department
College of Engineering, Design and Physical Sciences
Brunel University London
Uxbridge, London, UK

Maysam F. Abbod
Electronic and Computer Engineering Department
College of Engineering, Design and Physical Sciences
Brunel University London
Uxbridge, London, UK

***Abstract*—Compositions measurement is a vitally critical issue for the modelling and control of distillation process. The product compositions of distillation columns are traditionally measured using indirect techniques via inferring tray compositions from its temperature or by using an online analyser. These techniques were reported as inefficient and relatively slow methods. In this paper, an alternative procedure is presented to predict the compositions of a binary distillation column. Particle swarm optimisation based artificial neural network PSO-ANN is trained by different algorithms and tested by new unseen data to check the generality of the proposed method. Particle swarm optimisation is utilised, here, to choose the optimal topology of the network. The simulation results have indicated a reasonable accuracy of prediction with a minimal error between the predicted and simulated data of the column.**

***Keywords*—Hybrid Intelligence; Prediction; Distillation Column; Neural network; Particle swarm optimisation**

I. Introduction

Over the past few decades, great developments in online analysis, monitoring and measurement of dynamic processes were made in various applications [1]–[5]. This development is partially motivated by the desire to improve quality. Unambiguously, quality is a significant indication that has a substantial impact on productivity and economy of manufacture, particularly in the field of mass production [2]. Direct measurement of the product compositions of the distillation column is a crucial issue. However, its disadvantages at stream process lie in difficulty, unreliability and high capital and operational cost. These disadvantages will have an exponentially negative effect when more than one analysis are needed to obtain a clearer picture of the different streams involved. Consequently, indirect and inferential measurement techniques are being used to design and run many distillation columns. These columns operate widely in chemical and petrochemical plants as well as refineries to separate mixtures into their individual components.

Not only non-linearity and transit behaviour make distillation, as a process, complicated to control, but the product compositions also cannot be promptly measured, nor reliable. The delay caused by the measurement and analysis of compositions will negatively affect the effectiveness and robustness of control [6]. An indirect method is proposed to monitor the products compositions of the column by using tray temperature inside the column, albeit this feature is an unreliable indicator of product compositions [7], [8]. Moreover, other considerations like consistent maintenance, regular calibrations, and high-cost equipment make composition analysers an ineffective solution for precise online measurements. Consequently, soft sensing or inferential systems have been proposed recently as practical options to replace hardware measurement systems [9].

Artificial neural networks (ANNs) is one of the most attention-grabbing branches of artificial intelligence, which has grown rapidly in the recent years as an optimal solution for the modelling, and prediction of dynamic systems. ANNs have shown outstanding performance to learn the input-output relationship of nonlinear and complex systems. This relationship could be easily, quickly and efficiently found out via reducing the error between the network output(s) and the actual output(s). After the network is trained, the output can be predicted within few seconds. ANN-based models are still being applied successfully to overcome engineering problems in different fields such as adaptive control, pattern recognition, robotics, image processing, medical diagnostics, fault detection, process monitoring, renewable and sustainable energy, laser applications and nonlinear system identification [10]–[17].

The most crucial task which faces the neural network constructer is the proper selection of the network topology to solve a particular problem. The topology means, here, the number of nodes (neurones) and the number of layers in the hidden zone. Therefore, one of the most efficient methods to determine the optimal network structure is evolutionary algorithm EA methods such as genetic algorithm GA [18] and practical swarm optimisation PSO [19] and so on.

This paper proposed a PSO-based neural network as a predictor model for estimating the product compositions of a binary distillation column.

II. Distillation Column Modelling and Description

Distillation is, undoubtedly, one of the most important processes in chemical and petrochemical plants. Distillation columns are used as separators of chemical compounds in petroleum, natural gas, liquid and chemical industries [20]. The major disadvantage of using those columns is that they are considered as an intensive energy process. A report from the US Department of Energy has indicated that distillation is the

largest consumer of energy in the chemical industry; typically, it accounts 40% of the energy consumed by petrochemical plants. Despite its "thirst" for energy, distillation persists to be a widely utilised method for separations [21], [22].

Figure 1 is a schematic diagram of a binary distillation column, in which a feed mixture is separated into a distillate product (overhead) and a bottom product. Also, heat is transferred into the column via a reboiler (heat exchanger) to vaporise some of the liquid from the base of the column. The vapour travels up through trays inside the column to reach the top and, then, comes out to be liquefied in a condenser. Liquid from the condenser, at that point, drops into the reflux drum. Finally, the distillate is removed from this drum as a pure product. additionally, some liquid (reflux) is fed back near the top of the column while the impure product is produced at the bottom outlet.

