Extracting Medication Information from Clinical Text

Faisal Farooq F.FAROOQ@SIEMENS.COM
Siemens Medical Solutions, Malvern, PA

Manavender R. Malgireddy MRM42@BUFFALO.EDU
State University of New York, Buffalo, NY

Shipeng Yu SHIPENG.YU@SIEMENS.COM
Siemens Medical Solutions, Malvern, PA

Balaji Krishnapuram BALAJI.KRISHNAPURAM@SIEMENS.COM
Siemens Medical Solutions, Malvern, PA

Abstract

One of the important pieces of information in a patients clinical record is the information about medications. This information may consist of a drug name, its dosage, mode of administration, frequency, duration and reason for prescribing the medication. Unfortunately, with limited use of electronic Physician Order Entry (POE) systems, much of this information is present embedded in unstructured clinical notes. It is a manually intensive process to read the clinical notes and extract such information. With the penetration of Electronic Medical Record (EMR) systems in healthcare facilities continuously increasing, the need for such systems is also on the rise. This paper explains in detail a system developed to extract such information. We have trained a Conditional Random Field model to extract such information. We present results on a database made available by the Informatics for Integrating Biology and the Bedside (i2b2) Center as part of their Third NLP Shared Task.

1. Introduction

Hospitals store clinical information about patient visits in the form of medical records. These records are used by physicians, nurses and other hospital staff to extract information about the patient. Some of the most important visit level documents are the History and Physical Record (H&P) and the Discharge summary. The contain information about medications that were prescribed to the patient (either new or continued from past). This information is indispensable for various purposes notably diagnoses, treatment, retrospective analysis for quality and clinical adherence, research, clinical trials and so on. With the recent impetus of Electronic Medical Record systems, the trail of electronic documentation for a single patient visit has increased manifold. However, much of this information is available in unstructured (free) text. Typically, the medication information is embedded deep into the text in the type of documents mentioned above. In spite of the recommendations to push this information into structured data, much of it is still unstructured. This is often due to the fact that medications are reconciled (home, hospital stay etc.) at discharge time e.g. by stopping, continuing or changing. This reconciliation information is easy to enter using spoken text. It is a manually intensive process to read and extract such information from a record. If this process of extraction is automated and results are presented in a structured format, it will save effort and improve the quality of care. According to a report from Leapfrog Group, only 7% of US hospitals actually use POE (eMAR) systems (Leapfrog, 2008). Due to its significance, the Informatics for Integrating Biology and the Bedside (i2b2) Center (NIH funded effort) organized a challenge (I2B2) that targeted extracting medication information from a discharge summary. This paper explains in detail the system developed to extract such information. After briefly discussing the related work, we present the data, our methodology

Appearing in *Proceedings of the* 28^{th} *International Conference on Machine Learning*, Bellevue, WA, USA, 2011.

and the results obtained.

2. Related Work

Information extraction from clinical text has recently received a lot of attention. Researchers are applying machine learning and natural language processing for text mining in systems like BADGER (Soderland et al., 1995), MedIE (Zhou et al., 2006) and CLEF (Roberts et al., 2008). Friedman et al (Friedman & Hripcsak, 1999) discuss the potential of using NLP techniques in the medical domain, and also provides a comparative overview of the state-of-the-art NLP tools applied to biomedical text. Literature (Chapman et al., 2001; Hirschman et al., 2002; Friedman & Hripcsak, 1999) provides a survey of various approaches to information extraction from biomedical text including named entity tagging and extracting relationship between different entities and between different texts.

