Application of Future Reference Sentence Extraction in Support of Future Event Prediction

Yoko Nakajima† **Michal Ptaszynski‡** **Hirotoshi Honma†** **Fumito Masui‡**

†Department of Information Engineering
Kushiro National College of Technology
2-32-1 Otanoshike, Kushiro, 084-0916, Japan
{yoko,honma}@kushiro-ct.ac.jp

‡Department of Computer Science
Kitami Institute of Technology
165 Koen-cho, Kitami, 090-8507, Japan
{ptaszynski,f-masui}@cs.kitami-it.ac.jp

Abstract

One of the main tasks in the field of Artificial Intelligence (AI) is providing means for understanding of natural language. The understanding is often restricted to analysis of specific general meaning of particular instances of language behavior (sentences, documents, etc.). For example, in subfields of AI such as natural language processing, or sentiment analysis, speaker attitudes and emotions expressed in a sentence are in the focus of analysis. A different kind of task from the field of natural language processing, which we focus on in this research, is predicting trends of future events. From the point of view of understanding natural language it is a challenging task, since it assumes extracting information regarding future unfolding of events from the provided instances of language behavior. To make the task feasible we limited the focus of our research to detecting and extracting sentences which refer to the future as potentially the most useful in predicting future unfolding of events. We assumed that future reference sentences can contain such information as background of events, specific expert knowledge, etc., and thus can be useful in future prediction. We propose automatic extraction method for future reference sentences from news corpora using semantic and morphological information. Furthermore we discuss the experiment in which the sentences extracted with the proposed method are applied to predict future unfolding of a set of specific events.

1 Introduction

In everyday life people use past events and their own knowledge to predict future events. In such everyday predictions people use widely available resources (newspapers, Internet). In this study we focused on sentences referring to the future as potentially most useful in predicting future unfolding of events. Especially people who bare significant social responsibility, such as managing strategy, planing, or developing policies in large companies are in need of such prediction support tools, since the company's results and profits depend on how accurate is their competence in future trend prediction.

For example, when a company have just found out that a decision depends on the event 'X' and 'Y' will happen or not respectively, they can prepare four tactical options A, B, C or D for their company management, depending on the predictions on what would happen in the future when they select:

1. Tactical option 'A' when both events 'X' and 'Y' happen.
2. Tactical option 'B' when the event 'X' happens and the event 'Y' does not happen.
3. Tactical option 'C' when the event 'Y' happens and the event 'X' does not happen.
4. Tactical option 'D' when both events 'X' and 'Y' do not happen.

When people select their actions from a range of possible options, usually they consider and combine plenty of information, including one's and also other's experiences and expertise regarding the events. Obtaining such information for future predictions is a very challenging task requiring much time and labor with a lot of information to process and foresight ability before making a decision.

Previous studies have shown that data mining using statical techniques can support such predictions about future outcomes. However, to achieve that, one needs to process numerous numeric data, which requires professional skills.

On the other hand, there have been studies on predicting future outcomes of events with the use of Natural Language Processing (NLP) techniques. Some of them have proposed applying causality information and past events [Radinsky et al. 2012], which assume that "when the event A happens, the event B will usually also happen". However, such methods usually are limited to general events from the range of widely perceived common sense (e.g., "what will happed when an apple falls on ones head"). Others applied methods based on keyword extraction with their frequencies on a timeline using past events, temporal expressions and event-related keywords [Kanazawa et al. 2011].

As for a different research, [Nakajima et al. 2016] have proposed the method for automatic extraction of future reference sentences using morphosemantic information and indicated that future reference sentences could be useful in supporting predictions about future events.

Moreover, future reference sentences can be useful as significant knowledge base since they include various related information, such as background information regarding the event in question, which is also used as the source of knowledge by experts. Therefore, such sentences can be useful for making future predictions.

In this research we propose a prototype method which applies future reference sentences in the process of future prediction support and discuss its effectiveness.

The outline of this paper is as follows. In section 2 we describe previous research related to the prediction of future events. Section 3 describes the proposed method applying automatic extraction of references to future events. Section 4 describes the experiments evaluating the method. Section 5 proposes a discussion of the examined approach and future implications. Finally, section 6 contains conclusions and sets out plans for future improvement and applications of the proposed method.

2 Previous Research

Linguistically expressed references to the future has been studied by a number of researchers.

[Baeza-Yates 2005] investigated five hundred thousand sentences containing future events extracted from one day of Google News (http://news.google.com/), and found out that scheduled events occur with high probability and with correlation between the occurrence of an event and its time proximity. Therefore the information about upcoming events is of a high importance for predicting future outcomes. [Kanhabua et al. 2011] investigated newspaper articles, and found out that one-third of all sentences contains reference to the future. [Kanazawa et al. 2010] extracted implications for future information from the Web using explicit information, such as time expressions. [Omar et al. 2011] indicated that time information included in a document is effective for enhancing information retrieval applications. [Kanazawa et al. 2011] extracted unreferenced future time expressions from a large collection of text, and proposed a method for estimating the validity of the prediction by searching for a real-world event corresponding to the one predicted automatically. [Jatowt et al. 2013] studied relations between future news in English, Polish and Japanese by using keywords queried on the Web.

