UWM: A Simple Baseline Method for Identifying Attributes of Disease and Disorder Mentions in Clinical Text

Omid Ghiasvand
University of Wisconsin-Milwaukee
Milwaukee, Wisconsin
ghiasva2@uwm.edu

Rohit J. Kate
University of Wisconsin-Milwaukee
Milwaukee, Wisconsin
katerj@uwm.edu

Abstract

In this paper the system that was developed by Team UWM for the Task 14 of SemEval 2015 competition is described. Task 14 included two tasks: Task 1 was identification of disorder mentions and their normalization, and Task 2 was identification of the following attributes for disorder mentions: the CUI of the disorder, negation indicator, subject, uncertainty indicator, course, severity, conditional, generic indicator, and body location. For Task 1, an earlier system was applied that uses Conditional Random Fields (CRFs) for disorder recognition and learned edit distance patterns for normalization. Task 2 was implemented by a simple method that finds the attribute terms around the disease mentions by matching them in the training data. Among all participants Team UWM was ranked fourth in Task 1, fourth in Task 2A (over gold-standard mentions) and third in Task 2B (over extracted mentions).

1 Introduction

Automated extraction tools are crucial for managing huge amount of clinical texts. These tools have the potential to enable many automated applications in healthcare. The Task 14 of SemEval 2015 was designed to serve as a platform for evaluating one such extraction tool. Its Task 1 involved extracting and normalizing disorder mentions from clinical text and its Task 2 involved assertion identification for the mentions.

Task 1 is challenging because there is a lot of variability in which diseases and disorders are mentioned in clinical text and hence a pre-defined list of mentions is not sufficient to extract them. The task also required normalizing the extracted mentions by mapping them to UMLS CUIs if they exist in the SNOMED-CT part of UMLS and are marked as disease/disorder, otherwise they were to be declared as "CUI-less." This normalization process is also challenging because disorder names are frequently mentioned in modified forms which prevents them from exactly matching the concepts in UMLS. Task 2 required finding certain attributes for the mentions and finding the spans of these attributes in text. This task is also challenging due to the variability in which attributes are attributed to disease and disorder mentions in clinical text.

Our team, UWM, participated in both Task 1 and Task 2. For Task 1, we used the same system that we had previously developed for the Task 7 of SemEval 2014 (Ghiasvand and Kate 2014). For Task 2, we used a simple method that finds attributes of mentions by first collecting lists of attribute terms from the training data and then matching in this list. The nearest attribute terms to a mention are assigned to that mention. The attribute terms are normalized by finding their normalized values in the training data. Despite being simple, this method gave competitive results. The methods used in this paper are described in more details in the next section.

Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), pages 385–388,
Denver, Colorado, June 4-5, 2015.

2 Methods

2.1 Task 1

We briefly describe the system we had developed for Task 7 of SemEval 2014 (Ghiasvand and Kate 2014) which we used for Task 1. We treated disorder mention extraction as a standard sequence labeling task with "BIO" (Begin, Inside, Outside) labeling scheme. The model was trained using Conditional Random Fields (Lafferty et al., 2001) with various types of lexical and semantic features that included MetaMap (Aronson and Lang 2010) matches. These features are fully described in (Ghiasvand, 2014). This model was also inherently capable of extracting discontinuous disorder mentions. To normalize disorder mentions, our system first looked for exact matches with disorder mentions in the training data and then in the UMLS. If no exact match was found, then suitable variations of the disorder mentions were generated based on possible variations of disorder mentions learned from UMLS synonyms. These variations were learned in the form of edit distance patterns (Levenshtein 1966) using a novel method described in (Ghiasvand and Kate 2014).

2.2 Task 2

In this task, attributes related to disease or disorder mentions were to be identified along with their normalized values and spans in the text (Bodenreider, 2003). There were nine attributes related to each disorder mention for this task which were: the CUI of the disorder (same as Task 1), negation indicator, subject, uncertainty indicator, course, severity, conditional, generic indicator, and body location.

For identifying CUI attribute, we used the same normalization method that we had used for Task 1. For identifying the rest of the attributes, we used a simple matching method based on the training data for Task 2. The method first collects a list of attribute terms from the training data for each attribute type. For example, if "likely arising from", "lower suspicion of", and "possibly secondary" are marked as uncertainty terms in the training data then they will be included in our list of attribute terms for uncertainty. Table 1 lists the number of attribute terms thus collected from the training data for each of the attribute type. The only attribute that has many more values than other attributes is body location. For this attribute we used not only training data but also UMLS matches of body locations. Our training dataset consisted of combined training and development dataset parts, but when we collected these terms from only the training part, we found that a majority of these match in the development part. Thus we determined that only a small list of terms are frequently used to indicate most of the attributes of disease and disorder mentions and decided to use the simple matching method.

Our method identifies attributes of disease and disorder mentions as follows. Using the list of attribute terms, it first identifies attribute terms in the same sentence in which the mention is included. For each attribute type, the nearest attribute term (if present) is associated with the mention. The normalized value of the attribute is then simply obtained from the training data. For example the term "increasingly" in the course attribute type has normalized value "increased" in the training data, and the term "maternal aunt" in the subject attribute type has the normalized value "family_member". Hence if "increasingly" is the course attribute term found nearest to a disease mention in the test data then its course attribute will be assigned the value "increased". Similarly if "maternal aunt" is found as the nearest subject attribute term then its value will be assigned as "family_member".

Task 2 had two subtasks. In Subtask 2A, gold-standard disease and disorder mentions were provided and in Subtask 2B the mentions were to be first extracted by the system, hence it combined Task 1 and Subtask 2A.

