Information Extraction and Sentence Classification applied to Clinical Trial MEDLINE Abstracts

Kazuo Hara and Yuji Matsumoto

Graduate School of Information Science, Nara Institute of Science and Technology

8916-5 Takayama, Ikoma, Nara, Japan

Email : {kazuo-h, matsu}@is.naist.jp

ABSTRACT: In this paper, firstly we report experimental results on applying information extraction (IE) methodology to the task of summarizing clinical trial design information in focus on "Compared Treatment", "Endpoint" and "Patient Population" from clinical trial MEDLINE abstracts. From these results, we have come to see this problem as one that can be decomposed into a sentence classification subtask and an IE subtask. By classifying sentences from clinical trial abstracts and only performing IE on sentences that are most likely to contain relevant information, we hypothesize that the accuracy of information extracted from the abstracts can be increased. As preparation for testing this theory in the next stage, we conducted an experiment applying state-of-the-art sentence classification techniques to the clinical trial abstracts and evaluated its potential in the original task of the summarization of clinical trial design information.

1 INTRODUCION

Recently, people who practice medical treatment have been paying more attention to Evidence-Based Medicine (EBM). EBM is, "the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients" [1]. In practicing EBM, they are encouraged to be well informed on up-to-date sources of medical knowledge such as MEDLINE, the US National Library of Medicine's bibliographic database covering the fields of medical, pharmaceutical and biological sciences.

Among MEDLINE abstracts, those about clinical trials play one of the most important roles in EBM, for the results of clinical trials can provide firm evidence to support applying a certain therapy in actual medical treatments. However, since the rate at which new articles are being introduced into the MEDLINE database is fairly high, it takes patients or doctors who seek beneficial knowledge quite some time to read all of the articles that may contain clues in finding a suitable therapy. So, in order to assist members of the medical community, our goal is to summarize design information from MEDLINE clinical trial abstracts, in an effort to reduce the amount of time required to find relevant medical information.

In the research field of natural language processing, the task of information extraction (IE) has been pursued with a great deal of interest for decades. For example, in the series of Message Understanding Conferences (MUCs), participants developed methods for extracting information of the scenario-templates presented by the conference organizers. The focus of the study there was the construction of domain-specific lexicons and extraction patterns based on human judgment. Following the MUCs, the attention of researchers has shifted to automatic knowledge acquisition including lexicons and patterns.

In this paper, we report results of the two preliminary experiments for the summarization of clinical trial design information from MEDLINE abstracts. Firstly, we used conventional IE methods to conduct an experiment in extraction of clinical trial design information. Next, we performed sentence classification, using state-of-the-art sentence classification algorithm [2] with the future goal of using those results to determine when to carry out IE.

This paper is organized as follows. We describe the IE targets from clinical trial MEDLINE abstracts in section 2. We give results of the IE experiment in section 3 and of the sentence classification experiment in section 4, and finally conclude with some discussion about related work in section 5 and perspective on future work in section 6.

2 IE targets from clinical trial MEDLINE abstracts

In general, clinical trials should be carried out on predetermined conditions set for variables such as "Compared Treatment", "Endpoint", "Patient Population" and so on, to minimize bias and maximize precision of the results of the analysis [3]. Although other design information that describe clinical trials, which includes dosage or schedule of medicine, number of enrolled patients, and statistical significance of the results, are sometimes very important, we can at least say that the above three should be the extraction targets because few clinical trials are practiced without considering them. So in this paper, we consider "Compared Treatment", "Endpoint", and "Patient Population" as IE targets for summarizing the clinical trial MEDLINE abstracts. We give short comments about them as follows and show examples in Table 1.