When it comes to predicting the probability of an event to occur in the future, [Jatowt and Au Yeung 2011] have proposed a clustering algorithm for detecting future phenomena based on the information extracted from text corpus, and proposed a method of calculating the probability of an event to happen in the future. [Jatowt et al. 2009] used the rate of incidence of reconstructed news articles over time to forecast recurring events, and proposed a method for supporting human user analysis of occurring future phenomena. [Aramaki et al. 2011] used SVM-based classifier on Twitter to perform classification of information related to influenza and tried to predict the spread of the disease by using a truth validation method. [Kanazawa et al. 2011] proposed a method for estimation of validity of the prediction by automatically calculating cosine similarity between predicted relevant news and searching for the events that actually occurred. [Radinsky et al. 2012] proposed the Pundit system for prediction of future events in news based on causal reasoning derived from a similarity measure calculated using different ontologies.

The above findings have lead us to the idea that by using expressions referring to the future included in trend reports (newspaper articles, etc.), we could be able to support the future prediction process as one of the activities of people perform everyday. Such a method would be applicable in corporate management, trend foresight, and preventive measures, etc. Also, as indicated by previous research, when applied in real time analysis of Social Networking Services (SNS), such as Twitter or Facebook, it could also become helpful in disaster prevention or handling of disease outbreaks. This way the method would be useful in chance discovery [Ohsawa and McBurney 2003], by e.g., providing hints for a company regarding planning its future investments.

Methods using time referring information, such as "year", "hour", or "tomorrow", has been applied in extracting future information and retrieving relevant documents. It has also been indicated that it is useful to predict future outcomes by using information occurring in present documents. In our research we focused on more sophisticated expressions, namely, morphosemantic sentence patterns.

3 Automatic Extraction of Future Reference Sentences

In this section, we describe our method for extracting future reference sentences from news corpora.

Future reference sentences include both explicit as well as implicit expressions referring to the future. Explicit expressions include e.g., future temporal expressions, or words and phrases referring to the future (e.g. will∼, be expected that∼, plan to∼, etc.).

However, many important sentences do not contain such explicit expressions, but the information regarding future outcomes can be encoded as an implicit information. See the example below regarding the future of America's army troops dispatch to Afghanistan.

"He rejoiced that President Obama had reemphasized the need to focus on the War on Terror in Afghanistan, increasing the likelihood of an early withdrawal of U.S. troops from Iraq."

The sentence does not contain any future referring expressions. Moreover, the sentence is in past tense ("rejoiced", "had reemphasized"), and therefore it is not possible to specify that the sentence refers to the future by using standard methods. Yet, the sentence clearly presents potential future outcomes ("withdrawal of U.S. troops from Iraq") with the use of implicit information.

The method we propose to deal with both explicit as well as implicit information, such as above, consists of two stages. Firstly, the sentences are represented in a morphosemantic structure [Levin and Rappaport Hovav 1998] (combination of semantic role labeling and morphological information). Secondly, frequent combinations of such patterns are automatically extracted from training data and used in classification.

Morphosemantic patterns (MoPs) are useful for representing languages rich both morphologically and semantically, such as Japanese (language of datasets used in this research). We generated the morphosemantic model using semantic role labeling (SRL) supported with morphological information. SRL provides labels for words and phrases according to their role in the sentence, such as those represented in Table 1.

3.1 Morphosemantic Patterns

In the first stage, all sentences are represented in **morphosemantic structure** (MS) for further extraction of **morphosemantic patterns** (MoPs) in the second stage.

The idea of MS has been described widely in linguistics and structural linguistics. For example, [Levin and Rappa-

Table 1: An example of a sentence analyzed by ASA.

Example I: Romanized Japanese (RJ): *Ashita kare wa kanojo ni tegami o okuru darō.* **/ Glosses:** Tomorrow he TOP her DIR letter OBJ send will (TOP: topic particle, DIR: directional particle, OBJ: object particle.) **/ English translation (E):** He will [most probably] send her a letter tomorrow.

No.	Surface	Label
1	*ashita*	[Time-Point]
2	*kare ha*	[Agent]
3	*kanojo ni*	[Patient]
4	*tegami o*	[Object]
5	*okuru darou*	[State_change]-[Place_change]-[Change_of_place(physical)]

port Hovav 1998] distinguish them as one of the two basic types of morphological operations on words, which modify the Lexical Conceptual Structure (LCS), or the semantic representation of a word. As for practical application of the idea, [Kroeger 2007] applied MoPs to analyze an Indonesian suffix *–kan*. Later [Fellbaum et al. 2009] applied MoPs to improve links between the synsets in WordNet. More recently [Raffaelli 2013] used MoPs to analyze a lexicon in Croatian, a language rich both morphologically and semantically. In this research we used datasets in Japanese, and applied MoPs for the same reason. Using only one representation narrows the spectrum of analyzed information. Moreover, till now there has been no practical application of MoPs to solving real-world problems. With our research we present the first attempt of this kind.