Attribute	**Number of attribute terms in training data**
Conditional (CND)	154
Course (COU)	168
Generic (GEN)	45
Negation (NEG)	139
Severity (SEV)	92
Subject (SUB)	33
Uncertainty (UNC)	295
Body Location (BL)	1108

Table 1: Number of attribute terms for each attribute type in the training data.

3 Results

The training, development and test datasets of SemEval 2015 Task 14 had 298, 133 and 100 clinical notes respectively. We formed our training dataset by combining training and development datasets. The clinical notes contained different types of notes including de-identified discharge summaries, electrocardiogram, echocardiogram and radiology reports (Pradhan et al., 2013). The extraction and normalization performance in Task 1 was evaluated in terms of precision, recall and F-measure for strict (exact boundaries) and relaxed (overlapping boundaries) settings. Table 2 shows the results of this task. In this task, based on relaxed F-score, we got second rank, and based on strict F-score we got fourth rank considering only the best run of each participating team.

	Precision	**Recall**	**F-score**
Strict	0.773	0.699	0.734
Relaxed	0.809	0.731	0.768

Table 2: Results of Task1 (mention extraction and normalization).

For the Task 2A, unweighted and weighted accuracies were used as evaluation measures. For each disorder, a per-disorder, unweighted accuracy is computed, which represents the ability to identify all the slots for that disorder. The unweighted accuracy is the average of the per-disorder unweighted accuracy over all the disorders in the test set. For each disorder, a weighted per-disorder accuracy is computed, which represents the ability to identify all the slots for that disorder.

For Task 2B, the following evaluation measures were used: F-score for span identification, unweighted accuracy (which is same as the unweighted accuracy described in Task 2A computed over the true-positive identified disorders), and weighted accuracy (which is same as the weighted accuracy described in Task 2A computed over the true-positive identified disorders).

In Table 3 and 4, the results of these two subtasks are shown. Table 5 shows the results separately for each attribute type for both the subtasks. In Task 2A, except for the body location attribute our method got above eighty percent accuracy on all other attributes and above ninety percent on three of them. We also want point out that for the attribute type CUI we got 0.911 accuracy in Task 2A which is only slightly behind the best accuracy of 0.918 got by another team.

The reason our system got a very low accuracy for the body location attribute is because we forgot to include the CUI values for this attribute during the competition. This then also adversely affected our overall performance scores. Overall, in Task 2A our team ranked fourth and in Task 2B our team ranked third considering the best run of each participating team.

Our method for Task 2 was found to be competitive despite being very simple. For example, this simple matching scheme got 92.4% accuracy for negation attribute while the best team got 97.5% accuracy in Task 2A. Hence this method forms a very good baseline for comparing more sophisticated methods. It can also serve as a method that provides potential attributes which then can be tested and filtered by machine learning methods.

F-Score	**Accuracy**	**F*A**	**Weighted-Accuracy**	**F*WA**
1.00	0.859	0.859	0818	0.818

Table 3**:** Results of Task 2A.

F-Score	**Accuracy**	**F*A**	**Weighted-Accuracy**	**F*WA**
0.893	0.852	0.761	0.798	0.713

Table 4: Results of Task 2B.

Attribute	**Accuracy (Task 2A)**	**Accuracy (Task 2B)**
BL	0.531	0.551
CUI	0.911	0.858
CND	0838	0.839
COU	0.802	0.793
GEN	0.836	0.845
NEG	0.924	0.931
SEV	0.895	0.905
SUB	0.933	0.929
UNC	0.831	0.837

Table 5: Accuracy for each attribute type in Task 2A and Task 2B.

4 Conclusion and future work

We participated in Task 14 of SemEval 2015 which involved disorder mention extraction, normalization, and attribute identification. Our system used conditional random fields to extract

disorder mentions and edit distance patterns for normalization of the extracted mentions. For identifying attributes, we used a simple matching based method using the training data. Our team preformed competitively on all the subtasks. In future, we plan to combine machine learning methods with our simple matching method for attribute identification.

Acknowledgment

This work was supported by grant UL1RR031973 from the Clinical and Translational Science Award (CTSA) program of the National Center for Research Resources and the National Center for Advancing Translational Sciences.

References

Aronson A. R., and Lang F. M. An overview of MetaMap: historical perspectives and recent advances. *Journal of American Medical Informatics Association*. 2010;17(3):229–36.

Bodenreider, O. and McCray, A. 2003. *Exploring semantic groups through visual approaches*. Journal of Biomedical Informatics, 36(2203): pp. 414-432.

Omid Ghiasvand, 2014. *Disease Name Extraction from Clinical Text Using Conditional Random Fields*, Thesis and Dissertation, University of Wisconsin-Milwaukee, Milwaukee, USA.

Omid Ghiasvand and Rohit J. Kate, 2014. *UWM: Disorder Mention Extraction from Clinical Text Using CRFs and Normalization Using Learned Edit Distance Patterns*, in Proceeding of the Eight International Workshop on Semantic Evaluations (SemEval 2014), pages 828-832, Dublin, Ireland.

John Lafferty, Andrew McCallum, and Fernando Pereira. 2001. *Conditional Random Fields: Probabilistic models for segmenting and labeling sequence data*. In Proceedings of 18th International Conference on Machine Learning (ICML-2001), pages 282–289, Williamstown, MA.

Vladimir I Levenshtein. 1966. *Binary codes capable of correcting deletions, insertions and reversals.* In Soviet physics doklady, volume 10, page 707.

Sameer Pradhan, Noemie Elhadad, B South, David Martinez, Lee Christensen, Amy Vogel, Hanna Suominen, W Chapman, and Guergana Savova. 2013. *Task 1: ShARe/CLEF eHealth Evaluation Lab 2013*. Online Working Notes of CLEF, CLEF, 230.