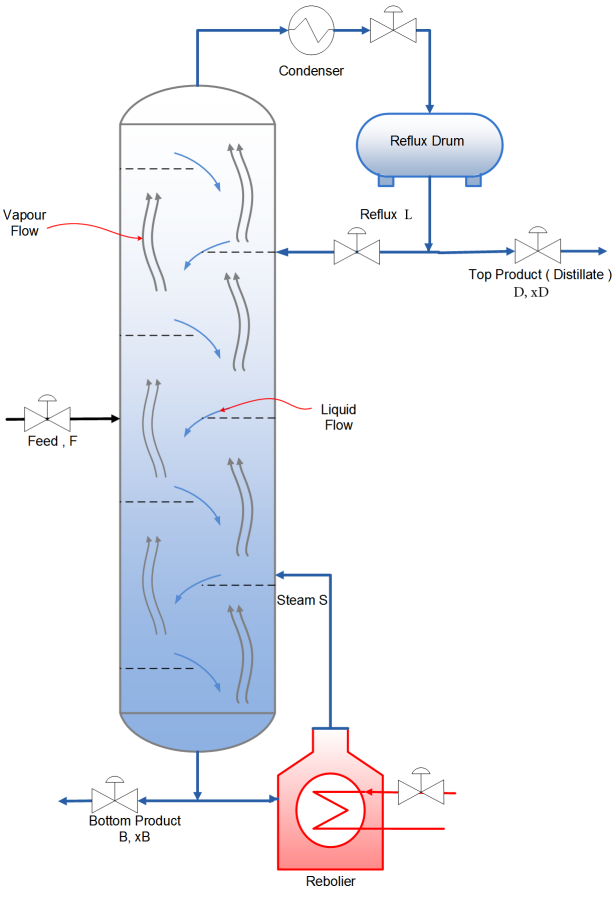


Fig. 1: Schematic diagram of a binary distillation column

The dynamic model can be simplified under the following assumptions:

- No chemical reactions occur in all stages of the column
- Constant pressure (open to atmosphere pressure)
- Binary mixture
- Constant relative volatility
- No vapour hold-up occurs
- Perfect mixing and equilibrium for vapour-liquid on all stages

While the operating conditions and technical aspects of the distillation column are detailed in the appendix at the end of this paper.

Accordingly, the mathematical expression of the model can be represented with the assumptions by the following equations: On each tray (excluding reboiler, feed and condenser stages):

- On each tray (excluding reboiler, feed and condenser stages):

$$M_i \frac{dx_i}{t} = L_{i+1}x_{i+1} + V_{i-1}y_{i-1} - L_i x_i - V_i y_i \quad (1)$$

- Above the feed stage $i = NF + 1$:

$$M_i \frac{dx_i}{t} = L_{i+1}x_{i+1} + V_{i-1}y_{i-1} - L_i x_i - V_i y_i + F_V y_F \quad (2)$$

- Below the feed stage $i = NF$:

$$M_i \frac{dx_i}{t} = L_{i+1}x_{i+1} + V_{i-1}y_{i-1} - L_i x_i - V_i y_i + F_L X_F \quad (3)$$

- In the reboiler and column base $i = 1$, $x_i = xB$:

$$M_B \frac{dx_i}{t} = L_{i+1}x_{i+1} - V_i y_i + Bx_B \quad (4)$$

- In the condenser, $i = N + 1$, $xD = xN + 1$:

$$M_D \frac{dx_i}{t} = V_{i-1}y_{i-1} - L_i x_D - Dx_D \quad (5)$$

- Vapour-liquid equilibrium relationship for each tray [23]:

$$y_i = \frac{\alpha x_i}{1 + (\alpha - 1)x_i} \quad (6)$$

- The flow rate at constant molar flow:

$$L_i = L, Vi = V + F_V \quad (7)$$

since

$$F_L = q_F \times F \quad (8)$$

$$F_v = F + F_L \quad (9)$$

- The flowrate of both condenser and reboiler as: Reboiler:

$$B = L + F_L - V \quad (10)$$

Condenser:

$$D = V + F_V - L \quad (11)$$

- The feed compositions xF and yF are found from the flash equation as:

$$F_{zF} = F_L \times x_F - F_V \times y_F \quad (12)$$

III. Hybrid Model Development and Optimisation

Recently, hybridization or combination of different learning and adaptation techniques has been employed to a large number of new intelligent system designs. The main aim of integrating these techniques is to overcome individual limitations and to achieve synergetic effects [24].

Therefore, a PSO-based artificial neural network is proposed as an estimator tool of a binary distillation column.

