Extraction of Daily Changing Words for Question Answering

Kugatsu Sadamitsu, Ryuichiro Higashinaka, Yoshihiro Matsuo

NTT Media Intelligence Laboratories, NTT Corporation
1-1 Hikari-no-oka, Yokosuka-shi, Kanagawa, 239-0847, Japan
{sadamitsu.kugatsu, higashinaka.ryuichiro, matsuo.yoshihiro}@lab.ntt.co.jp

Abstract

This paper proposes a method for extracting *Daily Changing Words* (DCWs), words that indicate which questions are real-time dependent. Our approach is based on two types of template matching using time and named entity slots from large size corpora and adding simple filtering methods from news corpora. Extracted DCWs are utilized for detecting and sorting real-time dependent questions. Experiments confirm that our DCW method achieves higher accuracy in detecting real-time dependent questions than existing word classes and a simple supervised machine learning approach.

Keywords Daily Changing Word, Temporal Slot, Question Answering

1. Introduction

In Question Answering (QA), it is important to recognize user intentions; that is, to determine what the user really wants to know. One of the problems is the occasional lack of time information in questions. For example, consider the question "*Tell me the guests of TV program XYZ?*" (テレビ番組 XYZ のゲストは 誰). It would be reasonable to recognize this question as asking about recent guests, not those in the past. As shown by this example, there are questions whose answers vary depending on the interpretation of time, and they need to be treated accordingly to provide reasonable answers. To this end, we focus on *detection for timely questions*, that is detecting whether a question is asking about a real-time dependent event or fact. By correctly detecting such questions we can correctly select appropriate resources for retrieving correct answers. For example, for non-timely questions, ordinary Web and Wikipedia are appropriate for retrieving answers. For timely questions, Twitter and other real-time-updated resources would be appropriate for retrieving answers by limiting the search span to a few hours or a few days.

Classification of answer types is the first process in QA systems. As answer types, Ferrucci et al. utilized lexical answer types (LATs) (Ferruci et al., 2012) and Higashinaka et al. utilized named entity types (Higashinaka et al., 2013). However, as far as we know, no study has focused on the classification of questions in terms of timeliness.

When questions contain *time-related words* such as "today" or "now", it is obvious that the question is about real-time events. When they contain *burst words* (words whose occurrence frequencies have shown a sudden increase in a short time span), it is also reasonable to identify them as timely questions. The problem is that some questions do not contain such words. Even when a question does not include time-related words or burst words, we can recognize that the question is timely, such as "Tell me the guest of TV program XYZ?" (テレビ番組 XYZ のゲストは誰？) by the clue, "guest" (ゲスト) because its referents vary day by day. On the other hand, the question "Tell me is a landmark of Tokyo? " (東京の有 名なランドマー クはどこ?) should not be identified as timely question because we know that landmarks do not vary their referents in short time spans. In this paper, we propose a new word class *Daily Changing Words* (DCWs) (such as *guest*); they indicate an association with timeliness. We also propose a method that can automatically extract DCWs from large corpora. Experiments show that the proposed method can extract DCWs and that they are effective for the detection of timely questions.

2. Daily Changing Words

We collected and analyzed Japanese questions that asked for timely information and discovered that words (DCW) such as *guest, starting pitcher, game*, and *temperature*, whose referents (entities being referred to) vary day by day, are used frequently. We named these words *Daily Changing Words* (DCWs). DCWs can be used as clues for detecting for timely questions, even when they do not include time related words or burst words. Finally, DCWs improves the accuracy of detecting timely question.

DCW is regarded as a kind of word class. In order to compare DCWs to existing word classes, we investigated several thesauri; of particular note, Japanese

Table 1: A list of templates for extracting Daily Changing Words. Upper line is original Japanese pattern and lower line is English translation. (TIME), (NE) and (DCW) indicate the slots of time, named entity, and Daily Changing Word, respectively.

Extraction template	Examples of extracted strings	DCW candidates
(TIME) の (DCW) は (NE)	(TIME=今日) の (DCW=ゲスト) は (NE=田中)	ゲスト
(TIME) 's (DCW) be (NE)	"(TIME=Today)'s (DCW=guest) is (NE=Tanaka) ."	guest
(TIME) (DCW) する (NE)	(TIME=昨日) (DCW=発表) された (NE=iPhone)	発表
(TIME), (NE) be (DCW)	"(TIME=Yesterday), (NE=iPhone) was (DCW=announced) ."	announce

Goi-Taikei (Ikehara et al., 1997) which has over 2700 word classes. We investigated the similarity and differences between DCWs and existing word classes. There is a verb class, the *attribute transformation* (属性変化) including "*lift*" (高揚する) and "*drop out*" (引退する). This word class partially overlaps with DCW, e.g. our DCW shares "*drop out*", although, *attribute transformation* class does not include noun type DCWs such as "*guest*". In the experimental section, we compare the words of *attribute transformation* to our DCWs in terms of discrimination of timely detection.