In general the systems that have been designed by various researchers can be classified into two broad categories i) Rule Based and ii) Learning Based. Rule based systems (Xu et al., 2010; Chhieng et al., 2007; Levin et al., 2007) encode knowledge into grammars, if-else rules and regular expressions. They can be easy to implement, however, suffer from major drawbacks like hard maintainability and very low generalization. The rules engine is hard to maintain and number of rules can increase exponentially. On the other hand learning based systems (Patrick & Li, 2009; Li et al., 2009) take advantage of data and the inherent statistical knowledge. They do not require any hand crafting of rules and are easy to maintain. However, they require hand labeled data on which these systems can be trained which can be laborious to create. Given the nature of the problem where it is not feasible to encode diverse rules into a small subset, it is better, however to invest on-time effort in creation of such datasets (often called ground truth). Jagannathan et al (Jagannathan et al., 2008) assessed four commercial NLP engines for their ability to extract medication information. In most systems it is noted that whereas they extract medication names with high accuracies, the associated information like dosage, frequency, duration and reasons etc. are hard to extract and relate to the corresponding medications. We propose a learning based method which uses statistical and natural language features that are easy to extract and do not require high degree of domain expertise. The method is highly scalable and can be easily trained as more data is available. We conducted experiments report results with the data made available by I2B2 so that there can be a fair comparison with other systems.

3. Pre-Processing

A conditional random field (CRF) (Lafferty et al., 2001) is a type of discriminative undirected probabilistic model most often used for the labeling or parsing of sequential data, such as natural language text. The CRF model works on a series of observed tokens X_i and outputs a distribution of the most-likely label sequence Y_i for the observed tokens. So the input for the CRF will be a document, which is converted into list of sentences. We used an in-house rule-based sentence detector that splits the input text into logical sentences. After detecting sentences, the text is classed for numbers, measurements (cubic, tablet, capsule, cm, lb etc), dates and temporal values (hours, seconds, etc). Thus, all occurrences of numbers are classified into a single class NUM, all dates into DATE and so on. After this, for training, each token in a sentence is assigned its correct label $Y_i \in f, m, mo, du, r, do, NA$ where f is frequency, m is medication, mo is mode, du is duration, r is reason and do is dosage. Each training document is thus a long list of (X_i, Y_i) pairs. In addition, we also mark the beginning and end of each medication entry (i.e. to mark the correspondence between a medication with all its additional information like frequency, mode, duration etc.).

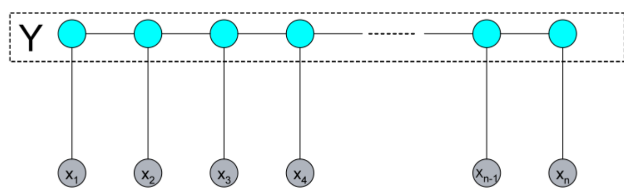

Figure 1. Graphical Representation of a CRF

4. Features and Training

This module takes a sentence detected by preprocessing step and extracts some intermediate features. These intermediate features are extracted at a token level. The sentence is broken into tokens and for each token, features are extracted. The features range from n-gram features to NLP feature, where the phrase type (Noun phrase, Verb phrase etc) for each token is extracted. We used the phrase parser from CTAKES (Savova et al., 2008), a tool available from Mayo Clinic to generate this feature. In addition we also used a state level feature whether a token is in a medication list or not. The list used for this purpose is RxNorm (Liu et al., 2005). Note that this is not a label, rather a feature. The reason is, for example, in context insulin dependent diabetes, the token insulin, even though a valid entry in RxNorm is not being used in context of prescribed medication. Thus, we only use this information as a feature, rather than a hard label.

CRF is a model used for segmentation/labeling sequential data and the features are presented in form of feature functions. It differs from a simple HMM as the features can not only be defined on observations, rather jointly on the observations and outputs (labels). They are also known as state and transition features respectively. The basic assumption in learning CRF parameters is that some variables X_i are always observed, whereas other variables Y_i are to be inferred. This makes it possible to maximize the conditional probability $p(Y_i|X_i)$ in contrast to maximizing the joint probability $p(Y_i, X_i)$. Since this is trained discriminatively, it is not necessary to model the distribution over always observed variables, which makes it possible to include arbitrarily complicated features of the observed variables into the mode e.g. n-grams on both observations (text) and phrasal values of output tags (labels). Each n-gram template automatically generates a set of feature functions. For the rest of the paper, label will mean $Y_i \in f, m, mo, du, r, do, NA$ and a token will mean a unit in the clinical text separated by blank spaces on both sides (e.g. a word).