PMID	Compared Treatment	Endpoint	Patient Population
15282352	(i) peginterferon alfa-2a plus ribavirin, (ii) interferon alfa-2a plus ribavirin	(i) sustained virologic response	(i) patients with chronic HCV infection and HIV
15235875	(i) iron reduction therapy	(i) serum alanine aminotransferase levels	(i) patients with CHC
15131791	(i) epoetin alfa, (ii) placebo	(i) RBV dose, (ii) quality of life, (iii) hemoglobin	(i) HCV-infect ed patients

Table 1: Examples of IE targets appeared in clinical trial MEDLINE abstracts; PMID = PubMed ID
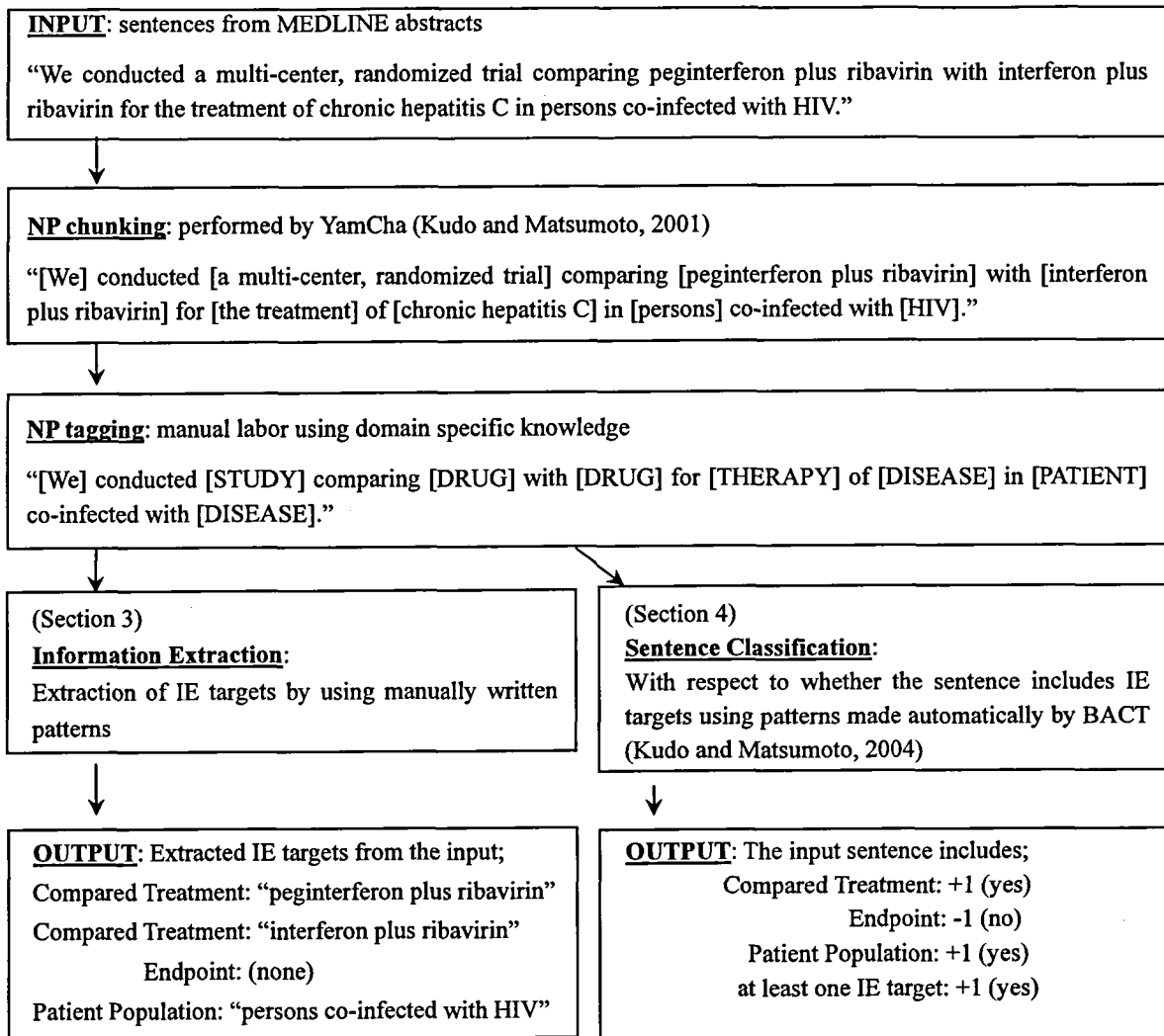
INPUT: sentences from MEDLINE abstracts

"We conducted a multi-center, randomized trial comparing peginterferon plus ribavirin with interferon plus ribavirin for the treatment of chronic hepatitis C in persons co-infected with HIV."

