Sentence Extraction in Recognition Textual Entailment Task

Yudi Wibisono, Dwi H. Widyantoro, Nur Ulfa Maulidevi
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
yudi@upi.edu, {dwi,ulfa}@stei.itb.ac.id

Recognizing textual entailment (RTE) is a task that predict whether a text fragment can be inferred from another text fragment. In this paper, we tackle RTE problem using sentence extraction to cover semantic variation and then extracting subject, predicate and object from each sentence without using external resources like Wordnet. Finally, similarity function is used to predict entailment relation. In sentence extraction phase, we used sentence detection, extract sentence in subordinate clause, prepositional phrase and passive sentence. Our system has accuracy of 0.63 which is comparable to other system that is not using external resources.

textual entailment, sentence extraction, text processing

I. INTRODUCTION

Recognizing Textual Entailment (RTE) has become important natural language processing task in recent year. The RTE goal is to detect entailment relation between two snippet text pairs <T (text), H (hypothesis)>. T entails H, if H can be inferred from T using common knowledge.

The following is an example extracted from first RTE challenge dataset showing Text (T) entails Hypothesis (H).

T: The body of Satomi Mitarai was found by a teacher after her attacker returned to class in bloody clothes.

H: Mitarai's body was found by a teacher after her killer returned to their classroom covered in blood.

RTE has been useful in various natural language processing application to handle variation of semantic expression, such as information extraction (IE), text summarization, question answering (QA), and machine translation (MT). In text summarization, textual entailment (TE) can be used to remove sentence redundancy. Lloret [7] showed that TE gave 6.78% performance improvement in a summarization task.

In the QA application, text entailment can be employed to select answer candidates. Harabagiu [5] ranked answer candidates by their entailment probability with the question. This method gave accuracy improvement of 20%. Magnini [8] used entailment graph to analyzing customer interactions across multiple channel like email chat and social media.

RTE also can be employed in automatic information credibility assessment task. Information is assumed more credible if there are other independent sources confirmed it and the information is consistent with ground truth information. RTE then can be used to compare information from various sources to check whether it confirm with each other and consistent with ground truth. Other important factor in information credibility is to asses the sources of information or informant credibility. We can asses sources credibility by compare their shared information in the past with the ground truth [9] using RTE.

To help user determine credible information from the internet, Murakami [12] used textual entailment to created statement map (**Fig. 1**). After user submitted a query, statement map shows summary of opinion related to the query and the contradicting opinions with their supporting evidence.

There are three groups of information in statement map: Focus, Conflict and Evidence. Focus is group of information that related to the query. Conflict contains information which contradicts with Focus. Both Focus and Conflict groups have Evidence group which contains information to support each group.

Some approaches have been employed to recognize textual entailment automatically, such as: 1) lexical similarity and syntactic alignment; 2) logic-based and 3) combination techniques. Some system use external lexical database such as Wordnet, DIRT, Wikipedia and verb oriented external databases like VerbNet, Framenet [6] [11]. Although external resources could increase accuracy, it needs more processing power and time.

Lexical similarity computes similarity between T and H by calculating lexical overlaps, like words overlap between T and H [2]. This methods assume that degree of lexical similarity indicate the entailment. Although this is simplest method using shallow feature, the accuracy is low. Other use syntactic alignment and relation between T-H is predicted by alignment of syntactical structure such as syntactic dependency graph [10].

Syntactic alignment approach for RTE uses syntactic structure of text. Wang [13] employed structure similarity function which used dependency structure. Dependency structure consists of triplet relations < relation, term1, term2 >. For example, dependency structure for a sentence "30 die in a

978-1-4799-7996-7/14/$31.00 ©2014 IEEE

bus collision in Uganda." is <nsubj,die,30>, <root,ROOT,die>, <det,collision,a>, <nn,collision,bus>, <prep_in,die,collision>, <prep_in, collision, Uganda>. Structure similarity then calculated by counting triplets of H contained in T.