We also consider the related task of "*temporal slot filling*" (Ji et al., 2011). This task's goal is adding temporal information to selected slots (attributes). One piece of temporal information is the start and end dates for any slots holding named entities. Temporal slot filling task competitions were held for Knowledge-Base Population (KBP) at TAC 2011 and 2013 (Ji et al., 2011) [1]. In KBP temporal slot filling, only eight slots are temporal (PERSON: spouse, PERSON: title, PERSON: employee_of, PERSON: member_of, PERSON: cities_of_residence, PERSON: state_or_province_of_residence, PERSON: countries_of_residence and ORGANIZATION: top_employee_member).

These slot names (not value in the slot) are similar to DCWs. The largest difference between DCWs and temporal slot names is that DCWs indicate changes within short periods whereas temporal slot names indicate NE attributes for any period. Therefore, not all temporal slot names indicate timeliness and so are inadequate for detecting timely questions. For example, "marriage" (結婚) is a DCW because "marriage" is a short-term event and does not describe a state of long duration. The temporal slot name "spouse" (配偶者) is a kind of attribute about an NE (PERSON) and does change in short periods. The difference is shown more clearly in the example of DCW "birth" (誕生). The DCW "birth" must not become a temporal slot because "birth" is one-time event for an entity; in other words, it's not temporal for an entity.

From another perspective, we could attempt to use tense information to directly discriminate of timely questions. In English, if a question begins with "Who is", not "Tell me", the question can be regarded as timely because the verb is in present tense. However, unlike English, Japanese expresses tense by verbal conjugation. Because there are several different forms in Japanese verbal conjugations such as conjunctive form, it is often the case that we do not use verbal conjugations and simply use the base form, which results in the lack of tense information. Several early studies focused on the analysis of tense in Japanese, especially for second language learners (Shirai and Kurono, 1998)(Ishida, 2004). From the perspective of automatic tense estimation, Imamura et al. proposed an automatic estimation method that utilizes a large dictionary of light verbs (Imamura et al., 2011). These studies mainly dealt with ordinary sentences and focused on verbs and light verbs. In contrast, our approach focuses on not only verbs, but also nouns, and is specialized in that it is directed toward timely question detection.

3. A Method for Extracting Daily Changing Words

In this section, we describe a method for extracting DCWs in Japanese based on simple rules. These rules are created on the basis of some characteristics of DCWs. Consider the DCW "*guest*", and two question examples, "*Yesterday's guest was X*" and "*Today's guest is Y*". In these examples, two factors are different, time information ("*yesterday*" and "*today*" and named entity (*X* and *Y*) (in short, NE. Note that, in this paper we treat IREX's eight NE types). For capturing these cases, we construct patterns including time information and named entity slots. The two patterns are shown in Table 1. The patterns in the upper cell and those in the lower cell can extract general nouns and verbal nouns, respectively. We utilize a named entity recognizer (Suzuki et al., 2006) for determining NEs and time information (date and time). After many candidates of DCWs are extracted by matching these patterns to large corpora (35M blog articles), we removed the candidates that had only a

---

[1]http://www.nist.gov/tac/2013/KBP/

Table 2: Examples of extracted Daily Changing Words.

Extracted Daily Changing Words (general noun) (samples from 44 correct extractions)
guest (ゲスト), starting pitcher (先発), game (試合), leading role (主役), topics (話題) sunset (日没), costume (衣装), MC (司会)
Extracted Daily Changing Words (verbal noun) (samples from 34 correct extraction)
launching (発売), announcement (発表), appearance (出演), release (公開), marriage (結婚), accomplishment (完成), visiting Japan(来日)
Extracted NOT Daily Changing Words (both are erroneous extraction)
obtaining (入手), establishment (創業)

small number of NE word types and time information by thresholding because such candidates do not have sufficient variations needed for the characteristics of DCWs.