↓

NP chunking: performed by YamCha (Kudo and Matsumoto, 2001)

"[We] conducted [a multi-center, randomized trial] comparing [peginterferon plus ribavirin] with [interferon plus ribavirin] for [the treatment] of [chronic hepatitis C] in [persons] co-infected with [HIV]."

↓

NP tagging: manual labor using domain specific knowledge

"[We] conducted [STUDY] comparing [DRUG] with [DRUG] for [THERAPY] of [DISEASE] in [PATIENT] co-infected with [DISEASE]."

↓ ↘

(Section 3)
Information Extraction:
Extraction of IE targets by using manually written patterns

(Section 4)
Sentence Classification:
With respect to whether the sentence includes IE targets using patterns made automatically by BACT (Kudo and Matsumoto, 2004)

↓ ↓

OUTPUT: Extracted IE targets from the input;
Compared Treatment: "peginterferon plus ribavirin"
Compared Treatment: "interferon plus ribavirin"
Endpoint: (none)
Patient Population: "persons co-infected with HIV"

OUTPUT: The input sentence includes;
Compared Treatment: +1 (yes)
Endpoint: -1 (no)
Patient Population: +1 (yes)
at least one IE target: +1 (yes)

Figure 1: The process of this paper's experiment

Compared Treatment	Procedures to extract IE targets
in titles	Extract NPs tagged as DRUG or THERAPY
in main texts	Extract NPs tagged as DRUG or THERAPY, which is included in the following patterns for listing or comparing drugs or therapy: "* DRUG (.* as non-NP) DRUG *" "* (compare* | between) * (DRUG | THERAPY) * (versus | with | and) * (DRUG | THERAPY) *"

Endpoint	Procedures to extract IE targets
in titles	Extract NPs tagged as TARGET
in main texts	Extract NPs tagged as TARGET, which is included in the following patterns for describing what performed in the clinical trial: "* (We | we) .* TARGET *" "* ((This | this) * as STUDY) .* TARGET *"

Patient Population	Procedures to extract IE targets
in titles	Extract NPs tagged as PATIENT Extract following noun phrases, which restricts patients with some clinical conditions: "PATIENT * with DISEASE" "PATIENT * with TARGET"
in main texts	Extract following noun phrases, which restricts patients with some clinical conditions: "PATIENT * with DISEASE" "PATIENT * with TARGET"

Table 2: Manually written procedures to extract IE targets

In this paper, "Compared Treatment" corresponds to the drug or therapy examined in the clinical trials. The final goal of clinical trials is to find out whether the therapy under investigation has clinical benefits. For example, in most confirmatory trials, the candidate product for a new drug is compared with control treatments or active comparators. As for "Endpoint", confirmatory clinical trials investigate the efficacy and safety of a new drug or therapy. Variables directly related to the patient's outcome, such as decrease in mortality or improvement in quality of life, are often employed. In contrast, indirect variables that include pharmacokinetics parameters are investigated in many exploratory trials. About "Patient Population", the inclusion criteria for patients are always set up in clinical trials. For example, only healthy male adults are enrolled in phase one clinical trials in many cases.