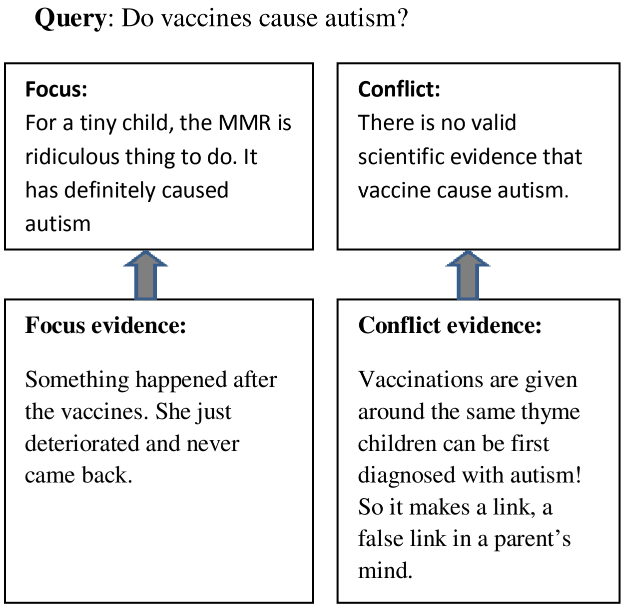

Fig. 1 Statement Map

In logic-based semantic approach, the task is to find logical semantic relation between T and H. T and H are transformed into some logic representation, and entailment between H and T is proved by inference [11][1].

In this paper, we describe a technique which enumerate sentence variations of T and H to capture as much as possible semantic variation between T and H using sentence extraction. Then we extract subject, predicate and object for each sentence for better alignment between T and H. Finally, we use similarity function to predict entailment.

II. APPROACH

Our textual entailment recognizer has five modules, as shown in **Fig. 2**. First, preprocessing module applies Stanford Parser [3] and transforms T and H into syntactical structure in the form of parse tree. Second, sentence extraction applied to H and T that may have more than one sentence or clause. Third, subject, predicate and object extracted from sentence. Fourth, features are extracted from sentence, subject, predicate and object and then classifier used to determine whether H entails T.

In the next section we will discuss each module in more detail.

A. Preprocessing

In the preprocessing step, we generated part of speech tag and syntactic parse tree from T and H using Stanford Parser [3]. We also removed symbol like "(", ")". '-".

B. Sentence Extraction

Sentence extraction is needed because hypothesis (H) can entail with only some part of the text (T), for example:

T: "At the same time the Italian digital rights group, Electronic Frontiers Italy, has asked the **nation's government to investigate Sony** over its use of anti-piracy software."

H: "Italy's government investigates Sony."

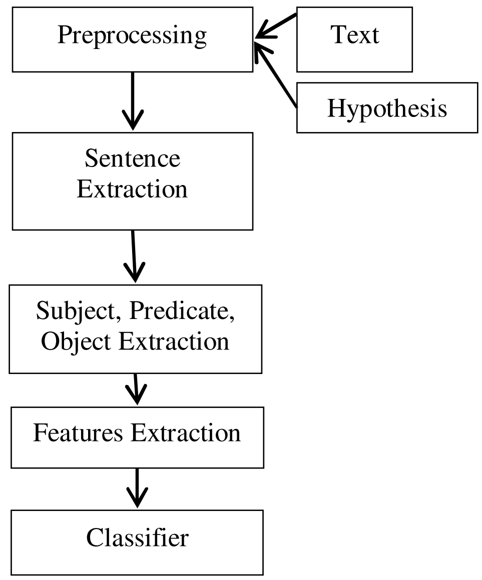

Fig. 2 : System architecture

Next, Part-of-speech tag and parse tree are used in the sentence extraction module. Steps in the sentence extraction for T and H are:

1. Sentence detection, split regular sentence which is separated by a period.
2. Extract subordinate clauses. "SBAR" part of speech tag used to extract subordinate clause.
3. Extract sentence in prepositional phrase. Detected using "S" or "VP" tag which is exist in "PP"
4. Generate active sentence from passive sentence.

For example, using above steps, we extracted the following text into six sentences (**Error! Reference source not found.**).

"*Ebola hemorrhagic fever is a fatal disease caused by a new virus which has no known cure. When a new epidemic was detected in Zaire in the spring of 1995, it was widely perceived as a threat to the West. Public attention was intense.*"

TABLE I. EXAMPLE OF EXTRACTED SENTENCES

Sentence	Step
Ebola hemorrhagic fever is a fatal disease caused by a new virus which has no known cure.	Sentence detection
When a new epidemic was detected in Zaire in the spring of 1995, it was widely perceived as a threat to the West	Sentence detection
Public attention was intense	Sentence detection
a new epidemic was detected in Zaire in the spring of 1995	Subordinate clause
a new virus which has no known cure	Sentence in prepositional phrase
[*] detected in Zaire in the spring of 1995	Passive to active sentence

After sentence extraction steps, number of sentences in T or H will be increased.

