Information Extraction and Sentence Classification Applied to Clinical Trial MEDLINE Abstracts

Computational Linguistics Laboratory

Kazuo Hara and Yuji Matsumoto

Background & Aim

Ubiquitous Medicine - a trend in the medical community -

- This trend is supported by popularization of ubiquitous technology such as
  - Remote Diagnostic Imaging, or
  - Electronic Health Records.
- The community is going to share comparable clinical information among medical sites.

This trend leads to a demand for high quality medical treatments.

- The concept, Evidence-Based Medicine (EBM), has become prevalent recently.
  - EBM requires medical practitioners to select appropriate treatments for individual patients based on the current best evidence.
- Where does the current best evidence come from?
  - One major source of evidence is clinical trial results.

What are the clinical trials?

- Phase I
  - Examination of the safety of the new treatment.
- Phase II
  - Exploration of the usage and dosage of the new treatment.
- Phase III
  - Verification of the new treatment compared to an active control or placebo.
- Phase IV
  - Post Marketing Surveillance of the new treatment.

Where to access the clinical trial results information?

- MEDLINE, the U.S. National Library of Medicine's (NLM) database of biomedical citations and abstracts that is searchable on the Web.
- MEDLINE search index includes:
  - clinical trial phases (phase I, II, III, and IV),
- but does not include important keys such as:
  - "compared treatments", "patient population", and "endpoints".

A clinical trial result is always summarized in a table.

- A typical example (phase III)

	Treatment A (New Drug)	**Treatment** B (Active Control)	statistical significance
Endpoint (Efficacy)	value or score	value or score	p-value
Endpoint (Safety)	frequency or count	frequency or count	p-value

MEDLINE abstracts are just the rewriting of the result tables.

【A MEDLINE abstract】

•TITLE: **Peginterferon Alfa-2a plus ribavirin** versus **interferon alfa-2a plus ribavirin** for chronic hepatitis C in HIV-coinfected persons.

•BACKGROUND: Chronic hepatitis C virus (HCV) infection is a cause of major … interferon plus ribavirin for the treatment of chronic hepatitis C in **persons coinfected with HIV**.

•METHODS: A total of 66 subjects were randomly assigned to receive … either a virologic response or histologic improvement.

•RESULTS: Treatment with peginterferon and ribavirin was associated with a significantly higher rate of **sustained virologic response** than was treatment with interferon and ribavirin. ‘‘‘

【Important Keys】

(1) Compared Treatment:
- **peginterferon alfa-2a plus ribavirin**
- **interferon alfa-2a plus ribavirin**

(2) Endpoint:
- **sustained virologic response**

(3) Patient Population:
- **persons coinfected with HIV**

Our research goal is:

- Extracting information with respect to important keys from each clinical trial MEDLINE abstract in order to construct a database which is easy to access.
- Information Extraction (IE) targets are:
  - “compared treatments”, “patient population”, “endpoints”, and so on.
- This can become a support for realizing EBM in the medical community.

Today’s presentation, we report ...

- Results of the two preliminary experiments for the summarization of clinical trial design information from MEDLINE abstracts.
  - Firstly, we used conventional **Information Extraction (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 with the future goal of using those results to determine when to carry out IE.

Experiment I

Information Extraction (IE)

What is Information Extraction (IE) in general?

- The goal is to extract pre-specified types of events, entities or relationships from the documents.
- Extracted information is usually entered into a database,
  - for the purpose of analyzing the data for trends, giving a natural language summary, or simply serving for on-line access.

Again, our IE task:

【A MEDLINE abstract】

•TITLE: **Peginterferon Alfa-2a plus ribavirin** versus **interferon alfa-2a plus ribavirin** for chronic hepatitis C in HIV-coinfected persons.

•BACKGROUND: Chronic hepatitis C virus (HCV) infection is a cause of major … interferon plus ribavirin for the treatment of chronic hepatitis C in **persons coinfected with HIV**.

•METHODS: A total of 66 subjects were randomly assigned to receive … either a virologic response or histologic improvement.