3 Information Extraction Experiment

3.1 Method

After tokenization, word sequences in the MEDLINE abstracts are part-of-speech (POS) tagged using the TnT tagger [4]. Then, noun phrase (NP) chunking is performed by the YamCha chunker [5]. Next, NPs are manually tagged according to our tag set described in Table 3. This can also be accomplished automatically by using machine learning, however, we chose to carry out tagging manually for this experiment due to the low accuracy of these automatic methods. Finally in order to extract IE targets (of the form of NPs), we follow the manually constructed procedures shown in Table 2 that includes regular expression patterns. We describe the procedures separately for titles and main texts because most of the titles in abstracts are condensed by important information. The process described above is illustrated in Figure 1.

Tag	Covered concept	Example
DISEASE	disease, complication, symptom, virus	"chronic hepatitis C"
DRUG	drug, chemical compound, nucleoside, placebo	"interferon plus ribavirin"
STUDY	clinical trial, statistical analysis	"randomized controlled trial"
THERAPY	therapy, treatment, regimen, surgery	"antiviral treatment"
PATIENT	participants in the clinical trial	"HbeAg-positive patients"
TARGET	endpoints, clinical laboratory evaluation	"Hb concentration"
SCHEDULE	time schedule of the clinical trial	"an additional 24 weeks"
VALUE	value or score of TARGET	"significantly higher rates"
NUMBER	numeral expression	"20 percent"

Table 3: Tag set for NPs in the domain of clinical trial

3.2 Data

We downloaded the 50 most recent abstracts of clinical trials from the MEDLINE database (http://www.ncbi.nlm.

nih.gov/entrez/query.fcgi) on October 2004. The important part of reference query was "hepatitis [MeSH Terms] AND Randomized Controlled Trial [ptyp]". To simplify the experiment, abstracts were selected from the medical area of hepatitis.

3.3 Evaluation Metrics

Authors of abstracts mention entities concerning clinical trial design information by the form of NPs, and we extract the entities by using the procedures that pick out the pattern matched NPs. In this IE experiment, we are interested in the precision and recall for extracting correct IE target entities, not those for extracting correct NPs. Also we are interested in precision and recall for the summarization of clinical trial design information for each abstract. We will give the definition of recall and precision for those below.

More formally, we have 50 abstracts, that is, A_i; $i = 1, \cdots,$ 50. In each A_i, there are entities, E_{ij}; $j = 1, \cdots, e(i)$, such that $e(i)$ is the number of entities mentioned in the abstract A_i. Also in each A_i, there are NPs, NP_{ik}; $k = 1, \cdots, n(i)$, such that $n(i)$ is the number of NPs in the abstract A_i. Take notice that the authors of abstract A_i mention an entity E_{ij} more than once in the abstract by the form of an NP. Actually, the authors often mention the same entity E_{ij} in the description of background, method, result, or conclusion repeatedly.

Here, we define three boolean functions to describe relations between A_i, E_{ij} and NP_{ik};

If NP_{ik} is a mention of E_{ij},

$Mention(E_{ij}, NP_{ik}) = True$, otherwise $False$.

If E_{ij} is a correct IE target entitiy in A_i,

$Target(E_{ij}, A_i) = True$, otherwise $False$.

And, if NP_{ik} was extracted by the system from A_i,

$Extracted(NP_{ik}, A_i) = True$, otherwise $False$.

Then, we define precision and recall for IE target entities (Ent_pre and Ent_rec) as follows:

$$Ent\_pre = \frac{\sum_{i=1}^{50}\sum_{j=1}^{e(i)} I_1(E_{ij}, A_i)}{\sum_{i=1}^{50}\sum_{j=1}^{e(i)} I_2(E_{ij}, A_i)},$$

$$Ent\_rec = \frac{\sum_{i=1}^{50}\sum_{j=1}^{e(i)} I_1(E_{ij}, A_i)}{\sum_{i=1}^{50}\sum_{j=1}^{e(i)} I_3(E_{ij}, A_i)}.$$

$I_1(E_{ij}, A_i) = 1$,

   if $\exists k \ Mention(E_{ij}, NP_{ik}) \wedge Target(E_{ij}, A_i) \wedge Extracted(NP_{ik}, A_i)$.