A. Artificial Neural Networks

Artificial neural network (ANNs) is a complicated system, which is composed of numerous neural nets. These nets fundamentally based on the principal understanding of the function, structure, and the mechanism of the brain of humankind [25]. In the last two decades or so, ANNs have been applied to a widespread range of applications due to their ability to analyse and capture the complexity and nonlinearity features of dynamic processes. One of the major applications of ANNs is a modelling or identification process of complex systems [26].

It is worth mentioning that the topology of the network is a crucial matter where as choosing the number of the neurones and layers in the hidden zone is not an easy task. So far, no systematic approach or automatic methods have been used to tackle this issue. Because the network structure depends on the nature and features of the process that would be modelled. Therefore, there are, probably, only two research methods to select from, a blind or heuristic. The blind approach, or trial and error, is an unguided and arbitrary search method, to which all possible alternatives are applied to find the optimal solution. Although this technique can eventually find the optimal ANN topology with limited search space, this method is not practical because it is considered highly expensive in terms of time and computations.

B. Particle Swarm Optimisation

Ever since particle swarm optimisation (PSO) has been proposed by Kennedy and Eberhart in 1995 [27] and 2001 [28], PSO algorithm turned to be vastly successful. The several of researchers have presented the merit of the implementation of PSO as an optimiser for various applications [29]. In PSO procedure, all individuals or particles (commonly between 10 and 100) are located at a random position and are supposed to move randomly in a defined direction in the search space. Each particle direction is then changed steadily to move assuredly along the direction of its best previous positions to discover even a new better position according to certain criteria or an objective function (fitness). The initial particle velocity and position are selected arbitrarily, and the following velocity equation can update them as

$$Vc_{i+1} = wV_i + C1R1 \times (Pb_i - x_i) + C2R2 \times (Gb - x_i) \quad (13)$$

Whereas the new particle is calculated by adding the previous one to the new velocity as shown in the following equation:

$$x_{i+1} = x_i + Vc_{i+1} \quad (14)$$

where: Vc: velocity of the particle, X: position of the particle, $R1, R2$: independent random variables uniformly distributed in [0, 1], $C1, C2$: acceleration coefficients as well as w: inertia weight. Eq. 13 is used to compute the new velocity of the particle according to its preceding velocity and the distances of its current position from its own best position (Pb) and the global best position (GB). Then, the particle moves to a new place in the search space, according to Eq. 14. The performance of each particle is measured according to a predefined objective function (performance index).

C. Hybrid System Design

Evolutionary-based optimisation, like PSO, can be applied by only simple mathematical operations with a few lines of code [30]. This feature provides a low-cost method concerning both memory and speed requirements. Thus, in this study, PSO is chosen to find the optimal network topology of the prediction model as depicted in Figure 2.

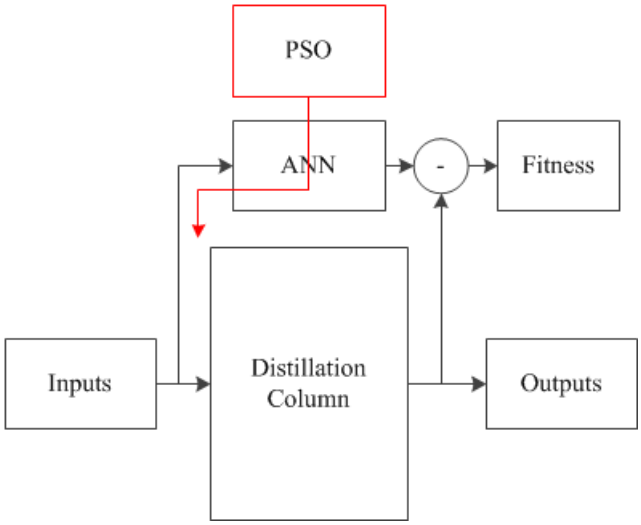


Fig. 2: Schematic diagram of the proposed hybrid system

IV. Simulation and Results

In this study, the reflux (L) and the boil-up (V) flow rates had been used as inputs to the network while distillate and bottom composition chosen as outputs. The dataset implemented for the training, validation and testing of ANN was generated by applying 40 distributed random values, each lasting 50 sampling time for (L) and (V) as shown in Figure 3. The distillate composition was approximately between 0.95 and 1 (mole fraction), while the bottom composition was around 0.005 to 0.12 (mole fraction) and a total of 2000 datasets were collected for identification, Figure 4 presents the simulated data of the column.

