

# UTH-CCB: The Participation of the SemEval 2015 Challenge – Task 14

**Jun Xu, Yaoyun Zhang, Jingqi Wang, Yonghui Wu, Min Jiang,  
Ergin Soysal, and Hua Xu**

School of Biomedical Informatics, The University of Texas Health Science Center at Houston  
Houston, TX, USA

{Jun.Xu, Yaoyun.Zhang, Jingqi.Wang, Yonghui.Wu, Min.Jiang,  
Ergin.SoySal, Hua.Xu}@uth.tmc.edu

## Abstract

This paper describes the system developed by the University of Texas Health Science Center at Houston (UTHealth), for the 2015 SemEval shared task on “Analysis of Clinical Text” (Task 14). We participated in both sub-tasks: Task 1 for “Disorder Identification”, which aims to detect disorder entities and encode them to UMLS (Unified Medical Language System) CUI (Concept Unique Identifier) and Task 2 for Disorder Slot Filling, where the task is to identify normalized value for modifiers of disorders. For Task 1, we developed an ensemble approach that combined machine learning based named entity recognition classifiers with MetaMap, an existing symbolic biomedical NLP system, to recognize disorder entities, and we used a general Vector Space Model-based approach for disorder encoding to UMLS CUIs. To identify modifiers of disorders (Task 2), we developed Support Vector Machines-based classifiers for each type of modifier, by exploring various types of features. Our system was ranked 3<sup>rd</sup> for Task 1 and 1<sup>st</sup> for the Task 2 (both 2A and 2B), demonstrating the effectiveness of machine learning-based approaches for extracting clinical entities and their modifiers from clinical narratives.

## 1 Introduction

Natural language processing (NLP) plays a critical role in unlocking important patient information from narrative clinical texts, to support various clinical applications such as decision support systems and translational research. One of the very important tasks for clinical NLP research is to extract clinical concepts such as diseases and treatments. Many clinical NLP systems such as

MedLEE system (Friedman et al., 1994), MetaMAP system (Aronson and Lang, 2010) and cTAKES system (Savova et al., 2010), have been developed to extract these important clinical concepts from text.

A number of shared tasks for clinical concepts extraction have been organized by different entities, including i2b2 (The Center for Informatics for Integrating Biology and the Bedside), ShARe/CLEF eHealth Evaluation Lab, and SemEval (International Workshop on Semantic Evaluation) (Kelly et al., 2014; Pradhan et al., 2014; Suominen et al., 2013; Uzuner et al., 2011). These challenges have greatly promoted clinical NLP research by building benchmark datasets and innovative methods. The 2015 SemEval Shared Task 14, entitled “Analysis of Clinical Text”, is to identify disorders and their modifiers from clinical text, which is an extension of the SemEval-2014 challenge. The 2015 SemEval challenge consists of two subtasks: Task 1 - disorder recognition, where disorder entities need to be detected and normalized to UMLS CUIs, and Task 2 - disorder slot filling, where the normalized value for nine types of modifiers of disorders are to be identified. Task 2 is further divided into two subtasks: 1) Task 2A – identifying modifiers based on gold standard disorders; and 2) Task 2B – identifying modifiers based on disorders recognized by our system, an end-to-end evaluation. In this paper, we describe our approaches and results for both tasks.

## 2 Methods

### 2.1 Datasets

For this shared task, organizers prepared three datasets: 1) training set - 298 clinical documents, 2) development set - 133 documents and 3) test set – 100 documents. We developed our models using the

training set and optimized parameters using the development set. For final submissions on the test set, we combined training and development sets to build the machine learning classifiers.

## 2.2 Task 1 – Disorder Identification

The disorder identification consists of two subtasks: 1) recognize disorder entities, and 2) encode recognized disorder entities to concept IDs (CUIs) in UMLS (limited to SNOMED-CT). We describe our approaches for both steps below:

**Disorder Entity Recognition** - The disorder recognition task is a typical named entity recognition (NER) task. We developed two machine learning based NER models, including the Conditional Random Fields (CRFs) (Lafferty et al., 2001) and the Structural Support Vector Machines (SSVMs). The CRFsuite package (Okazaki, 2007) and SVM<sup>hmm</sup>\* are used for CRFs and SSVMs implementations, respectively. In addition, we also developed hybrid models that combine the two machine learning models with an existing symbolic biomedical NLP system – MetaMap. We developed hybrid systems for disorder recognition by adopting two previously developed ensemble learning strategies, including ensemble<sup>ML</sup> and ensemble<sup>MV</sup>, which were originally developed in our participation of the SemEval-2014 (Zhang et al., 2014). The ensemble<sup>MV</sup> approach follows the majority voting strategy to combine the three systems. The ensemble<sup>ML</sup> approach trains an SVM classifier to combine the predictions from the three systems.