$I_2(E_{ij}, A_i) = 1$,

   if $\exists k \ Mention(E_{ij}, NP_{ik}) \wedge Extracted(NP_{ik}, A_i)$.

$I_3(E_{ij}, A_i) = 1$,

   if $Target(E_{ij}, A_i)$.

$I_i(.)$; $i = 1, \cdots, 5$ are indicator functions. Those shown above and below are non-zero conditions.

We also define precision and recall for the summarization of each abstract (*Sum_pre* and *Sum_rec*) as follows (here, notice that #abstract = 50) :

$$Sum\_pre = \frac{\sum_{i=1}^{50} I_4(A_i)}{\#abstract},$$

$$Sum\_rec = \frac{\sum_{i=1}^{50} I_5(A_i)}{\#abstract}.$$

$I_4(A_i) = 1,$

    if $\forall k \exists j$ $Mention(E_{ij}, NP_{ik}) \wedge Target(E_{ij}, A_i) \wedge$

        $Extracted(NP_{ik}, A_i).$

$I_5(A_i) = 1,$

    if $\forall j \exists k$ $Mention(E_{ij}, NP_{ik}) \wedge Target(E_{ij}, A_i) \wedge$

        $Extracted(NP_{ik}, A_i).$

Ent_pre represents the percentage of correctly extracted IE target entities in comparison to the total number of entities extracted by the IE system, and *Ent_rec* means the percentage of correctly extracted IE target entities by the system compared with the total number of target entities. As for the evaluation metrics for the summarization, *Sum_pre* represents the proportion of abstracts without any incorrect extractions, and *Sum_rec* represents the proportion of abstracts where all IE targets were successfully extracted.

3.4 Results and Discussion

The results of IE from titles alone and from titles and main texts are summarized in Table 4. The results from titles alone can be considered as the baseline, because just putting together the titles is close to summarizing the articles. Judging from the results of experiment, although the results of IE from all text are superior to those from titles alone, the procedure we manually developed for this experiment is not powerful enough for practical use. Another remark is that the lower recall on the results from titles alone implies that few authors seem to describe "Endpoint" in the title.

		Compared Treatment	Endpoint	Patient Population
IE from titles only	Ent_pre	76.9%	66.7%	88.2%
	Ent_rec	60.2%	29.0%	53.6%
	Sum_pre	86.0%	96.0%	94.0%
	Sum_rec	40.0%	24.0%	50.0%
IE from titles and main texts	Ent_pre	71.4%	71.9%	68.6%
	Ent_rec	78.3%	59.4%	85.7%
	Sum_pre	70.0%	82.0%	68.0%
	Sum_rec	66.0%	52.0%	84.0%

Table 4: The results of IE experiment

Here, we discuss the reason for the low performance in this experiment. Firstly, the patterns we manually developed might be too simple to be effective measures, especially for "Compared Treatment" and "Endpoint". Of course we know that clinical trial experts will be able to construct more sophisticated and practical patterns. However, those based on heuristics have no theoretical guarantee that they are correct in the long run. Secondly, the task in this IE experiment is essentially different from the task of named entity (NE) recognition. For example, even finding an NP tagged by "DRUG" does not always lead to a correct answer of the target "Compared Treatment", because it is a common case that the author just introduces general drugs as background information of the article. In the next section, we show experimental results of sentence classification that might overcome the difficulties found in this IE experiment.

4 Sentence Classification Experiment

In this section, we show the experimental results of classifying each sentence of clinical trial MEDLINE abstracts into "yes" or "no" with respect to whether the sentence includes IE targets or not. By classifying sentences and only performing IE on sentences that are most likely to contain relevant information, we hypothesize that the accuracy of information extracted from the abstracts can be increased. Furthermore, in the respect that our final goal is summarization of clinical trial design information, classifying or filtering sentences could be the rational step to get to the goal.