The dataset obtained by the simulation was randomly divided into 70%, 15% and 15% for training, validation and testing respectively. Feedforward multilayer network had been implemented to predict the product compositions of the distillation column. In addition, various backpropagation training algorithms, namely; Gradient Descent (GD), Scaled Conjugate Gradient (SCG) and LevenbergMarquardt (LM) were separately applied to decide which one performs better

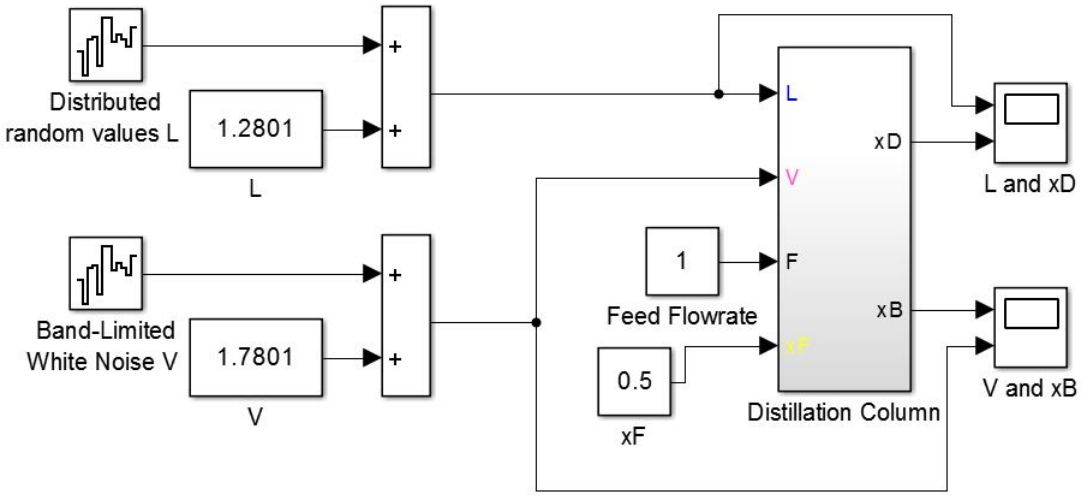
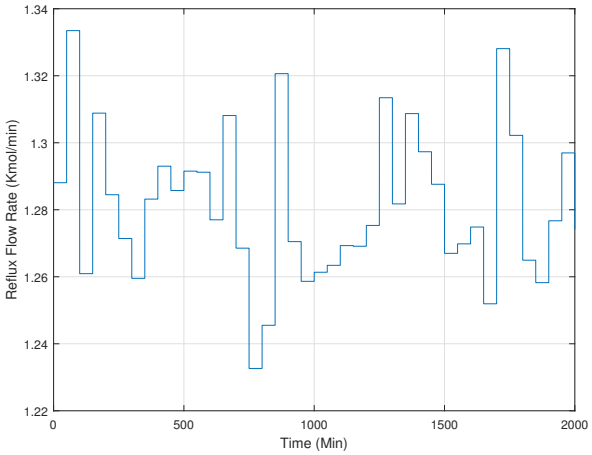
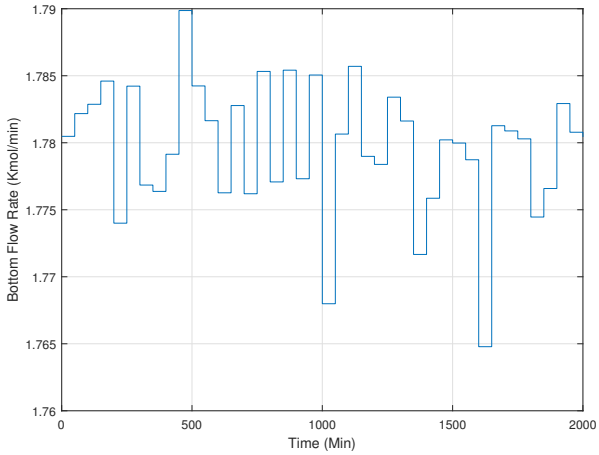


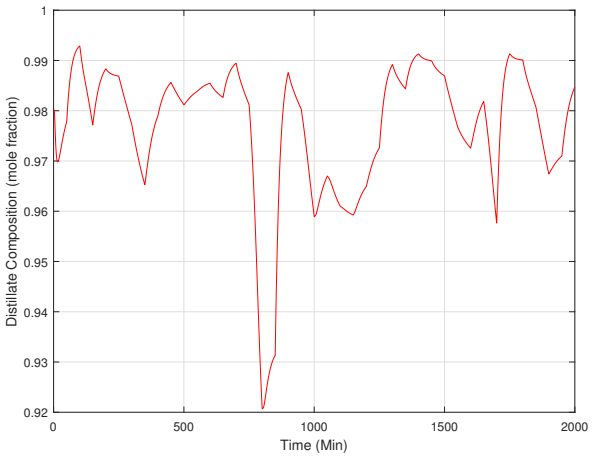
Fig. 3: MATLAB/Simulink model of the distillation column