We adopted the features engineered in the previous participation of SemEval 2014 (Zhang et al., 2014), including: word-level features, such as bag-of-words; linguistic features; and discourse features, such as section name in the clinical notes and type of the notes (e.g. ‘DISCHARGE\_SUMMARY’). In this challenge, we further explored the deep neural network (DNN) based word embeddings. We obtained word embeddings by training a deep neural network (Collobert et al., 2011) from the unlabeled MIMIC II corpus (about 3G clinical notes) provided by the SemEval organizers.

**Disorder Entity Encoding** - We adopted the same Vector Space Model (VSM) approach developed for the SemEval-2014 to encode the disorder to UMLS/SNOMED-CT CUIs (Zhang et al., 2014).

This is a general approach to encode clinical entities to UMLS CUIs, without utilizing training samples provided by this task.

## 2.3 Task 2 – Disorder Slot Filling

The task is to identify eight types of disorder modifiers, including negation indicator (NI), subject class (SC), uncertainty indicator (UI), course class (CC), severity class (SV), conditional class (CO), generic class (GC) and body location (BL). For each of the first seven types of modifiers, we built SVMs-based individual classifiers. The implementation of SVMs in LibShortText package (Yu et al., 2013) was used for this purpose. The LibShortText package is an open source library for large-scale short-text classification.

We systematically extracted the following features to train SVMs classifiers, including:

1). N-gram features. All unigrams and bigrams in the sentence were extracted as features.

2). Context words with position and direction (left or right) information. Here we describe the features using the following sentence: “*patient said he has no acute distress before*”. There is one disorder (‘distress’) in this sentence.

**Group-1 features:** context words within the window size of 1 to disorder: [‘acute\_L1’, ‘before\_R1’]

**Group-2 features:** context words within the window size of 4 to disorder: [‘he\_L4’, ‘has\_L4’, ‘no\_L4’, ‘acute\_L4’, ‘before\_R4’]

**Group-3 features:** context words within window size range of 5 to 8: [‘patient\_L8’, ‘said\_L8’]

3). Lexicon features, including word lists for negation, pseudo-negation, conjunction, condition, uncertainty, subject, severity, and course.

4). Dependency relation features. We used the Stanford Parser to generate dependency relations of a sentence. We only counted dependency relations where a target disorder is the governor or the dependent in the relation. We extracted all these syntactic relations as features.

5). Section names, e.g. ‘Family History’.

The final set of features was optimized based on the performance of cross-validation of the training set for each modifier.

\* [http://www.cs.cornell.edu/people/tj/svm\\_light/svm\\_hmm.html](http://www.cs.cornell.edu/people/tj/svm_light/svm_hmm.html)

The body location modifiers require specification of the text spans and the corresponding UMLS CUIs. Therefore, we first built a NER system for body location entities and then applied the same encoding approach, similar to the methods used in disorder identification task. We also constructed a comprehensive body location dictionary from UMLS and WordNet (Miller, 1995). The relative positions of the target disorder and the candidate body location were extracted as features (e.g., whether the body location is part of the target disorder). For body location encoding, we extended VSM-based lookup method by adding a regression-based re-ranking layer trained from the training corpus.

## 2.4 Submissions and Evaluation

We combined training and development datasets to build our final models for all tasks. Since each task allows for three submissions, we tried different strategies for the three runs. For Task 1, run 0 and run 1 used the ensemble<sup>MV</sup> method to get better F1; while run 2 used the ensemble<sup>ML</sup> method to get higher precision, in disorder entity recognition. For Task 2A and Task 2B, run 0 and run 2 used two sets of parameters optimized for better weighted performances; while run 1 used a set of parameters optimized for un-weighted performance. For body location recognition, only run 2 of Task 2A used SSVMs model, all other runs used CRFs models for better prediction.

The evaluation metrics for this task include F-1 score (strict vs. relaxed), un-weighted accuracy, and weighted accuracy etc., as defined by the organizers. For more details, please refer to the task description paper or the task website<sup>†</sup>.

## 3 Results and Discussion

For Task 1, the main evaluation scores were strict F1. Table 1 shows the overall performance of three runs of our system in Task 1 as reported by the organizer, where ‘P’, ‘R’, ‘F’ denotes precision, recall, and F1 score respectively. Our best run of Task 1 ranked 3<sup>rd</sup> among all participants. Our disorder entity recognition step actually achieved the highest F1 of 0.927 under ‘relaxed’ criterion (please see Table 2). The performance of disorder encoding was not as good as other top performed teams in task

1, because we used a general encoding module that did not use the CUI annotations in the training/development set for training.