The unigram, i.e. per single token, features in our model depend on

a. current token and its current label (e.g. token can be aspirin and its label medication)

b. current token, features of the current token and its current label (feature can be present in a medication list)

c. previous token, features of current token and the current label and so on

In bigram, i.e. taking pair-wise relationship between tokens or labels, the feature depends on the current label and previous label. Such a feature would help because in case the label of the previous token is a medication, the next token has a higher probability of being a dose if its feature value is a NUM. In addition, the longer n-grams also enable us to find the relation between a the main label (m) and its related labels (f, mo, du, r, do) i.e. for a found medication, its frequency, mode etc. This template automatically generates a set of feature function depending on the total number of output tokens.

As the result of the training, the tool outputs a model file. For testing, the files are processed similarly and using the model file and the features extracted from the test files, the tokens are labeled with the corresponding tags.

5. Pose-Processing

The need for post-processing is to convert the output of CRF model to i2b2 format so that we could report results in the same format. One should note that this post processing step is not necessary for the information extraction system as such. CRF outputs a label for each token of the test document. To use the standard evaluation tools provided by i2b2, we performed this post-processing step to conform to i2b2 format, where each entry is a medication name its offset, dosage and its offset, duration and its offset, reason and its offset, frequency and its offset, mode of administration and its offset, and marker which specifies if the medication is found in narrative of the text or in the list. For each medication in the CRF output, its corresponding mode, duration, frequency, dosage and reason need to be found in the text. This task becomes more difficult, since one single medication can have multiple entries in the documents. A medication having multiple entries means that, it has multiple references for dosages, reasons or multiple other fields. Even though this can be achieved using the same models by introducing additional variables, due to time constraints, we have not accounted for referential resolution in our model at this time.

6. Results

The dataset consists of 251 gold standard documents with ground truth provided by the i2b2 challenge organizers. Each document has an average of 800 tokens. We split the data into training and test files. We report results on a model trained on 188 files and tested on 63. Table 1 describes the results as acquired from the evaluation tool made available by the organizers. We have not used any other training data and no domain knowledge or experts were used. Based on the results, we are at par with the top entry that was submitted to the I2B2 challenge. Unlike the winning system, we have not used of any clinical experts and additional annotations.

7. Analysis of Results

As noted from the results section, the reason and duration were the hardest piece of the puzzle. The system performed extremely well on list type medication entries e.g. Percocet 1-2 tablets p.o. q 4 prn, Colace 100 mg p.o.75 b.i.d Our system was also able to extract medications and their dosages, modes and frequencies

Table 1. Comparison of F-measures for various entities. Bold is the better one

Tag	Winner	Proposed
Overall	0.84	0.84
do	**0.89**	0.87
f	0.89	0.89
mo	0.88	**0.90**
m	0.85	**0.86**
r	**0.44**	0.38
du	0.44	0.44

from the narrative text very well e.g. she was started on a course of Coumadin to reserve patency of 68 her graft. The challenging pieces of reason and duration were cases of reference resolution e.g. One of the discharge summaries had the following information.

PRINCIPAL DISCHARGE DIAGNOSIS Esophagitis OTHER DIAGNOSIS; NIDDM, MDD, SSI, Infections Med: continue remeron.

In this case the remeron (a tetracyclic antidepressant) has been continued due to MDD (major depressive disorder). Thus the reason would be MDD. However, the system yielded an nm (not mentioned). These kinds of entries prove to be daunting challenges that the system was unable to handle.

8. Conclusion

We achieved highly encouraging results comparable to top performing system. The overall effort in the development was only 2 person weeks. We were not able to account for the co-reference resolution in our model; however, with a little more effort this can be achieved using the same model as well. We plan to perform this as a future extension of our work. In addition, since the I2B2 organizers released the same data annotated in a community based effort, we plan to investigate using the multiple annotation for each record. This is an interesting area of machine learning called the phenomenon of learning from crowds where noisy but multiple labels are available and the model has to account during training for the best ones.

