Rule-based Assertion Classification of Medical Problems in Clinical Narratives

Sunghwan Sohn, PhD, Sean P. Murphy, James J. Masanz, Vinod C. Kaggal, Jiaping Zheng
Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN

Abstract

This paper describes the rule-based approach of the Mayo Clinic team's work in the 2010 i2b2 Assertion Challenge. The system builds on Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES) which is available open source at www.ohnlp.org. The engineering framework for cTAKES is the open-source Unstructured Information Management Architecture (UIMA) platform. The original negation and status annotators in cTAKES have been modified or replaced by customized rule-based methods to accommodate the i2b2 assertion specifics. The training data consisting of 349 clinical records provided by the i2b2 challenge organizer have been used to develop assertion rules. Our highest micro average F-score of all assertion categories is 0.92. The system results suggest possible future improvement especially for the assertion categories "conditional" and "possible".

Introduction

The clinical narratives often contain medical problems of patients. An assertion is an attribute of the medical problem roughly classified as present, absent, and some uncertain categories. In this study we implemented rule-based methods to identify medical problems into one of six assertion categories as defined in i2b2 2010 assertion task[1]. Our system builds on a variation of Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES). cTAKES is a comprehensive, modular, extensible information extraction system geared specifically for the clinical narrative and released open source at www.ohnlp.org [1, 2]. The cTAKES pipeline consists of a sentence boundary detector, tokenizer, part-of-speech tagger, shallow parser and a named entity recognition module, all trained on clinical free-text data. A version of cTAKES has been in production at the Mayo clinic to process a repository of 60M+ clinical notes generated across three Mayo Clinic sites. A variation of cTAKES has been applied to a number of use cases, e.g. the extraction of smoking status for a peripheral artery disease study [3], prospective cohort identification for an unstable angina study, and medication extraction for a breast cancer treatment pharmacogenomics study [4 for preliminary results]. We based our entries for the i2b2 NLP challenges in 2006, 2008, and 2009 [5-7] on cTAKES. For the 2010 challenge, we also build upon cTAKES.

Methods

The i2b2 assertion classification has six categories (i.e., *present*, *absent*, *possible*, *conditional*, *hypothetical*, *associated with someone else*) and each is mutually exclusive. We have implemented UIMA annotators, using manually developed rules for each category except *present* since *present* was assigned as a default category. Each medical problem concept was assigned an assertion category whenever an assertion annotator satisfied the given rule conditions. The annotated training data for 349 patient records provided by the i2b2 organizer were used to develop a set of customized rules. Then the system applied a priority rule to determine the final assertion category. Based on the i2b2 organizer guideline, *associated with someone else* has the highest priority, followed by *hypothetical*, *conditional*, *possible*, and *absent*. If no assertion category is found, *present* is assigned as the default. Alternatively, we have investigated the different order for priorities, such as *absent* followed by *possible*

The first step of the assertion annotator is to determine the window that is anchored by the medical problem concept in order to limit context for searching for assertion indication words. For each assertion category we implemented customized rules to find certain patterns of indication words. The summaries of each rule set are set forth in the following discussion:

<u>Associated with someone else</u>

The window spans up to six words to the left of a given concept without counting stopwords. If predefined boundary words/punctuation (e.g., "patient" "his" "her" ';' '(' ')' etc.) exist within this span, the window is terminated before these boundary words/punctuation. Then, if indication words (e.g.,

[1] https://www.i2b2.org/NLP/Relations/Main.php

"father" "mother" "son" "daughter" etc.) exist within the window, *associated with someone else* is assigned. Additionally if a given concept is in a family history section, it is also assigned this same assertion category.