Run	Strict			Relaxed		
	P	R	F	P	R	F
0	.748	.713	.730	.777	.741	.759
1	.748	.713	.730	.777	.741	.759
2	.778	.696	.735	.797	.714	.753

Table 1. The performances of the three runs of our system on Task 1.

As reported by the organizers, our system achieved the best performance in Task 2, both for Task 2A - slot filling given gold-standard disorder spans and Task 2B - end-to-end system for disorder span identification and slot filling. Table 2 shows the overall performance of our systems in Task 2A and Task 2B. ‘F’, ‘A’, and ‘WA’ denotes ‘relaxed’ F1 score for disorder entity recognition, overall un-weighted and weighted accuracy respectively.

Task	Run	F	A	F*A	WA	F*WA
2A	0	1.00	.943	.943	<b>.886</b>	<b>.886</b>
	1	1.00	<b>.953</b>	<b>.953</b>	.876	.876
	2	1.00	.943	.943	<b>.886</b>	<b>.886</b>
2B	0	<b>.927</b>	.940	.872	.872	<b>.808</b>
	1	<b>.927</b>	<b>.949</b>	<b>.880</b>	.862	.800
	2	.907	.943	.855	<b>.880</b>	.798

Table 2. The overall performances of our system on Task 2.

## 4 Conclusion

In this paper, we described our participation in the SemEval-2015 challenge – Task 14 “Analysis of Clinical Text”. Our system was among the top ranked systems (ranked 3<sup>rd</sup> for Task 1, 1<sup>st</sup> for Task 2A and Task 2B). These results show that machine learning based methods, integrated with medical domain specific features, could reasonably identify disorders and associated modifiers from clinical narratives.

## Acknowledgments

This study is supported in part by grants from NLM 2R01LM010681-05, NIGMS 1R01GM103859 and

<sup>†</sup> <http://alt.qcri.org/semEval2015/task14/index.php>

1R01GM102282, and CPRIT R1307. The first author is partially supported by NSFC 61203378.

In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. pp.802-806. Dublin, Ireland.

## References

- Aronson, A. R., and Lang, F. M. 2010. An Overview of MetaMap: Historical Perspective and Recent Advances. *Journal of the American Medical Informatics Association*, 17(3):229-236.
- Collobert, R., Weston, J., et al. 2011. Natural Language Processing (Almost) from Scratch. *Journal of Machine Learning Research*, 12:2493-2537.
- Friedman, C., Alderson, P. O., et al. 1994. A General Natural-Language Text Processor for Clinical Radiology. *Journal of the American Medical Informatics Association*, 1(2):161-174.
- Kelly, L., Goeuriot, L., et al. (2014). Overview of the ShARe/CLEF eHealth Evaluation Lab 2014. In E. Kanoulas, M. Lupu, et al. (Eds.), *Information Access Evaluation. Multilinguality, Multimodality, and Interaction* (Vol. 8685, pp.172-191): Springer International Publishing.
- Lafferty, J. D., McCallum, A., et al. 2001. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proceedings of the Eighteenth International Conference on Machine Learning*. pp.282-289.
- Miller, G. A. 1995. WordNet: A Lexical Database for English. *Communications of the ACM*, 38(11):39-41.
- Okazaki, N. 2007. *CRFsuite: a Fast Implementation of Conditional Random Fields (CRFs)*. <http://www.chokkan.org/software/crfsuite/>
- Pradhan, S., Elhadad, N., et al. 2014. SemEval-2014 Task 7: Analysis of Clinical Text. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. pp.54-62. Dublin, Ireland.
- Savova, G. K., Masanz, J. J., et al. 2010. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): Architecture, Component Evaluation and Applications. *Journal of the American Medical Informatics Association*, 17(5):507-513.
- Suominen, H., Salanterä, S., et al. (2013). Overview of the ShARe/CLEF eHealth Evaluation Lab 2013. In P. Forner, H. Müller, et al. (Eds.), *Information Access Evaluation. Multilinguality, Multimodality, and Visualization* (Vol. 8138, pp.212-231): Springer Berlin Heidelberg.
- Uzuner, O., South, B. R., et al. 2011. 2010 i2b2/VA Challenge on Concepts, Assertions, and Relations in Clinical Text. *Journal of the American Medical Informatics Association*, 18(5):552-556.
- Yu, H.-F., Ho, C.-H., et al. 2013. *LibShortText: A Library for Short-text Classification and Analysis*. <http://www.csie.ntu.edu.tw/~cjlin/libshorttext>
- Zhang, Y., Wang, J., et al. 2014. UTH\_CCB: A Report for SemEval 2014 – Task 7 Analysis of Clinical Text.