We generated the morphosemantic model using semantic role labeling with additional morphological information. Below we describe in detail the process of morphosemantic representation of sentences.

At first, the sentences from the datasets are analyzed using semantic role labeling (SRL). SRL provides labels for words and phrases according to their role in sentence context. For example, in a sentence "John killed Mary" the labels for words are as follows: John=`actor`, kill[past]=`action`, Mary=`patient`. Thus the semantic representation of the sentence is "`[actor][action][patient]`".

For semantic role labeling in Japanese we used **ASA**[1], a system, developed by [Takeuchi et al. 2010], which provides semantic roles for words and generalizes their semantic representation using an originally developed thesaurus. An example of SRL provided by ASA is represented in Table 1.

Moreover, not all words are semantically labeled by ASA. The omitted words include those not present in the thesaurus, as well as grammatical particles, or function words not having a direct influence on the semantic structure of the sentence, but in practice contributing to the overall meaning. For such cases we used a morphological analyzer MeCab[2] in combination with ASA to provide morphological information, such as "Proper Noun", or "Verb". However, in its basic form MeCab provides morphological information for all words separately, which causes compound words to be unnecessarily divided. For example "Japan health policy" is one morphosemantic concept, but in grammatical representation it takes form of "Noun Noun Noun". Therefore as a post-processing procedure we added a set of linguistic rules for specifying compound words in cases where only morphological information was provided.

Moreover, as it is shown on Table 1, some labels provided by ASA are too specific. Therefore in order to normalize and simplify the patterns, we specified the priority of label groups in the following way.

1. Semantic role (Agent, Patient, Object, etc.)
2. Semantic meaning (State_change, etc.)
3. Category (Dog → Living animal → Animated object)
4. In case ASA does not provide any of the above labels, perform compound word clustering for parts of speech (e.g., "International Joint Conference on Artificial Intelligence" → `Adjective Adjective Noun Preposition Adjective Noun` → `Proper_Noun`)

Furthermore, post-processing in the case of no semantic information is organized as follows.

- If a compound word can be specified, output the part-of-speech cluster (point 4 above).
- If it is not a compound word, output part-of-speech for each word.

Below is an example of a sentence generalized with the morphosemantic tagging method applied in this research.

Romanized Japanese: *Nihon unagi ga zetsumetsu kigushu ni shitei sare, kanzen yōshoku ni yoru unagi no ryōsan ni kitai ga takamatte iru.*

English: As Japanese eel has been specified as an endangered species, the expectations grow towards mass production of eel in full aquaculture.

SRL: `[Object][Agent][State_change][Action][Noun][State_change][Object][State_change]`

3.2 Future Reference Pattern Extraction

From sentences represented this way we extract frequent MoPs using SPEC [Ptaszynski et al. 2011]. Firstly, we generate ordered non-repeated combinations from all sentence elements. In every n-element sentence there is k-number of combination groups, such as that $1 \leq k \leq n$. All combinations for all values of k are generated. Additionally, all non-subsequent elements are separated with a wildcard ("*", asterisk). Pattern lists extracted this way from training set are then used in classification of test and validation set.

SPEC uses all patterns generated this way to extract frequent patterns appearing in a given corpus and calculates their weight. Two features are important in weight calculation. A pattern is the more representative for a corpus when, the longer it is (length k), and the more often it appears in the corpus (occurrence O). Thus the weight can be calculated by

- awarding length (`LA`),
- awarding length and occurrence (`LOA`),
- awarding none (normalized weight, `NW`).

The generated list of frequent patterns can be also further modified. When two collections of sentences of opposite features (such as "future-related vs. non-future-related") is compared, the list will contain patterns that appear uniquely in only one of the sides (e.g., uniquely positive patterns and

[1]http://cl.it.okayama-u.ac.jp/study/project/asa

[2]http://code.google.com/p/mecab/

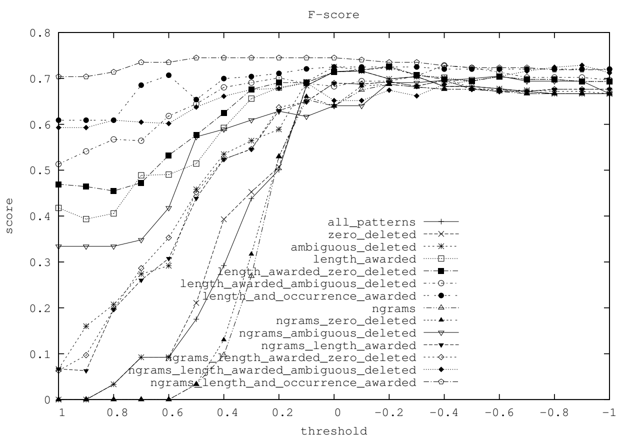

Figure 1: F-score for all tested classifier versions.

uniquely negative patterns) or in both (ambiguous patterns). Thus pattern list can be modified by

- using all patterns (ALL),
- erasing all ambiguous patterns (AMB),
- erasing only those ambiguous patterns which appear in the same number in both sides (zero patterns, 0P).