Absent

The window spans up to five words to the left or right of a given concept without counting stopwords. The left or right direction depends on *absent* indication words (e.g., left indication words: "not" "no" etc., right indication words: "resolved" "subsided" etc.)—i.e., left indication words must be within the left window and right indication words must be within the right window. If predefined boundary words/punctuation exist within this span, then the window span is terminated before these boundary words/punctuation. The determination of whether to assign the *absent* category is done in two parts. The first is based on indication words like the other annotators. If indication words (e.g., "not" "deny" "no" "negative" etc.) exist within the window, *absent* is assigned. The second is a set of rules that were developed based on occurrences of negation within the training set that the first annotator missed: 1) identifies words and phrases that are negated based solely on a string match to words and phrases that had been found in the training set such as "uncomplicated", "sclera anicteric", 2) identifies negative signs, such as in "- rash" as an indicator that "rash" is negated, 3) identifies abbreviations and acronyms that have an inherent negative meaning such as NAD for "no acute distress", 4) handles terms that include a negative prefix such as "afebrile" and "non-tender". To attempt to differentiate between words that simply start with the letter '*a*' and words where that letter is the prefix meaning *without*, the rest of the word is looked up in a dictionary (a subset of UMLS). If the word is found in the dictionary without the prefix, it is assumed the full word should be negated.

Hypothetical

The window span depends on indication words and could be the left side of a given concept, the left side of a given concept up to a semi-colon, or a sentence containing a given concept. The rules search for indication words (e.g., "prn" "if" "return" "report", "hold for" "to prevent" "risk for" etc.) with conditions attached (e.g., not allowing certain POS tagged indication words). Also, sentences starting with certain phrases, such as "Call to doctor", are used to identify zones for hypothetical concepts. Concepts in these zones are classified as *hypothetical* even though no indication word is found.

Possible

The window spans up to two or six words to the left or right of a given concept without counting stopwords. The window direction and size depends on indication words. For certain indication words and punctuation (i.e., "versus", "vs", and "?"), both left and right windows are examined. If predefined boundary words/punctuation exist within the span, the window span is terminated before these boundary words/punctuation. Then, if left indication words (e.g., "rule out" "may be" "question of" "probably" "possibly" "suspected" "likely" etc.) or right indication words (e.g., "not excluded" "is suspected" "is likely" etc.) exist within the window, *possible* is assigned.

Conditional

The window is simply the sentence that contains a given concept. A variety of rules that search for certain patterns between concept and indication words (e.g., conjunction: "when" "while" "after" "before", preposition: "on" "at" "to", verb: "cause" lead" "induce" "give") with specified conditions that consider location and POS tag of words were developed. For example, "Penicillin for which she reports a rash and *dyspepsia* when taking uncoated aspirin". Here the concept *dyspepsia* is classified as *conditional* based on the pattern of indication words (underlined). An additional rule finds concepts that themselves contain an allergy indication string. For example, the concept, *droperidol allergy* contains "allergy", and so it is classified as *conditional*.

Results and Discussion

Table 1 shows the evaluation results of our system's best performance on test data of 477 patient records provided by the i2b2 organizer. The results were obtained by running the evaluation script provided by the i2b2 organizer. The system produced a micro average F-score of 0.92. Interestingly, the priority rule used for this result was slightly different from the i2b2 guideline—i.e., here we used *absent* followed by *possible* (In the priority rule based on the i2b2 guideline, *possible* has a priority over *absent*). When we used the original i2b2 priority rule, the micro average F-score was 0.917 (*absent*: 0.921, *possible*: 0.625).

In Table 1, the *present* category shows the highest F-score, and the *conditional* category shows the lowest F-score. The major reasons for low recall in *conditional* were: 1) overly specialized rules using the limited set of words along with indicative conjunctions and preposition, 2) concepts that appear without any indication words, but appeared in the allergies section. However, many FP cases produced by our system seemed to be questionable (i.e., seemed to be correct). For example (*italic* denotes concept), "*Allergies* to penicillin", "Because the patient had *previously known poor reactions* to statin…", "*Dyspneic* at rest" "*Pain* worse with movement", "he does have *a thrombocytopenic reaction* to Heparin" "*Shortness of breath* occurring with activity and at rest." "He had no *gag* to tube manipulation". If all questionable cases were treated as being correct, the precision of *conditional* would be much higher. The low recall of *possible* was largely due to: 1) not covering all variations of indication words (e.g., our system used "suspicion" but not used "suspicious"), 2) having incorrectly associated boundary word, 3) not using some indication words that showed inconsistent annotation in the gold standard (e.g., the concepts occurring with "consistent with" were sometimes annotated as *possible*, but sometimes not annotated as such.).