•RESULTS: Treatment with peginterferon and ribavirin was associated with a significantly higher rate of **sustained virologic response** than was treatment with interferon and ribavirin. ・・・

【IE targets】

(1) Compared Treatment:
・**peginterferon alfa-2a plus ribavirin**
・**interferon alfa-2a plus ribavirin**

(2) Endpoint:
・**sustained virologic response**

(3) Patient Population:
・**persons coinfected with HIV**

We used conventional IE methods to estimate the difficulty of our task.

- Part-of-speech tagging
  - TnT tagger (Brants, 2000)
- Noun Phrase chunking
  - YamCha (Kudo and Matsumoto, 2001)
- Noun Phrase tagging
  - Manual labor using domain specific knowledge
- Extraction of IE targets by using manually written patterns

Part-of-speech tagging:

- TnT tagger (Brants, 2000)

An example sentence:

"We conducted a multi-center, randomized trial comparing …"

TOKEN	Part-of-speech
We	PRP
conducted	VBD
a	DT
multi-center	NN
,	,
randomized	VBN
trial	NN
comparing	VBG

Noun Phrase chunking:

- YamCha (Kudo and Matsumoto, 2001) -

TOKEN	POS	NP chunk
We	PRP	B (Begin)
conducted	VBD	O (Outside)
a	DT	B (Begin)
multi-center	NN	I (Inside)
,	,	I (Inside)
randomized	VBN	I (Inside)
trial	NN	I (Inside)
comparing	VBG	O (Outside)

An example sentence:

"[We] conducted [a multi-center, randomized trial] comparing …"

Noun Phrase tagging:

- Manual labor using domain specific knowledge -

NP chunked sentence by YamCha:

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

NP tagging:

"[We] conducted [STUDY] comparing [DRUG] with [DRUG] for [THERAPY] of [DISEASE]."

Our Noun Phrase tag set:

【tag】	【covered concept】	【example】
DISEASE:	disease, symptom, virus	chronic hepatitis C
DRUG:	drug, chemical compound	interferon
STUDY:	clinical trial	clinical trial
THERAPY:	treatment, regimen	antiviral treatment
PATIENT:	participants in the trial	HBeAg-positive patients
TARGET:	endpoints	efficacy and safety
SCHEDULE:	time schedule of the trial	an additional 24 weeks
VALUE:	value of TARGET	significantly higher rates
NUMBER:	numeral expression	20 percent

Extraction of IE targets:

- using our manually written patterns -

A regular expression pattern for "Compared Treatment":
"compar∗" [DRUG] "with" [DRUG]

⬇ Apply the pattern to the example sentence.

"[We] conducted [STUDY]
comparing **[DRUG] with [DRUG] ..."**

⬇ Extract matched NPs and recover original texts.

IE Result:
Compared Treatment : **"peginterferon plus ribavirin"**
Compared Treatment : **"interferon plus ribavirin"**

Data used in our experiment:

- We downloaded the 50 most recent abstracts of clinical trials from the MEDLINE database on October 2004.
  - http://www.ncbi.nlm.nih.gov/entrez/query.fcgi
- To simplify the experiment, abstracts were selected from the medical area of hepatitis.

Evaluation Metrics in our IE experiment:

- We used four measures.
- For IE target entities,
  - Precision for entity extraction (Ent_pre)
  - Recall for entity extraction (Ent_rec)
- For Abstracts,
  - Precision for abstract summarization (Sum_pre)
  - Recall for abstract summarization (Sum_rec)

Results:

- We show two types of results:
  - IE from titles alone, and
  - IE from titles and main texts.
- The results from titles alone can be considered as the baseline,
  - because just putting together the titles is close to summarizing the articles.

Results of IE from titles alone and from titles and main texts:

		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%

Why performance isn't good?

- The patterns based on heuristics have no theoretical guarantee that they are correct.
- In the next, we show experimental results of sentence classification that might overcome the difficulties found in this IE experiment.

Experiment II

Sentence Classification

Our Sentence Classification task:

- Classifying sentences into "yes" or "no" with respect to whether the sentence includes IE targets or not.

NP tagged sentence:

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

Classification Result:

Compared Treatment: +1 (Yes)

Endpoint: -1 (No)

Patient Population: +1 (Yes)

Why Sentence Classification?

- We hypothesize that the IE accuracy can be increased
  - by classifying sentences and only performing IE on sentences that are most likely to contain relevant information.
- Furthermore, classifying or filtering sentences could be the rational step to get to our goal
  - in the respect that our final goal is summarization of clinical trial design information.