(a) Reflux folw rate



(c) Boil-up flow rate

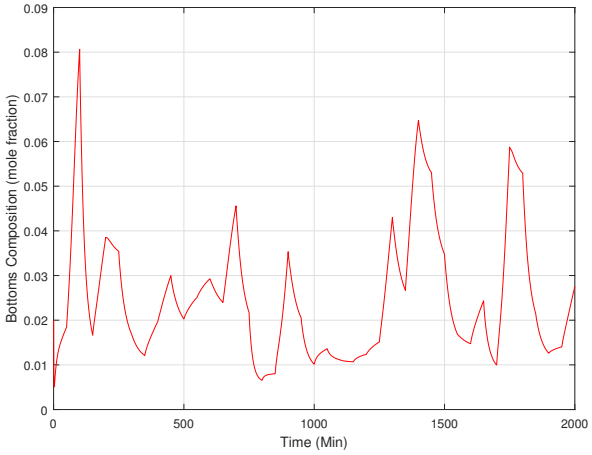


(b) Distillate composition

than the others. Moreover, Log-sigmoid activation function was embedded in the neurones of the hidden layer(s) because of its differentiability. PSO was employed to find the optimal structure of the network, the best operational parameters of PSO algorithm were chosen after extensive simulations and were set as following: For each of the network architecture,

No. of variables (dimensions)	2
Size of the swarm (no. of particles)	50
Maximum iterations (max)	100
Cognitive acceleration ($C1$)	1.2
Social acceleration ($C2$)	0.12
Momentum or inertia (w)	0.9
Minimum search space	1
Maximum search space	25

the training algorithms had been run ten times with different random initial weights and biases. After investigating the performance of different architectures using the PSO technique, a network with two hidden layers (including 23 neurons in the first and 25 in the second) trained by LM algorithm have



(d) Bottoms composition

Fig. 4: Inputs/outputs of the simulation of the distillation column

indicated reasonably good results. Figures 5 and 6 show the performance of the network as a mean square error (MSE) versus the network architecture of the single and double hidden layer respectively. Table I demonstrates the training, testing and validation performances of different training approaches of both one and two layers in the hidden zone.

It is clearly indicated that much better results are found using LM as training algorithm with two hidden layers topology because LM uses Hessian matrix approximation as a second-order method to calculate the change in gradient. Figures 7 and 8 display regression plots of the network outputs on both compositions of training and test sets. For a perfect fit, the data must fall along a 45-degree line, where the network outputs are equal to the targets. The network with two hidden layers trained by LM algorithm, the fit, is reasonably good of both datasets, with R values in each case of 0.99 or above. Checking the test set is importantly required for examining the generalisation of the network of unseen data in the learning stage. It is worth noting that the network training and simulation was performed using MATLAB® and Simulink® platform.

Training Method	No. of Hidden Layers	No. of Neurones	Training MSE	Testing MSE	Validation MSE	Time (sec)
GD	1	4	3.972×10^{-4}	3.952×10^{-4}	3.996×10^{-4}	12.40
	2	$9-2$	2.047×10^{-4}	2.112×10^{-4}	2.015×10^{-4}	11.84
SCG	1	19	1.529×10^{-5}	1.6×10^{-5}	1.5×10^{-5}	12.39
	2	$20-25$	9.392×10^{-7}	9.429×10^{-7}	9.743×10^{-7}	38.39
LM	1	25	1.12×10^{-6}	1.225×10^{-6}	1.195×10^{-6}	18.58
	2	$23-25$	4.365×10^{-9}	3.984×10^{-9}	7.8×10^{-9}	11.37