Moreover, a list of patterns will contain both the sophisticated patterns (with disjoint elements) as well as more common n-grams. Therefore the system can be trained on a model using

- patterns (PAT), or
- only n-grams (NGR).

All combinations of those modification are tested in the experiment.

3.3 Future Reference Sentence Extraction with Morphosemantic Patterns

From three newspaper corpora[3] We collected and annotated a dataset containing equal number of (1) sentences referring to future events and (2) other (describing past, or present events). We conducted an evaluation experiment with training dataset containing 130 sentences each, furthermore as the test data we used randomly extracted additional 170 sentences from the news corpora.

The test datasets were applied in a text classification task on 10-fold cross validation. Each classified test sentence was given a score calculated as a sum of weights of patterns extracted from training data and found in the input sentence. The results were calculated with Precision, Recall and balanced F-measure. We compared fourteen classifier versions shown F-measure in Figure 1. The results indicated that the highest overall performance was obtained by the version using pattern list containing all patterns (including ambiguous patterns and n-grams). We looked at top scores within the threshold, checked which version got the highest break-even point (BEP) of Precision and Recall, and calculated statistical significance of the results.

Finally, we compared the proposed method to [Jatowt et al. 2013], who extracted future reference sentences with 10 words unambiguously referring to the future, such as "will" or "is likely to", etc. In comparison, the proposed method obtained better results even when only 10 most frequent MoPs were used (Table 2).

Moreover, We verified the performance of the fully optimized model (FOM). We retrained the best model using all sentences from the initial dataset and verified the performance by classifying the new validation set. The final overall performance is represented in Figure 2. Finally, the obtained break-even point (BEP) was 0.76.

Table 2: Comparison of results for validation set between different pattern groups and the state-of-the-art.

Pattern set	Precision	Recall	F-score
10 patterns	0.39	**0.49**	**0.43**
10 pattern with only over 3 elements	**0.42**	0.37	0.40
5 patterns	0.35	0.35	0.35
Optimized (see Fig. 2)	**0.76**	**0.76**	**0.76**
[Jatowt et al. 2013] (10 phrases)	**0.50**	0.05	0.10

4 Future Prediction Support Experiment

In this section we present a validation experiment for the effectiveness of using future reference sentences (FRS) in the task of supporting predictions regarding future events.

4.1 Experiment Setup

In the experiment for supporting future trend prediction we used the fully optimized model of **future reference sentences** (FRS) trained on **morphosemantic patterns** (MoPs) described in Section. 3.3. The model was applied to extract new FRS concerning a specific topic, from the available newspaper data. Such sentences are further called **future prediction support sentences** (FPSS). Future prediction was performed by a group of thirty laypeople (balanced gender distribution, age groups from university students to their fifties), who were told to read the FPSS and reply to questions asking them to predict the future in 1–2 years from now, or from the starting point of prediction.

The questions were taken from the **Future Prediction Competence Test** (**FPCT**, jap.: *Senken-ryoku Kentei*), released by the **Language Responsibility Assurance Association** (**LRAA**, jap.: *Genron Sekinin Hoshō Kyōkai*)[4], a non-profit organization focused on supporting people of increased public responsibility (managers, politicians) and people responsible of making decisions influencing civic life. Such

[3] *Nihon Keizai Shimbun*, *Asahi Shimbun*, *Hokkaido Shimbun*.

[4] http://homepage3.nifty.com/genseki/kentei.html

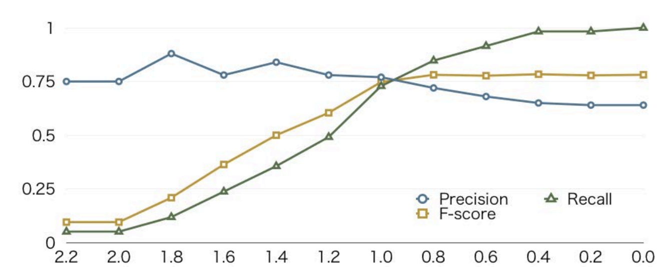

Figure 2: Final overall results of fully optimized model.

people often need to perform public speeches in which they reveal details or opinions regarding future events. In such situations they are obliged to express some contents (e.g., objective facts), while restraining from revealing others (one's fears towards the future or negative thoughts, disturbing public opinion, etc.). Thus the association helps preparing one's public speeches and responsibility bound presentations.

The Future Prediction Competence Test is an examination that measures prediction abilities in humans regarding specific events that are to happen in 1–2 years in the future. It has been initiated in 2006 and from that time it has been performed six times. The test consists of various questions, including multiple choice questions (e.g., "Will US Army contingent in Afghanistan increase or decrease during next year?"), essay questions (e.g., "Describe economic situation of a country after next two years"), and questions that must be answered using numbers (e.g., "What will be the exchange rate of Japanese Yen to US Dollar after two years"), and they are scored after those particular events have come to light.