References

Chapman, W., Bridewell, W., Hanbury, P., Cooper, G., and Buchanan, B. G. Evaluation of negation phrases in narrative clinical reports. In *Proceedings of the American Medical Informatics Association (AMIA) Symposium*, pp. 105–109, 2001.

Chhieng, D., Day, T., and G. Gordon, et al. Use of natural language programming to extract medication from unstructured electronic medical records. In *AMIA Annu Symp*, pp. 907, 2007.

Friedman, C. and Hripcsak, G. Natural language processing and its future in medicine: Can computers make sense out of natural language text. *Academic Medicine*, 74(8):890–895, August 1999.

Gardner, R. M., Golubjatnikov, O. K., Laub, R. M., Jacobson, J. T., and Evans, R. S. Computer-critiqued blood ordering using the help system. *Comput Biomed Res*, 23:514–28, 1990.

Hirschman, L., Park, J. C., Tsujii, J., Wong, L., and Wu, C. H. Accomplishments and challenges in literature data mining for biology. *Bioinformatics*, 18 (12):1553–1561, December 2002.

I2B2. Medication extraction challenge, third i2b2 shared-task and workshop challenges in natural language processing for clinical data. URL https://www.i2b2.org/NLP/Medication.

Jagannathan, V., Mullett, C. J., and J. G. Arbogast, et al. Assessment of commercial nlp engines for medication information extraction from dictated clinical notes. In *Int J Med Inform*, volume 78, pp. 284–91, 2008.

Lafferty, J., McCallum, A., and Pereira, F. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the 18th International Conference on Machine Learning (ICML 2001)*, pp. 282–289, San Francisco, CA, 2001. Morgan Kaufmann.

Leapfrog. Leapfrog hospital survey results, 2008. URL http://www.leapfroggroup.org/media/file/leapfrogreportfinal.pdf.

Levin, M. A., Krol, M., and A. M. Doshi, et al. Extraction and mapping of drug names from free text to a standardized nomenclature. In *AMIA Annu Symp*, pp. 438–42, 2007.

Li, Z., Cao, Y., Zhang, Q., Agarwal, S., Ramesh, B. P., Kim, D., and Yu, H. Extracting medication information from patient discharge summaries. In *i2b2 NLP Workshop, AMIA*, 2009.

Liu, S., Ma, W., Moore, R., Ganesan, V., and Nelson, S. Rxnorm: Prescription for electronic drug information exchange. *IT Professional*, 7(5):17–23, 2005.

Patrick, J. and Li, M. A cascade approach to extract medication event. In *i2b2 NLP Workshop, AMIA*, 2009.

Roberts, A., Gaizauskas, R., and Hepple, M. Extracting clinical relationships from patient narratives. In *BioNLP 2008: Current Trends in Biomedical Natural Language Processing*, pp. 10–18, Columbus, Ohio, USA, June 2008.

Savova, G., Kipper-Schuler, K., Buntrock, J., and Chute, C. Uima-based clinical information extraction system. In *Language Resources and Evaluation: LREC: Towards enhanced interoperability for large HLT systems: UIMA for NLP*, Marrakech, Morocco, 2008.

Soderland, S., Aronow, D., Fisher, D., Aseltine, J., and Lehnert, W. Machine learning of text analysis rules for clinical records. Technical report, CIIR Technical Report, University of Massachusetts, Amherst, 1995.

Uzuner, O., Goldstein, I., Luo, Y., and Kohane, I. Identifying patient smoking status from medical discharge records. *J Am Med Inform Assoc*, 15(1):15–24, 2008.

Xu, H., Stenner, S. P., Doan, S., Johnson, K. B., Waitman, L. R., and Denny, J. C. Medex: a medication information extraction system for clinical narratives. *JAMIA*, 17:19–24, 2010.

Zhou, X., Han, H., Chankai, I., Prestrud, A., and Brooks, A. Approaches to text mining for clinical medical records. In *Proceedings of the 2006 ACM Symposium on Applied Computing*, pp. 235–239, 2006.