We used BACT, a state-of-the-art sentence classification algorithm.

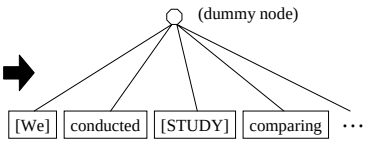
- Up to Noun Phrase tagging, the experiment process is same as IE experiment.
- We used BACT (Kudo and Matsumoto, 2004)
  - to acquire optimal classification patterns by machine learning, and
  - to classify sentences according to those automatically constructed patterns.

BACT learns from training data as ordered trees.

- In case of Bag-of-words (BOW) assumption:

An example sentence:

"[We] conducted [STUDY] comparing …"

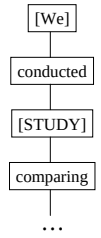


BACT learns from training data as ordered trees.

- In case of N-gram assumption:

An example sentence:

"[We] conducted [STUDY] comparing …"

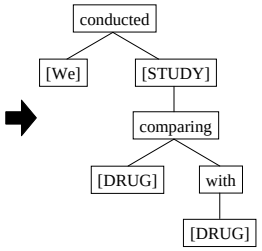


BACT learns from training data as ordered trees.

- In case of dependency grammar restriction:

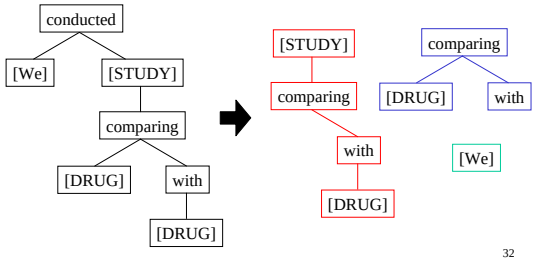
An example sentence:

"[We] conducted [STUDY] comparing [DRUG] with [DRUG] …"



BACT searches for sub-trees in each ordered tree.

- Searches all sub-trees comprehensively.



BACT ranks the sub-trees (calculates weights) :

- An example of dependency grammar restriction.

automatically constructed patterns by BACT that include "DRUG"	Compared Treatment	Endpoint	Patient Population
"PATIENT received DRUG"	0.048	–	–
"DRUG"	0.046	–	–
"TARGET of DRUG"	–	0.035	–
"DRUG, DRUG"	0.013	–	–
"received DRUG"	0.01	0.023	–
"of DRUG"	0.006	0.012	–
"with DRUG"	-0.004	–	-0.026
"to DRUG"	-0.013	–	-0.012
"in DRUG"	-0.019	–	–

BACT classifies sentences according to automatically constructed patterns.

NP tagged sentence:

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

Classification Result:

Compared Treatment: +1 (Yes)

Endpoint: -1 (No)

Patient Population: +1 (Yes)

Data used in our experiment:

- Same data as IE experiment.
- We downloaded the 50 most recent abstracts of clinical trials from the MEDLINE database on October 2004.
  - http://www.ncbi.nlm.nih.gov/entrez/query.fcgi
- To simplify the experiment, abstracts were selected from the medical area of hepatitis.

Evaluation Metrics:

- We used two measures.
- Precision = tp / (tp + fp)
- Recall = tp / (tp + tn)
  - tp means true positive, fp means false positive, and tn means true negative.
  - Precision is the correctness of the system when it classifies sentences to "yes".
  - Recall is the proportion of "yes" sentences that the system classifies to "yes".

Results (five-fold cross validation):

		Compared Treatment	Endpoint	Patient Population
# total sentence		562	562	562
# total "yes" sentence		90	76	55
BOW	precision	82.3%	81.5%	71.7%
	recall	70.8%	69.1%	64.7%
N-gram	precision	82.6%	85.7%	81.5%
	recall	71.7%	73.2%	81.5%
dependency	precision	86.8%	84.7%	75.2%
	recall	78.5%	72.2%	71.4%

Conclusions

Today's presentation, 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.

Future Work

- Construction of bigger corpora.
- Automate NP tagging.
- In the IE subtask,
  - identification of correspondence between entities and mentions.
- In the subtask of sentence classification using BACT,
  - improving parsing accuracy such that come from coordination structure or PP attachment ambiguity.

Thank you!