The questions for the experiment to benchmark our future trend prediction support method were selected from the 4th of the past six future prediction tests[5], as it had the largest total number of questions, and respondents, which would assure the highest possible objectivity of the evaluation. Implemented in 2009, the 4th Future Prediction Competence Test contained questions regarding predictions for 2010 and 2011, and the scoring was performed in 2011. Respondents were to choose to answer at least 15 questions from a total of 25 questions in six areas, namely, politics, economics, international events, science and technology, society, and leisure. The test contained a large number of multiple choice questions and several questions requiring predicting specific numbers. There was also a small number of questions requiring a written explanation of the reasoning for the prediction. When participating in the Test, respondents can browse any and all materials, and are free to seek the opinions of others in answering the question, but the submission deadline was fixed and set at December 31st, 2009 (end of the year). The scoring is set at 90 total points on prediction questions and 30 total points for descriptive questions, with a total of 120 points.

The prediction support method we developed in this study is intended to apply future prediction support sentences (FPSS) related to a given question and provide assistance for humans on which answer to choose. Therefore for its evaluation we limited the questions to multiple-choice questions. Questions with two or more (multiple) choices were selected from the 4th Future Prediction Competence Test and applied as questions for the experiment. One example of such question is represented in Figure 3.

```
Question 3: Predict the stationing status
of US forces in Afghanistan at the end of
June 2011.

(A) The US forces will be still present and
    further reinforced comparing to October
    2009.
(B) The US forces will be still present on
    similar level comparing to October 2009.
(C) The US forces will be still present but
    in decreased number comparing to October
    2009.
(D) The US forces will be completely
    withdrawn.

Answer:
[ 1st candidate:       / 2nd candidate:
/ 3rd candidate    ]

Specify which sentence (number ID) from
the prepared Future Prediction Support
Sentences was most useful in making the
above decision:

1st candidate:  [    ]
2nd candidate:  [    ]
3rd candidate:  [    ]
```

Figure 3: An example of one multiple choice question from the 4th Future Prediction Competence Test with additional question inquiring which of the prepared automatically extracted sentences was most useful.

4.2 Data Preparation

In this section we describe the way of data preparation for the experiment. Firstly, a total of 7 multiple-choice questions were selected from the 4th FPCT test. Laypeople read the FPSSs presented to them and were given some time to respond. However, differently to the original settings of the test we did not give the participants one or two years for the answer, but required them to answer on site.

The FPSSs for each question presented to the laypeople participants were gathered by in the following way. At first we extracted from *Mainichi* Newspaper's entire 2009 year all sentences related to the questions on the basis of topic keywords (for Q3 from Table 3 it would be for example "US Army", or "Afghanistan"). Those sentences were then analyzed by the proposed method using the fully optimized model trained on MoPs, and sorted in a descending order.

This way the sentences that appeared on top of the list were in the highest probability to be future reference sentences. We retained only those FPSS with scores over 0.0 and presented the highest 30 of them to the subjects in chronological order. We decided to present the subjects FPSS in the order they appeared in newspapers instead of descending order of scores so that the subjects could have a better image of how the events unfolded, which would make the prediction more natural. We also decided to limit the number of sentences for the subjects to read to thirty so that the subjects did not become bored or tired too quickly. However, we stored the rest of the sentences in case the subjects insisted on further reading. Moreover, there were also situations in which the list of initial sentences extracted with topic keywords was less than thirty. In such situations we presented to the subjects all sentences which had a probability of being future reference sentences.

In addition, for questions for which less than 30 FPSSs were extracted in general, We presented all of the sentences that were classified into FRS.

[5] http://homepage3.nifty.com/genseki/senken/index4.html

As an example, some of the FPSSs for Question 3 are presented below. The questions were answered directly after reading only the FPSSs. Additionally, the respondents were asked to report the ID number of the FPSSs they referred to in their answer (or the FPSS that was the most informative and useful in their opinion).

The questions were collected with the following keywords. We tried several kinds of keywords on each question, and after careful examination of the sentences, decided to use the noun keywords that appeared in the original questions. Where it was possible we added closely related keywords by hand to cover the widest scope of possible search topics that a normal user would perform if they wanted to search for newspaper articles related to a specific topic. In particular we used the below topic keywords for each question.

Q1-1: (Participation in regional government by foreigners with permanent resident status) | (Participation in regional government permanent resident alien)

Q1-2: Husband and wife retaining separate family names

Q2: midterm elections | (Republican | Democrat) & (United States | America)

Q3: Afghanistan

Q4: Analog broadcasting | Digital broadcasting

Q5: Child allowance

Q6: (Democratic Party | Ruling Party | Liberal Democratic Party) & elections

In the evaluation of subjects choices we retained the original scoring schema. Namely, each of the questions 1, 2, and 7 were allocated 3 points. Moreover, in questions 2–5 the participants were allowed to make up to three candidate choice answers: primary candidate, secondary candidate and tertiary candidate, allocated 3 points, 2 points and 1 point, respectively if selected correctly. Additionally, to make the evaluation more strict and objective, for comparison, we also used a different scoring, based strictly on only one point per question.

Examples (translated in English) of FPSS for Question 3. (ordered chronologically in the form [Month/Day] in 2009) are represented below.

01/18 Other newspapers are also carrying out the Mainichi Newspaper's three-part feature reportage on trilateral coordination between Japan, Korea, and the US regarding North Korean nuclear arms, cooperation between Japan and Korea on reconstruction aid to Afghanistan, and the establishment of regular meetings or "shuttle diplomacy" between the respective leaders of these countries.