C. Part of sentence extraction: subject, predicate, and object

We extracted subject, predicate and object for every sentence generated in sentence extraction module. To extract those part of sentence, we used part of speech tag and syntactic parse tree information and apply heuristic rules.

Error! Reference source not found. shows examples for subject, predicate and object extracted from sentences.

TABLE II. EXAMPLE OF SUBJECT, PREDICATE OBJECT EXTRACTION

Sentence	Subject	Predicate	Object
A unique feature of previous Ebola outbreaks has been the relative sparing of children.	A unique feature of previous Ebola outbreaks	has been	sparing of children
The number of the confirmed Ebola cases has risen slightly to 26 in Gabon and to 16 in Congo Brazzaville	The number of the confirmed Ebola cases	has risen	26 in Gabon and to 16 in Congo Brazzaville

To extract subject, we choose the highest level noun phrase in the parse tree. Predicate is extracted from the first verb phrase after the subject. Finally, object is extracted from the highest level noun phrase after the predicate.

D. Feature Extraction and Classifier

We use Term Frequency-Inverse Document Frequency (TF-IDF) for words weighting. Words with high TF-IDF have strong relationship with sentence they are appeared in.

After sentence extraction, each T and H could be transformed into several sentences. Cosine similarity is used for calculating distance between parts of sentence (subject, predicate, object) from all H's sentences to all T's sentences.

We assumed value of cosine similarity between T and H is proportional with entailment relationship between them. We also use cosine similarity to calculate distance between T and H sentences directly and not using part of sentence.

Table 1 shows complete features.

TABLE III. TABLE 1 FEATURE LIST

Feature	Description
best_t_h_spo	Best cosine similarity between all T's and H's sentences using part of sentence (subject-predicate-object). Subject is compare with pair's subject and so on.
best_t_h	Best cosine similarity without part of sentence
avg_t_h_spo	Average cosine similarity using part of sentence
t_h	Cosine similarity between H and T directly (not using sentence extraction)

After all features extracted from dataset, we use machine learning classifier to automatically classify whether T entails H. We tested several classifier algorithm using WEKA [14] and Naïve Bayes gave the best accuracy (0.63).

III. EXPERIMENT AND RESULT

A. Data

We used dataset from Third Pascal Recognizing Textual Entailment Challenge (RTE-3) [4]. RTE-3 has two datasets: development set and test set. Each consists of 800 pairs text (T) and hypothesis (H), all manually annotated. RTE-3 dataset is the last RTE challenge dataset that available for direct download without needing special permission.

RTE-3 pairs taken from various source from web and was reviewed by three human judges. Average agreement between judges is 87.8% with Kappa level 0.75. Table 2 shows some RTE3 pairs.

Main measurement for RTE-3 challenge is accuracy.

B. Result and Discussion

Using all the features in

Table 1, system accuracy is 0.62, but if we use only two features: best_t_h and t_h, which not use the subject-predicate-object extraction, the accuracy is slightly higher at 0.63. This shows that sentence extraction is useful but there is problem with the subject-predicate-object extraction.

Result analysis show that sentence extraction is useful. For example the T-H pair:

T: They cite scholarly research showing that young women at co-ed schools tend to defer to men in intellectual debate.

H: Young women at co-ed colleges tend to defer to men.

T was extracted into two sentences:

T1: They cite scholarly research showing that young women at co-ed schools tend to defer to men in intellectual debate.

T2: young women at co-ed schools tend to defer to men in intellectual debate

TABLE IV. TABLE 2 EXAMPLE OF RTE-3 DATASET

Text (T)	Hypothesis (H)	Entails?
Mars is considered the red planet because of its bright red color.	Mars is called 'the red planet'.	YES
The policy focused on coca eradication from the territories of Peru and Bolivia. During the years of 1996 and 2000, much of the resources of "Andean Initiative" where utilized in intense aerial fumigation in coca producing areas.	Peru debates border region with Ecuador.	NO
Nival was founded in 1996 by Sergey Orlovskiy. In early 2005, the company was bought by Ener1 Group, a Florida-based holdings company, for around US$10 million.	Nival was sold in 2005.	YES
NPF said that young people displaying onset symptoms of Parkinson's disease have special needs, because they often have children to raise, jobs to keep and other issues to address that are different than those patients in retirement years.	The symptoms of Parkinson's disease are: tremor, rigidity (stiffness; increased muscle tone), bradykinesia/akinesia, also "dysdiadokinesia", and postural instability.	NO

and H was extracted into one sentence:

H1: Young women at co-ed colleges tend to defer to men.