Here, containing an NE that is often extracted as an IE target is just the necessary condition to classify the sentence "yes". It can be determined by the constituent words or phrases in the sentence, and sometimes by the neighboring sentences or discourse. Another important clue to classify a sentence correctly is the heading of the sentence, such as BACKGROUND, OBJECTIVE, METHOD, RESULT and CONCLUSION, which are explicitly written in some MEDLINE abstracts. However, in this experiment, we use only the information belonging to the sentence for classification. The reader should also take notice that some sentences contain more than one type of IE targets, as the example sentence shown in Figure 1.

4.1 Method

We use BACT, a state-of-the-art sentence classification algorithm [2], that uses machine learning to acquire optimal classification patterns and classify sentences according to them. We use NP-tagged sentences as the input for BACT.

BACT takes a sentence as input in the form of an ordered tree. The form of the trees varies according to how the connectivity of words in a sentence is defined. If the sentence is seen as a flat tree having a dummy root and all the words in the sentence being the leaves, a sub-tree can be thought of as a bag-of-words (BOW), although strictly speaking, this structure differs from general BOW in that the original word order of the sentence is preserved. Regardless, we will use the term bag-of-words to refer to this flat tree structure throughout this paper. If the sentence is seen as a sequence of words, a sub-tree represents an N-gram. If the sentence is parsed into a word dependency tree, a sub-tree represents a set of words that are directly connected by syntactic relations.

In this experiment, we compare performance between a BOW assumption, an N-gram assumption, and a

dependency grammar restriction. After constructing an ordered tree for each training sentence, BACT searches for sub-trees in each ordered tree and ranks the sub-trees that are found according to their applicability in sentence classification.

Both the BOW and N-gram assumptions are simple: BOW assumes interesting patterns exist within any combinations of words in a sentence with keeping just the order of them, and N-gram assumes interesting patterns exist within any sequences of neighboring words in a sentence, so they can be expressed using regular expressions. On the other hand, a dependency grammar restriction is available only if lexical knowledge can determine a dependency relationship between the words in a sentence correctly enough. In this experiment, we converted an NP-tagged sentence into a phrase-structure tree using the phrase-structure analyzer proposed in [6], and then converted the phrase-structure tree into a dependency tree using the head rules described in Collins' doctoral dissertation [7]. Although BACT can also handle a phrase-structure tree as input, we select the BOW, N-gram, and dependency tree for its ease in interpreting patterns.

4.2 Data

We use the same 50 clinical trial MEDLINE abstracts as shown in section 3.2. A total of 562 sentences were examined to determine whether they include the IE targets "Compared Treatment", "Endpoint", and "Patient Population". In addition, we classified sentences "yes" or "no" according to whether it has "at least one IE target". We can expect that a sentence having "at least one IE target" contains clinical trial design information, although other information such as dosage or schedule of medicine are not included.

4.3 Results and Discussion

In order to evaluate the effectiveness of automatically constructed patterns by BACT, we performed 5-fold cross validation test. The results are summarized in Table 5; $precision = (tp)/(tp+fp)$ and $recall = (tp)/(tp+tn)$, in that tp means true positive, fp means false positive, and tn means true negative. We also examined the significance of sentence classification disagreements between assumptions by McNemar's paired test. We observe a significance of (P < 0.05) between dependency and BOW for "at least one IE target" and "Compared Treatment", and between N-gram and BOW for "Patient Population".

(a) at least one IE target (b) Compared Treatment (c) Endpoint (d) Patient Population		(a)	(b)	(c)	(d)
# total sentence		562	562	562	562
# total "yes" sentence		160	90	76	55
BOW	precision	83.7%	82.3%	81.5%	71.7%
	recall	73.2%	70.8%	69.1%	64.7%
N-gram	precision	81.5%	82.6%	85.7%	81.5%
	recall	76.3%	71.7%	73.2%	81.5%
dependency	precision	84.2%	86.8%	84.7%	75.2%
	recall	82.4%	78.5%	72.2%	71.4%

Table 5: The results of Sentence Classification

Assuming ideal circumstances in which the parser always works correctly, we might expect dependency trees to outperform N-grams and BOWs because dependency trees make use of more lexical knowledge. For example, N-grams can only deal with multiple neighboring word expressions and BOWs cannot even handle fixed expressions. However, in this experiment, dependency trees are outperformed by N-grams regarding "Patient Population". We can guess the reason here: the fixed expression of "PATIENT with DISEASE" is frequently observed in the "yes" sentences, and parse errors occurred in many of the dependency trees caused by PP attachment ambiguity ("with DISEASE").