01/21 Additionally, it revealed their intention to finish the Iraq War through the gradual withdrawal of US combat troops stationed there, and put full force into the War on Terror in Afghanistan.

01/22 Substantial negotiations toward realizing the campaign pledge to reduce the number of stationed US forces "within 16 months of inauguration" have begun, aiming for an early formulation of a comprehensive plan that includes sending more U.S. troops to Afghanistan, a key battleground in the War on Terror.

01/22 Ahmad Saif (29), an engineer in Baghdad, rejoiced that President Obama had reemphasized the need to focus on the War on Terror in Afghanistan, increasing the likelihood of an early withdrawal of U.S. troops from Iraq.

02/07 At a cabinet-level meeting between Finance and Foreign Ministers of each country, in addition to the steps to be taken on the deterioration of public order in Afghanistan caused by formerly dominant Taliban forces, the agenda featured discussion on water resource development policies in response to the ongoing drought, and negotiations over assistance measures.

4.3 Experiment Results

The result of performance obtained by the subjects in future prediction task when supported only with the proposed method, with comparison to original results of Future Prediction Competence Test were represented in Figure 4.

At first the scoring was performed in accordance with future prediction test scoring procedure, wherein each question is worth up to 3 points with a total of 21 possible points.

For questions with up to 3 choices possible, we awarded 3 points when the subjects' first choice was correct, 2 points when the second was correct, and 1 point when the third choice was correct. Apart from the experiment with 30 respondents, we analyzed the original responses of the participants in the 4th Future Prediction Competence Test. The total possible score of was equal to 120 points. The test was taken by 11 people. From the total of 120 points, prediction questions accounted for 90 points, while essay questions accounted for 30 points. The comparison was based on prediction questions with a maximum score of 90 points.

In the performed experiment, the average score of our participants was 42.9% (see Figure 4). In comparison, the average score of the test participants was 33.4%. The results were similar, which indicates several things. Firstly, even though the events for prediction in our experiment were in fact from the past, the experiment participants performed similarly to original test participants. Therefore it can be said that the participants did not use (or did not have) the knowledge about the predicted events and that they based their judgments only on the provided FPSS. Furthermore, in comparison with original test results, an improvement of approximately 10% was noticed. This can be considered as the contribution of our method. However, the greatest contribution of our method for future prediction support is the following. Even if we assume that the improvement was not sufficient, and that our subjects performed similarly to original test participants, it must be remembered that our subjects made their decision based only on about thirty specifically extracted sentences and were given only short time for decision, whereas original test participants had over one year for preparing the answer, unlimited access to all available data and help from experts.

Additionally, the highest score in our experiment was 85.7%, while the lowest was 14.29%. In comparison with the participants of the 4th Future Prediction Competence Test these results indicate an improvement of about 8%-points for the lowest score range to even 25%-points for the highest range. The accuracy of the results (when any of the options 1–3 were correct) is shown in Figure 4.

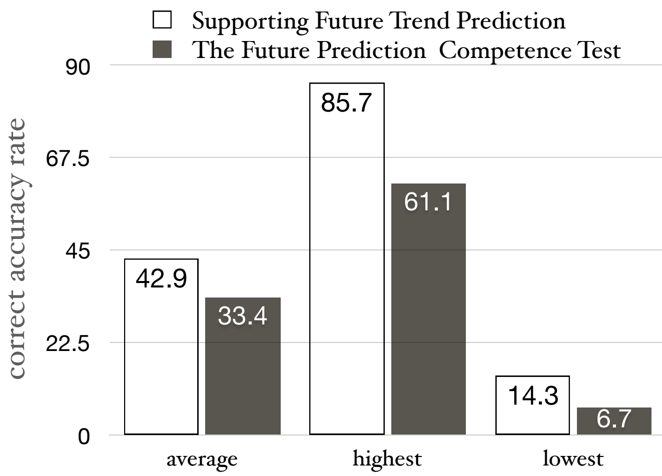

Figure 4: Performance of subjects in future prediction task supported with the proposed method with comparison to original results of Future Prediction Competence Test.

Moreover, the Future Prediction Competence Test has an established ranking system based on the number of points a participant received. On the 4th Future Prediction Competence Test, a final score covering over 60% of all points gives the participant a title of the 1st Class Future Prediction Competence; scores within 50-60% account for 2nd Class; 40-50% account for 3rd Class. This refers to the level of competence a participant is said to have when it comes to prediction of future unfolding of events. On the 4th Future Prediction Competence Test, 2 people earned 1st Class, none earned 2nd, and 2 people earned 3rd Class. In comparison, experiment participants performing predictions with the use of FPSS produced significantly more accurate results, if their scores were calculated at the time of test submission: 6 people would earn 1st Class, 6 people would earn 2nd, and 4 would earn 3rd Class of Future Prediction Competence.

Hence, supporting future prediction with future reference sentences extracted for specific topics can be considered as much more efficient than collecting available information by oneself for a year.