Furthermore, we also utilized a newspaper corpus for verifying that DCW candidates were truly useful indicators of changing events. That is, the contents of news articles intrinsically change every day. If a DCW candidate appears in sufficient number of news articles, we can regard it as reliable. Our implementation filters (drops) DCW candidates that appear only a few times (under a fixed threshold) in newspaper articles. The experiment (describes in the next section) used the threshold of 10 as per a preliminary examination. This simple filtering approach increases the reliability of the remaining DCWs.

4.   Experiments and Results

This section describes two experiments. The first assessed the quality of DCWs extracted by the proposed method. The second evaluated the accuracy of detecting timely questions. Each decision by the system was evaluated as correct or incorrect by us. The experimental settings are as follows: we used 35M blog articles crawled in Feb. 2012 for candidate extraction and 30,465 news articles crawled in June-July 2012 for filtering. As the test corpus, we created a total of 86 question sentences from those generated by multiple annotators (each question is independent, not sequential). The test corpus included 55 timely questions, none of which included time-related words, and 31 that were not timely sentences.

4.1.   Daily Changing Word Extraction

We adapted our extraction method to corpora holding blog entries and news articles. Our algorithm could

extract 82 DCWs automatically. The accuracy was 97.6% (80 / 82). Examples of extracted DCWs are shown in Table 2. These results confirm our method's performance in terms of timeliness detection and false detection. The two wrong extractions were "*establishment*" (創業) and "*obtaining*" (入手). To suppress such wrong extraction, we could constrain the [TIME] slot to words indicating recent time spans like "today" and "yesterday". On the other hand, this strategy would decrease the total number of extractions, and so was not adopted.

More specific results are described below. Using only the extraction rules (step 1) extracted 154 candidates. Of these, 91 were correct and 63 were incorrect as DCWs. The accuracy of filtering by newspaper corpus (step 2) is 81.9%. Of the 72 candidates, 59 were incorrect and 13 were correct for DCWs including "*MVP*" (MVP), "*TV*" (テレビ), "*news*" (ニュース) and "*pass away suddenly*" (急逝).

4.2.   Detecting Timeliness Questions

We compared four methods for the detection of timeliness questions. They were

- *Japanese Goi-Taikei word class matching in each question (baseline: described in Section 2.)* ,

- *Daily Changing Word matching in each question (proposed),*

- *Supervised approach using a logistic regression classifier (baseline),*

- *The result of supervised approach with Daily Changing Words matching (proposed).*

As the DCWs, TV program entries in Wikipedia that were broadcast during the experiments were added to DCWs because these entries change their content every day or every week. To construct our Goi-Taikei word class, we extracted 645 words belonging to the *attribute transformation* (属性変化) verb category of Goi-Taikei based on the discussion in Section 2.. Because each entry is registered as a verb with "*do*" (する), stemming was performed. In the supervised approach experiments, we utilized the standard bag of words features and performed 5-fold cross validation using LIBLINEAR (Fan et al., 2008).

Table 3 shows the accuracies and the confusion matrix for each method. The keyword matching methods (first and second columns) yielded high precision, although DCWs (second column) achieved much higher recall than Japanese Goi-Taikei (first column). This result indicates that Japanese Goi-Taikei word classes are not sufficient to detect timely questions as

Table 3: Results for the detection of timeliness questions. "sys.=T" and "sys.=N" indicate the system estimation is timeliness question and non-timeliness question respectively. "ref.=T" and "ref.=N" indicate the annotated reference is timeliness question and not timeliness question. The difference in accuracy between the proposed method (fourth) and the baseline method(third) is significant by binominal test with $p < 0.01$.

	Keyword matching method				Machine learning method			
	Goi-Taikei (baseline)		DCW (proposed)		Supervised (baseline)		Supervised + DCW (proposed)	
	sys.=T	sys.=N	sys.=T	sys.=N	sys.=T	sys.=N	sys.=T	sys.=N
ref.=T	2	56	23	35	40	18	49	9
ref.=N	0	31	1	30	14	17	14	17
total	Prec.=1.000, Rec.=0.035, F val.=0.068		Prec.=0.958, Rec.=0.397, F val.=0.561		Prec.=0.740, Rec.=0.690, F val.=0.714		Prec.=0.778, Rec.=0.845, F val.=0.810	

Table 4: Examples of correct and incorrect detection of timeliness questions using a supervised method with Daily Changing Word (DCW) in Table 3. The asterisk (*) indicates improved example by adding Daily Changing Words to the result of the supervised method from only the supervised method. The words underlined are DCWs.