TABLE I: Performances (MSE) of different PSO-ANN topologies

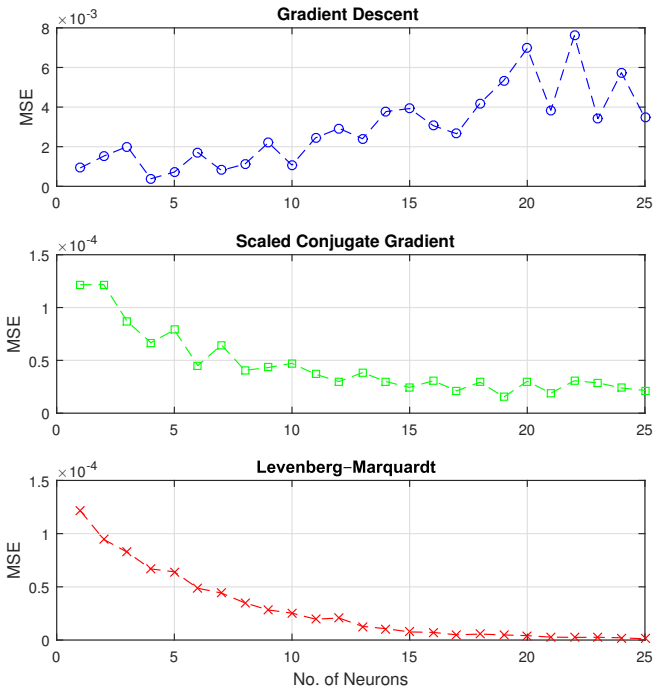
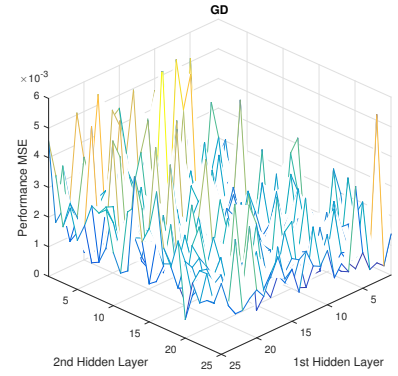
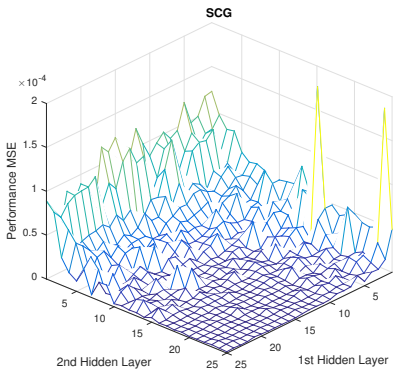


Fig. 5: Performance curves of one hidden layer PSO-ANN of different algorithms





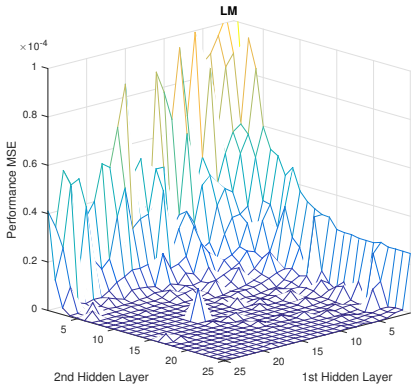
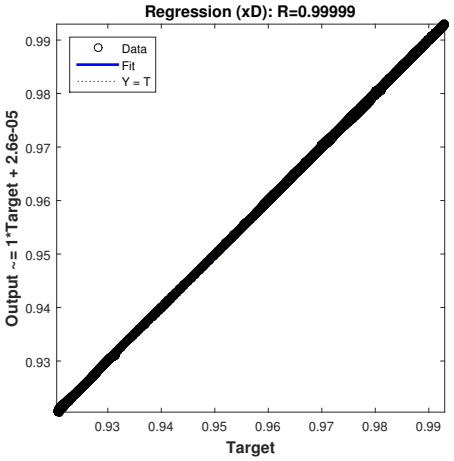
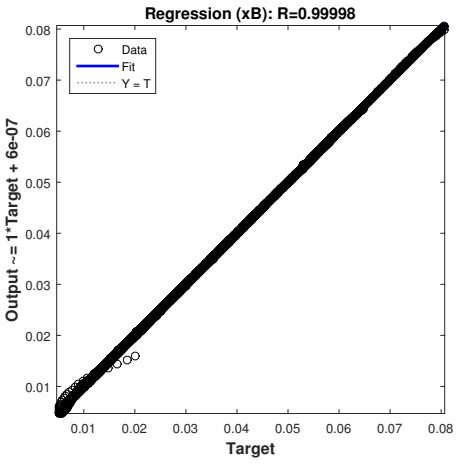


Fig. 6: Performance 3D surfaces of two hidden layers PSO-ANN of different algorithms

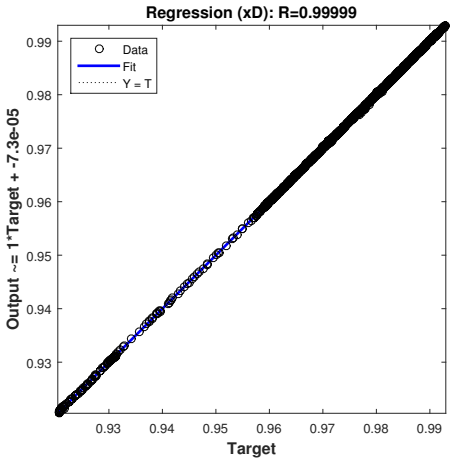


(a) Distillate composition prediction

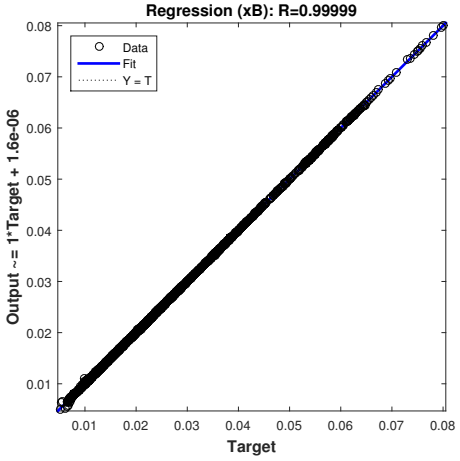


(b) Bottoms composition prediction

Fig. 7: The regression line between predicted and simulated compositions of training set