Clearly the similar pair is T2 and H1, with same subject and predicate, and similar object. Similarity score between T2 and H1 is 0.62, but average similarity subject-predicate-object is only 0.32. This is because number of words is much smaller in subject, predicate and object compare with whole sentence. We tried vector normalization but the performance did not improved. In the future, we will try other similarity measures for subject-predicate-object alignment.

Compared to other methods which are not used external resources and use same dataset, our method is more simpler than [15], which use state of the art NLP, and has better accuracy than [16] [17].

IV. CONCLUSION

In this paper, we present our method that employs sentence extraction and part of speech extraction (subject, predicate, object) in recognition textual entailment task. For sentence extraction, we use sentence detection extract subordinate clause, sentence in prepositional phrase, and transform passive voice into active.

Our system accuracy is comparable with other system that did not use external resources, with only one system have accuracy better then ours.

REFERENCES

[1] Bar-Haim, Roy, et al. "Semantic inference at the lexical-syntactic level." *PROCEEDINGS OF THE NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE*. Vol. 22. No. 1. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 2007.

[2] Bos, Johan, and Katja Markert. "Combining shallow and deep NLP methods for recognizing textual entailment." *Proceedings of the First PASCAL Challenges Workshop on Recognising Textual Entailment, Southampton, UK*. 2005.

[3] Dan Klein and Christopher D. Manning. 2003. Accurate Unlexicalized Parsing. *Proceedings of the 41st Meeting of the Association for Computational Linguistics*, pp. 423-430.

[4] Giampiccolo, Danilo, et al. "The third pascal recognizing textual entailment challenge." *Proceedings of the ACL-PASCAL workshop on textual entailment and paraphrasing*. Association for Computational Linguistics, 2007.

[5] Harabagiu, Sanda, and Andrew Hickl. "Methods for using textual entailment in open-domain question answering." Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2006.

[6] Iftene, Adrian, and Alexandra Balahur-Dobrescu. "Hypothesis transformation and semantic variability rules used in recognizing textual entailment." *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*. Association for Computational Linguistics, 2007.

[7] Lloret, Elena, et al. "A Text Summarization Approach under the Influence of Textual Entailment." *NLPCS*. 2008.

[8] Magnini, Bernardo, et al. "Entailment Graphs for Text Analytics in the Excitement Project." *Text, Speech and Dialogue*. Springer International Publishing, 2014.

[9] Ministry of Defence, Joint Doctrine Publication 2-00, Understanding and Intelligence Support to Join Operations, 2011

[10] Snow, Rion, Lucy Vanderwende, and Arul Menezes. "Effectively using syntax for recognizing false entailment." Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics. Association for Computational Linguistics, 2006.

[11] Tatu, Marta, and Dan Moldovan. "Cogex at RTE3." *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*. Association for Computational Linguistics, 2007.

[12] Murakami, Koji, et al. "Statement map: assisting information crediblity analysis by visualizing arguments." *Proceedings of the 3rd workshop on Information credibility on the web*. ACM, 2009.

[13] Wang, Rui, and Günter Neumann. "Recognizing textual entailment using sentence similarity based on dependency tree skeletons." *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*. Association for Computational Linguistics, 2007.

[14] Witten, I. H. and Frank, E. (2000) Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann

[15] Li, Baoli, et al. "Machine learning based semantic inference: Experiments and Observations at RTE-3." *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*. Association for Computational Linguistics, 2007.

[16] Malakasiotis, Prodromos, and Ion Androutsopoulos. "Learning textual entailment using SVMs and string similarity measures." *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*. Association for Computational Linguistics, 2007.

[17] Marsi, Erwin, Emiel Krahmer, and Wauter Bosma. "Dependency-based paraphrasing for recognizing textual entailment." *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*. Association for Computational Linguistics, 2007.