Next, we discuss the patterns automatically constructed by BACT on the assumption of dependency grammar restriction. In Table 6, we show patterns containing "DRUG" and their weight for classification (a hyphen indicates that the pattern is not selected by BACT). According to the table, "DRUG" is unfavorable for "Patient Population", favorable for "Endpoint", and context dependent for "Compared Treatment" and "at least one IE target". It is straightforward that the pattern, both "PATIENT" and "DRUG" depend on "received", that includes the expression "PATIENT received DRUG", has a high weight (+ 0.048) for "Compared Treatment" because the drugs received by patients are usually identical to the drugs examined in clinical trials. The pattern, "DRUG" depends on "of" and "of" depends on "TARGET", that includes the expression "TARGET of DRUG", has a high weight (+ 0.035) for "Endpoint" is also consistent in that "TARGET of DRUG" normally stands for objects that drugs take effect on. Table 7 shows how BACT classifies a sentence using constructed patterns and its weight. Because we use different training sentences, the weights don't correspond to each other in Table 6 and Table 7.

Patterns automatically constructed by BACT: (a) at least one IE target (b) Compared Treatment (c) Endpoint (d) Patient Population	(a)	(b)	(c)	(d)
both "PATIENT" and "DRUG" depend on "received"	–	+ 0.048	–	–
* DRUG *	+ 0.009	+ 0.046	–	–
"DRUG" depends on "of" and "of" depends on "TARGET"	+ 0.009	–	+ 0.035	–
"DRUG" depends on "DRUG"	+ 0.008	+ 0.013	–	–
"DRUG" depends on "received"	–	+ 0.010	+ 0.023	–
"DRUG" depends on "of"	+ 0.011	+ 0.006	+ 0.012	–
"DRUG" depends on "with"	–	- 0.004	–	- 0.026
"DRUG" depends on "to"	- 0.004	- 0.013	–	- 0.012
"DRUG" depends on "in"	- 0.011	- 0.019	–	–

Table 6: Automatically constructed patterns containing "DRUG"

Patterns automatically constructed by BACT: (a) at least one IE target (b) Compared Treatment (c) Endpoint (d) Patient Population	(a)	(b)	(c)	(d)
(default)	-0.033	-0.079	-0.141	-0.210
* We *	+0.019	+0.051	+0.016	+0.105
* STUDY *	+0.028	+0.013	+0.065	+0.081
* DRUG *	+0.014	+0.045	+0.009	-0.003
* with *	+0.000	+0.008	-0.002	+0.037
"DRUG" depends on "with"	-0.004	-0.003	-	-0.050
* for *	-	-	-	-0.006
* THERAPY *	+0.002	+0.014	-	-0.001
"THERAPY" depends on "for"	-0.002	-0.006	-	-
* of *	-	+0.002	-	-
* DISEASE *	-0.002	-0.000	-	+0.034
"DISEASE" depends on "of"	-	-0.009	-	-
* in *	+0.001	-0.013	+0.012	-0.000
* PATIENT *	-0.005	+0.007	-0.028	+0.070
"PATIENT" depends on "in"	-0.007	-	+0.000	-
"DISEASE" depends on "with"	+0.004	+0.006	+0.005	+0.018
Total Weight	+0.015	+0.035	-0.065	+0.074
Classification Result	+1 (yes)	+1 (yes)	-1 (no)	+1 (yes)