(a) Distillate composition prediction



(b) Bottoms composition prediction

Fig. 8: The regression line between predicted and simulated compositions of testing set

V. Conclusion

A PSO-based artificial neural network has been proposed as an intelligent prediction approach to estimate product compositions of a binary distillation column; boil-up and reflux were used as inputs to the network. A double layer in the hidden zone with 23-25 neurones architecture was presented the optimal performance of the prediction model after examining different training algorithms and topologies using particle swarm optimisation. the network trained by Levenbergmarquardt algorithm gave more accurate results with less MSE compared to Gradient Descent and Scaled Conjugate Gradient. Therefore, the precision of predicted compositions of the distillation column using LM algorithm has shown to be high, and the estimated compositions have approximately been in agreement with the simulation results. The proposed ANN could be used efficiently to improve the performance of the different neural network controllers like NARMA-L2, direct inverses and NN predictive controller, which mainly depend

on the prediction performance, which is to be the subject of future work.

ACKNOWLEDGMENT

The corresponding author is grateful to the Iraqi Ministry of Higher Education and Scientific Research for supporting the current research.

REFERENCES

[1] R. Teti, K. Jemielniak, G. ODonnell, and D. Dornfeld, "Advanced monitoring of machining operations," *CIRP Annals-Manufacturing Technology*, vol. 59, no. 2, pp. 717–739, 2010.

[2] R. X. Gao, X. Tang, G. Gordon, and D. O. Kazmer, "Online product quality monitoring through in-process measurement," *CIRP Annals-Manufacturing Technology*, vol. 63, no. 1, pp. 493–496, 2014.

[3] A. Vijayakumari, A. Devarajan, and N. Devarajan, "Decoupled control of grid connected inverter with dynamic online grid impedance measurements for micro grid applications," *International Journal of Electrical Power & Energy Systems*, vol. 68, pp. 1–14, 2015.

[4] H. Wang, P. Senior, R. Mann, and W. Yang, "Online measurement and control of solids moisture in fluidised bed dryers," *Chemical Engineering Science*, vol. 64, no. 12, pp. 2893–2902, 2009.

[5] A. K. Pani and H. K. Mohanta, "Online monitoring and control of particle size in the grinding process using least square support vector regression and resilient back propagation neural network," *ISA transactions*, vol. 56, pp. 206–221, 2015.

[6] A. M. F. Fileti, L. S. Pedrosa, and J. A. Pereira, "A self tuning controller for multicomponent batch distillation with soft sensor inference based on a neural network," *Computers & Chemical Engineering*, vol. 23, pp. S261–S264, 1999.

[7] W. Luyben, "Feedback control of distillation columns by double differential temperature control," *Industrial & Engineering Chemistry Fundamentals*, vol. 8, no. 4, pp. 739–744, 1969.

[8] T. Mejdell and S. Skogestad, "Estimation of distillation compositions from multiple temperature measurements using partial-least-squares regression," *Industrial & Engineering Chemistry Research*, vol. 30, no. 12, pp. 2543–2555, 1991.

[9] L. Fortuna, S. Graziani, and M. G. Xibilia, "Soft sensors for product quality monitoring in debutanizer distillation columns," *Control Engineering Practice*, vol. 13, no. 4, pp. 499–508, 2005.

[10] M. A. Hussain, "Review of the applications of neural networks in chemical process controlsimulation and online implementation," *Artificial intelligence in engineering*, vol. 13, no. 1, pp. 55–68, 1999.

[11] N. D. Ramirez-Beltran and H. Jackson, "Application of neural networks to chemical process control," *Computers & industrial engineering*, vol. 37, no. 1, pp. 387–390, 1999.

[12] P. J. Drew and J. R. Monson, "Artificial neural networks," *Surgery*, vol. 127, no. 1, pp. 3–11, 2000.

[13] J. W. Catto, D. A. Linkens, M. F. Abbod, M. Chen, J. L. Burton, K. M. Feeley, and F. C. Hamdy, "Artificial intelligence in predicting bladder cancer outcome a comparison of neuro-fuzzy modeling and artificial neural networks," *Clinical Cancer Research*, vol. 9, no. 11, pp. 4172–4177, 2003.