5 Discussion

In this section, we discuss the effectiveness of FRS for future trend prediction while comparing in detail experiment results with the Future Prediction Competence Test.

As shown in Figure 4, if we look at the accuracy of the 4th Future Prediction Competence Test, the average was 33.4%, demonstrating that when people have every means at their disposal, they still only accurately predict the future on around one third of the time. [Kurokawa and Kakeya 2009] analyzed trends in the answer results of the 1st Future Prediction Competence Test and verified whether the idea of such collective intelligence is useful or not in the context of future prediction. The accuracy rate at that time was 33.17 %. Moreover, Kakeya et al. concluded that the collective intelligence is not possible when it comes to future prediction. It means predicting future trends is not easy for people, even when they have access to all available resources.

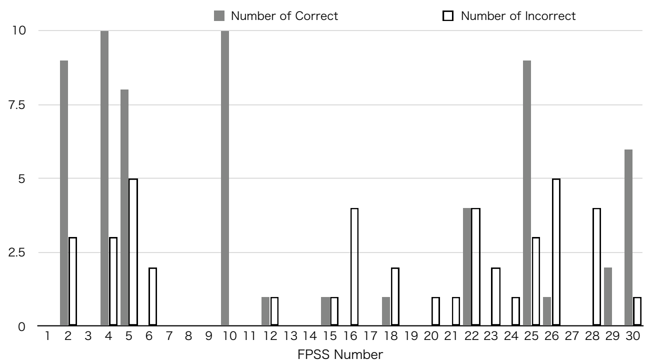

Figure 5: FPSS referred to by experiment participants as most useful with comparison between successful and unsuccessful participants.

On the other hand, the accuracy of the proposed method was 43%, an improvement of 13 percentage points over that average. Furthermore, a consideration of the certification breakdown from 1st Class to 3rd shows that only one third of all Future Prediction Competence Test got a certification, while if our experiment subjects using FPSS participated in the test, over half of them would get the certification (reach over 60% of all points). Thus, it is evident that when predicting future trends, FRS can dramatically reduce time and effort spent gathering information and achieve above-average predictive accuracy. Therefore we can say that using FRS to support future trend forecasting is both effective and efficient.

Next, we analyzed the FPSS referred to by experiment participants as most useful. As an example, Figure 5, shows for graphs for Question 3. Gray bars indicate the number of statements referred to by successful respondents, while white bars indicate the number of statements referred to by respondents who failed the task of prediction.

The contribution of these statements to choosing correct answers can be analyzed by focusing on gray bars. It is possible that differences in prediction accuracy depend on which of the 30 FPSS statements were referred to. Taking Question 3 as an example, We analyze both the content of statements that were only referred to by incorrect answers as well as those that contributed to correct responses.

The values on the horizontal axis of Figure 5 correspond to FRS numbers. In the experiment, 83.33% of responses to Question 3 were accurate. The examples of FPSS extracted for this question presented in previous section indicate that although all sentences contained the keyword “Afghanistan”, some sentences also contained references to US Army troops (see example 01/22), whereas others (02/07) contained the word “Afghanistan” but did not refer to the troops. Therefore, in order to improve the prediction accuracy, it is necessary to devise a better keyword setting for selecting FPSS from newspaper corpora.

6 Conclusions and Future Works

In this paper we conducted a validation experiment to determine whether future reference sentences are effective in supporting future trend prediction. We drew questions from the official Future Prediction Competence Test and, using topic keywords from those questions, gathered newspaper articles from the entire applicable year (2009). Then we extracted future prediction support sentences (FPSS) from those articles, and had thirty laypeople read these sentences and make predictions regarding unfolding of the events. The results yielded an average of 10 percentage point improvement over the results of the original Future Prediction Competence Test. However, the original test allowed respondents to prepare their answers for over a year and use any available information source, as well as seek the opinions of others. On the other hand, in the subjects of the experiment replied immediately after reading the provided support material, which consisted of only thirty FPSS. Therefore, although further experiments are needed, we can say that within the scope of present experiment, the significance of obtained results for prediction support has been sufficiently demonstrated.

In the experiment only separate future reference sentences were extracted for support from whole articles. However, the experiment showed that if such sentences are extracted with accurately set topic keywords they yield very detailed information sufficient to make the prediction. Furthermore, we believe that a combination of the proposed method with inference from statistical data would further increase the potential for obtaining information useful for future trend predictions.

In the future, we are planning to use this method with other corpora to conduct experiments on real-world problems, such as company management support or economic trend prediction. Carrying out a chronological analysis of FRS and the addition of sentiment analysis could lead to the discovery of new knowledge. We also plan to take part in the next Future Prediction Competence Test. The morphosemantic pattern extraction method alone could also be useful in tasks other than future reference sentence extraction, such as cyberbullying detection.

References

[Aramaki et al. 2011] Eiji Aramaki, Sachiko Maskawa, Mizuki Morita. 2011. Twitter catches the flu: Detecting influenza epidemics using twitter. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 1568–1576.

[Asur and Huberman 2010] Sitaram Asur, Bernardo Huberman. 2010. Predicting the future with social media. *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pp. 492–499.