Assertion	**Recall**	**Precision**	**F-score**
Present	0.945	0.949	0.947
Absent	0.952	0.903	0.927
Possible	0.606	0.681	0.641
Conditional	0.298	0.418	0.348
Hypothetical	0.816	0.842	0.829
Associated	0.959	0.837	0.894
Micro avg.	0.920	0.920	0.920

Table 1. Assertion evaluation results on the test data

		system					
		pres	**abs**	**poss**	**cond**	**hypo**	**asso**
	pres	12314	334	222	60	77	18
g	**abs**	134	3435	20	5	9	6
o	**poss**	299	21	535	5	23	
l	**cond**	109	8	2	51	1	
d	**hypo**	118	3	7	1	585	3
	asso	5	1				139

Table 2. Confusion matrix of assertion evaluation results on the test data

Table 2 shows the confusion matrix for the results in Table 1. In all assertion categories except *present* most incorrect cases were confused with *present* concepts (i.e., most false negative and false positive cases occurred in the *present* concepts). In certainty-type assertion, *possible* and *hypothetical* were confused with each other somewhat (i.e., 23 false negative *possible* cases misclassified as *hypothetical*).

One interesting avenue to explore is to use more sophisticated window that is based on syntactics/semantics rather than simply using word tokens. Another avenue is the use of machine learning techniques for certainty-type assertion categories such as *possible*, *conditional*, and *hypothetical* since there are many varied patterns to express these categories.

Conclusion

We developed annotators for a variation of cTAKES for the assertion task of medical problems from the clinical narratives to accommodate the i2b2 clinical NLP challenge. The portability of cTAKES allowed us to implement customized UIMA annotators for each assertion category defined in the i2b2 assertion task.

References

[1] Savova G, Masanz J, Ogren P, Zheng J, Sohn S, Kipper-Schuler K, et al. Mayo Clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc. 2010;In press.
[2] Savova GK, Kipper-Schuler K, Buntrock JD, Chute CG. UIMA-based clinical information extraction system. LREC 2008: Towards enhanced interoperability for large HLT systems: UIMA for NLP; 2008.
[3] Sohn S, Savova GK. Mayo Clinic Smoking Status Classification System: Extensions and Improvements. AMIA Annual Symposium; San Francisco, CA: 619-623; 2009.
[4] Savova GK, Olson J, Murphy SP, Couch F, Goetz M, Ingle J, et al. Pharmacogenomic study of tamoxifen and aromatase inhibitors in women with breast cancer: drug extraction through natural language processing. *PGRN meeting*. Rochester, MN 2009.
[5] Savova GK, Clark C, Zheng J, Cohen K, Murphy SP, Wellner B, et al. The Mayo/MITRE system for discovery of obesity and its comorbidities. *AMIA 2nd i2b2 challenge workshop*. Washington DC 2008.
[6] Savova GK, Murphy SP, Masanz JJ, Zheng J, Sohn S. Mayo Clinic NLP System for Medication Extraction. Third I2B2 Shared-Task and Workshop - Challenge in NLP for Clinical Data - Medication Extraction Challenge; San Francisco, CA; 2009.
[7] Savova GK, Ogren PV, Duffy PH, Buntrock JD, Chute CG. Mayo Clinic NLP system for patient smoking status identification. J Am Med Inform Assoc. 2008;15:25-8.