[14] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Comprehensive review of neural network-based prediction intervals and new advances," *Neural Networks, IEEE Transactions on*, vol. 22, no. 9, pp. 1341–1356, 2011.

[15] F. Amato, A. López, E. M. Peña-Méndez, P. Vaňhara, A. Hampl, and J. Havel, "Artificial neural networks in medical diagnosis," *Journal of applied biomedicine*, vol. 11, no. 2, pp. 47–58, 2013.

[16] S. Mondal, A. Bandyopadhyay, and P. K. Pal, "Application of artificial neural network for the prediction of laser cladding process characteristics at taguchi-based optimized condition," *The International Journal of Advanced Manufacturing Technology*, vol. 70, no. 9-12, pp. 2151–2158, 2014.

[17] M. Maarouf, A. Sosa, B. Galván, D. Greiner, G. Winter, M. Mendez, and R. Aguasca, "The role of artificial neural networks in evolutionary optimisation: A review," in *Advances in Evolutionary and Deterministic Methods for Design, Optimization and Control in Engineering and Sciences*. Springer, 2015, pp. 59–76.

[18] L. M. Lima de Campos, L. de Oliveira, and R. Celio, "A comparative analysis of methodologies for automatic design of artificial neural networks from the beginnings until today," in *Computational Intelligence and 11th Brazilian Congress on Computational Intelligence (BRICS-CCI & CBIC), 2013 BRICS Congress on*. IEEE, 2013, pp. 453–458.

[19] R. K. Agrawal and N. G. Bawane, "Multiobjective pso based adaption of neural network topology for pixel classification in satellite imagery," *Applied Soft Computing*, vol. 28, pp. 217–225, 2015.

[20] M. L. Darby and M. Nikolaou, "Identification test design for multivariable model-based control: an industrial perspective," *Control Engineering Practice*, vol. 22, pp. 165–180, 2014.

[21] R. W. Baker, *Membrane separation systems: recent developments and future directions*. Noyes Publications, 1991.

[22] C. L. Smith, *Distillation control: An engineering perspective*. John Wiley & Sons, 2012.

[23] W. L. Luyben, *Distillation design and control using Aspen simulation*. John Wiley & Sons, 2013.

[24] A. Abraham, "Hybrid intelligent systems: evolving intelligence in hierarchical layers," in *Do Smart Adaptive Systems Exist?* Springer, 2005, pp. 159–179.

[25] Y. Yousif, K. Daws, and B. Kazem, "Prediction of friction stir welding characteristic using neural network," *Jordan Journal of Mechanical and Industrial Engineering*, vol. 2, no. 3, 2008.

[26] G. Dreyfus, *Neural networks: methodology and applications*. Springer Science & Business Media, 2005.

[27] R. C. Eberhart, J. Kennedy *et al.*, "A new optimizer using particle swarm theory," in *Proceedings of the sixth international symposium on micro machine and human science*, vol. 1. New York, NY, 1995, pp. 39–43.

[28] J. Kennedy, J. F. Kennedy, R. C. Eberhart, and Y. Shi, *Swarm intelligence*. Morgan Kaufmann, 2001.

[29] R. Poli, "Analysis of the publications on the applications of particle swarm optimisation," *Journal of Artificial Evolution and Applications*, vol. 2008, p. 3, 2008.

[30] J. Kennedy, "Particle swarm optimization," in *Encyclopedia of machine learning*. Springer, 2011, pp. 760–766.

APPENDIX

Abbreviations, the operating conditions and technical aspects of the distillation column are detailed in following table.

Symbol	Description	Value	Unit
N	Number of trays	20	-
N_F	Feed stage location	11	-
F	Typical inlet flow rate to the column	1	kmol/min
D	Typical distillate flow rate	0.5	kmol/min
B	Typical bottoms flow rate	0.5	kmol/min
zF	Light component in the feed (mole fraction)	0.5	-
q_F	Mole fraction of the liquid in the feed	1	-
L	Typical reflux flow rate	1.28	kmol/min
V	Typical boil-up flow rate	1.78	kmol/min
α	Relative volatility	2	-
xD	Distillate composition (mole fraction)	0.98	-
xB	Bottoms composition (mole fraction)	0.02	-
i	Stage number during distillation	-	-
x	Mole fraction of light component in liquid	-	-
y	Mole fraction of light component in vapour	-	-
M	Tray hold-up liquid	0.5	kmol
MD	Condenser hold-up liquid	0.5	kmol
MB	Reboiler hold-up liquid	0.5	kmol