	sys.=T	sys.=N
ref.=T	ソニーが 発表 した新製品は？ –*Tell me the new product underlined announced by SONY?* NTT の 終値 は？* –*Tell me the closing stock price of NTT? * 次の iPad はいつ 発売 ？* –*Tell me the release date of next iPad? * 読売新聞朝刊、一面記事は何？ –*Tell me the article on the front-page of Yomiuri's morning paper?*	誕生花は何？ –*Tell me the birth flower (of today)?* アントラーズ戦の審判は誰？ –*Tell me the umpier of the game of Antlers?* 巨人のマジックは？ –*Tell me the magic number of Giants?*
ref.=N	一番長い氷河はどれ –*Tell me the longest glacier?* ドラえもんの誕生日はいつ？ –*Tell me the birthday of Doraemon?*	秋の花といえば –*Tell me a flower of fall?* 島根県出身の首相は何人 –*Tell me the number of prime ministers elected from Shimane?*

we noted in Section 2. On the other hand, DCWs increased the recall score in detecting timely questions. The third and fourth methods were compared in a supervised situation without/with DCWs. Combining the supervised approach and our approach is likely to improve recall performance. This expectation was confirmed because the recall improved about 15% (0.690 to 0.845) by adding just 80 DCWs in Table 2. Some questions detected by using Supervised + DCW in Table 3 are shown in Table 4. Correct examples (intersection of "sys.=T" and "ref.=T" in Table 4), shown with an asterisk (*), indicate improved examples by adding DCWs from only the supervised method. We can confirm that DCWs (underlined) are effective for improving and reducing false negative examples (18 to 9 examples in Table 3). Remaining false negative examples (intersection of "sys.=N" and "ref.=T") include DCWs such as " *umpier*" (審判), "*magic number*" (マジック) and "*birth flower*" (誕生花) which

were not extracted by automatic extraction because these words are rarely used in the corpora. All 14 false positive examples (intersection of "sys.=T" and "ref.=N") are derived by the supervised method as shown in Table 3.
Remaining goals are to raise the recall of DCW extraction for reducing false negatives and to better integrate the supervised methods and DCWs.

5. Conclusion

We proposed a method for extracting Daily Changing Words (DCWs) for detecting timely questions. 82 DCWs are automatically extracted from blog articles by two templates (general nouns and verbal nouns) and a simple filtering approach. Experiments confirmed that DCWs are more effective for detecting timely questions than existing word classes taken from Japanese Goi-Taikei and a simple supervised approach. Our future work includes expanding the list

of DCWs and evaluating them in end-to-end QA systems.

6. References

Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874.

David Ferruci, Anthony Levas, Sugato Bagchi, David Gondek, and Erik Mueller. 2012. Watson: Beyond Jeopardy! *Artificial Intelligence*.

Ryuichiro Higashinaka, Kugatsu Sadamitsu, Kuniko Saito, and Nozomi Kobayashi. 2013. Question answering technology for pinpointing answers to a wide range of questions. *NTT Technical Review*, 11(7).

Satoru Ikehara, Masahiro Miyazaki, Satoshi Shirai, Akio Yokoo, Hiromi Nakaiwa, Kentaro Ogura, Yoshifumi Ooyama, and Yoshihiko Hayashi. 1997. Goi-Taikei: A Japanese Lexicon.

Kenji Imamura, Tomoko Izumi, Genichiro Kikui, and Satoshi Sato. 2011. Semantic label tagging to functional expressions. In *Proceedings of the 17th Annual Meeting of the Association for Natural Language Processing*, pages 308–311.

Midori Ishida. 2004. Effects of recasts on the acquisition of the aspectual form -te i-(ru) by learners of japanese as a foreign language. *Language Learning*, 54(2):311–394.

Heng Ji, Ralph Grishman, and Hoa Trang Dang. 2011. Overview of the TAC2011 knowledge base population track. In *In Fourth Text Analysis Conference (TAC 2011)*.

Yasuhiro Shirai and Atsuko Kurono. 1998. The acquisition of tense-aspect marking in japanese as a second language. *Language learning*, 48(2):279–244.

Jun Suzuki, Erik McDermott, and Hideki Isozaki. 2006. Training Conditional Random Fields with Multivariate Evaluation Measures. In *Proceedings of the COLING-ACL Conference*, pages 217–224.