Table 7: Sentence classification in case of the sentence
shown in Figure 1

5 Relation to Previous Work

As described so far, we conducted experiments in IE and sentence classification experiment with the final goal of summarizing clinical trial design information. The idea that the IE benefits from text classification is not a new one [8, 9]. In bioinformatics, H. Yu and E. Agichtein have used a classifier based on SVM for extracting gene and protein synonyms from biological journal articles [10]. M. Craven and J. Kumlien have applied a naïve Bayes classifier to the task of protein subcellular-localization, and also have approached using a shallow phrase-structure parser, an area that is similar to our approach [11]. The approaches differ in that, firstly, we used dependency trees, and secondly, we employed BACT that learns comprehensively from all sub-trees of the training sentences.

6 Conclusions

In this paper, we have reported results of the two preliminary experiments to estimate the difficulty of our task. These preliminary experiments show that the combination of IE methodology and sentence classification can be the solution to the summarization task in clinical trial MEDLINE abstracts. So we plan to construct a complete pipeline from sentence classification to IE. However, some problems still remain. One of those is the construction of bigger corpora. In order to automate NP tagging, public sources of clinical trial information such as UMLS, MeSH or RxNorm, the thesauruses from US National Library of

Medicine, might prove useful. Also there may be a problem in identification of correspondence between entities and mentions in the IE subtask, and in the subtask of sentence classification using BACT, there are problems involving parsing accuracy such that come from coordination structure or PP attachment ambiguity. Recently the entity mention identification problem has been investigated in detail using machine learning techniques [12, 13]. In future research, we plan to investigate applicability of those researches to the summarization task of clinical trial MEDLINE abstracts.

REFERENCES

[1] D. L. Sackett, W. M. Rosenberg, J. M. Gray, R. B. Haynes, and W. S. Richardson. Evidence based medicine: what it is and what it isn't. BMJ 312 (7023), 13 January, pages 71--72,1996.

[2] T. Kudo and Y. Matsumoto. A Boosting Algorithm for Classification of Semi-Structured Text. In Proc. of Empirical Methods in Natural Language Processing, pages 301--308, 2004.

[3] ICH E9. Statistical Principles for Clinical Trials. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, 1998.

[4] T. Brants. TnT - a statistical part-of-speech tagger. In Proc. of the 6th Applied Natural Language Processing Conference, pages 224--231, 2000.

[5] T. Kudo and Y. Matsumoto. Chunking with Support Vector Machines. In Proc. of North American Chapter of the Association for Computational Linguistics, pages 192--199, 2001.

[6] E. Charniak. A Maximum-Entropy-Inspired Parser. In Proc. of North American Chapter of the Association for Computational Linguistics, pages 132--139, 2000.

[7] M. Collins. Head-Driven Statistical Models for Natural Language Parsing. PhD dissertation, University of Pennsylvania, 1999.

[8] D. Lewis. Data extraction as text categorization: An experiment with the MUC-3 corpus. In Proceedings MUC3, May 1991.

[9] E. Riloff and W. Lehnert. Information Extraction as a Basis for High-Precision Text Classification. ACM Transactions on Information Systems, 12 No.3: pages 296--333, 1994.

[10] H. Yu and E. Agichtein. Extracting synonymous gene and protein terms from biological literature. Bioinformatics 19, Suppl. 1: pages I340--I349, 2003.

[11] M. Craven and J. Kumlien. Constructing Biological Knowledge Bases by Extracting Information from Text Sources. In Proc. of the AAAI Conference on Intelligent Systems in Microbiology (ISMB-99), Heidelberg, Germany, AAAI, Menlo Park, CA, pages 77--86, 1999.

[12] A. McCallum and B. Wellner. Toward conditional models of identity uncertainty with application to proper noun coreference. In Proc. of IJCAI-03 Workshop on Information Integration on the Web (IIWeb-03), pages 79--84, 2003.

[13] X. Li, P. Morie, D. Roth. Robust Reading: Identification and Tracing of Ambiguous Names. In Proc. of North American Chapter of the Association for Computational Linguistics, pages 17--24, 2004.