[Baeza-Yates 2005] R. Baeza-Yates. 2005. Searching the Future. *SIGIR Workshop on MF/IR*.

[Fellbaum et al. 2009] Christiane Fellbaum, Anne Osherson, and Peter E. Clark. 2009. Putting semantics into WordNet's "morphosemantic" links. *Human Language Technology. Challenges of the Information Society*, Springer, pp. 350-358.

[Jatowt et al. 2009] Adam Jatowt, Kensuke Kanazawa, Satoshi Oyama, Katsumi Tanaka. 2009. Supporting analysis of future-related information in news archives and the Web. In *9th ACM/IEEE-CS Joint Conference on Digital Libraries*, pp. 115-124.

[Jatowt and Au Yeung 2011] Adam Jatowt, Ching-man Au Yeung. 2011. Extracting collective expectations about the future from large text collections. *The 20th ACM international conference on Information and knowledge management*, pp. 1259–1264.

[Jatowt et al. 2013] Adom Jatowt, Hideki Kawai, Kensuke Kanazawa, Katsumi Tanaka, Kazuo Kunieda, Keiji Yamada. 2013. Multilingual, Longitudinal Analysis of Future-related Information on the Web. *Proceedings of the 4th International conference on Culture and Computing 2013*, IEEE Press.

[Kanazawa et al. 2010] Kensuke Kanazawa, Adam Jatowt, Satoshi Oyama, Katsumi Tanaka. 2010. Exracting Explicit and Implicit future-related information from the Web(O) (in Japenese). *DEIM Forum 2010*, paper ID: A9-1.

[Kanazawa et al. 2011] Kensuke Kanazawa, Adam Jatowt, Katsumi Tanaka. 2011. Improving Retrieval of Future-Related Information in Text Collections. *2011 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pp. 278–283.

[Kanhabua et al. 2011] Nattiya Kanhabua, Roi Blanco, Michael Matthews. 2011. Ranking related news predictions. *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pp. 755-764.

[Kroeger 2007] Paul Kroeger. 2007. Morphosyntactic vs. morphosemantic functions of Indonesian *-kan*. In *Architectures, Rules, and Preferences: Variations on Themes of Joan Bresnan*, edited by A. Zaenen, J. Simpson, T. H. King, J. Grimshaw, J. Maling and C. Manning, pp. 229-251. Stanford, CA: CSLI Publ.

[Kurokawa and Kakeya 2009] T. Kurokawa and H. Kakeya. 2009. Analysis of a tendency to answer in foresight official approval (in japanese), *The fifth media information inspection arts and sciences meeting*, pp.50-55, 2009.

[Levin and Rappaport Hovav 1998] Beth Levin and Malka Rappaport Hovav. 1998. *Morphology and Lexical Semantics*. In Spencer and Zwicky, eds., pp. 248-271.

[Matthews et al. 2010] Michael Matthews, Pancho Tolchinsky, Roi Blanco, Jordi Atserias, Peter Mika, Hugo Zaragoza. 2010. Searching through time in the New York Times. *Proc. of the 4th Workshop on Human-Computer Interaction and Information Retrieval*, pp. 41–44.

[Nakajima et al. 2016] Yoko Nakajima, Michal Ptaszynski, Fumito Masui, Hirotoshi Honma. 2016. A Method for Extraction of Future Reference Sentences Based on Semantic Role Labeling, *IEICE Transactions on Information and Systems*, Vol.E99-D, No.2, pp.514-524.

[Ohsawa and McBurney 2003] Ohsawa Y., and McBurney P., eds. 2003. *Chance Discovery*, Springer Verlag.

[Omar et al. 2011] Omar Alonso, Jannik Strötgen, Ricardo Baeza-Yates, Michael Gertz. 2011. Temporal Information Retrieval: Challenges and Opportunities. *TWAW*, Vol. 11, pp. 1–8.

[Popescu and Strapparava] Octavian Popescu and Carlo Strapparava. 2013. Behind the Times: Detecting Epoch Changes using Large Corpora, *Proc. of IJCNLP 2013*.

[Ptaszynski et al. 2011] Michal Ptaszynski, Rafal Rzepka, Kenji Araki and Yoshio Momouchi. 2011. Language combinatorics: A sentence pattern extraction architecture based on combinatorial explosion. *International Journal of Computational Linguistics (IJCL)*, Vol. 2, Issue 1, pp. 24-36.

[Radinsky et al. 2012] Kira Radinsky, Sagie Davidovich and Shaul Markovitch. 2012. Learning causality for news events prediction. *The 21st International Conference on World Wide Web*.

[Raffaelli 2013] Ida Raffaelli. 2013. The model of morphosemantic patterns in the description of lexical architecture. *Lingue e linguaggio*, Vol. 12, No. 1 (2013), pp. 47-72.

[Takeuchi et al. 2010] Koichi Takeuchi, Suguru Tsuchiyama, Masato Moriya, Yuuki Moriyasu. 2010. Construction of Argument Structure Analyzer Toward Searching Same Situations and Actions. *IEICE Technical Report*, Vol. 109, No. 390, pp. 1-6.