Mody University International Journal of Computing and Engineering Research
2 (2), 2018, 61-63
http://www.publishingindia.com/micer

Medical Natural Language Systems: A Review

A. Senthil[1] and Anjali Kedawat[2*]

[1]Associate Professor and Head, Dept. of Computer Science, Mody University of Science and Technology, Laxmangarh, Sikar, Rajasthan, India. Email: asenthil.cet@modyuniversity.ac.in
[2]Research Scholar, SET, Mody University of Science and Technology, Laxmangarh, Sikar, Rajasthan, India. Email: anj_ked@yahoo.co.in
*Corresponding Author

Abstract: Medication information is one of the free text clinical data in medical records. It is difficult to access medical records due to healthcare safety and patient information security. Clinical narratives are differing due to multilingualism, clinical report formats. Clinical information can be extracted with Natural Language Processing System based on medical domain. This paper contains a short review on NLP systems used for medical domain. Medical natural language systems are different due to their use of different applications in medical domain.

Keywords: Clinical report, Natural language processing.

I. Introduction

Natural language processing has an important role for extract medical information. Clinical data available in free text format which is not useful. Medical NLP systems are useful to convert unstructured data in to natural language form which can be readable and help for improvement of electronic medical records. These systems are capable of reading various radiology reports and convert it in structured format. MNLP (Medical NLP) are differ in terms of technology used, size, modularity, accuracy, efficiency, network used, framework, etc.

II. Review

Friedman *et al.* [1] build MedLEE (Medical Language Extraction and Encoding) NLP system which is used for decision support. This system initially applied to radiology reports of the chest X -ray report. The first extension of MedLEE is mammography report due to language similarity of chest X-ray report. Second extension of MedLEE is discharge summaries. This system written in PROLOG which support Windows, UNIX and AIX platform. MedLEE have four components: pre-processor which reads original clinical text, parser which is used to determine structure of sentence, phrase regularization and encoder.

Koehler and B. Spencer [2] develop an SymText NLP system which combine syntax and semantic approach for encoding of free text. It uses Bayesian n/w and used for chest x-ray report. This system achieves 86.1% recall and 85.6% precision. It contains a parser which have augmented transition network grammar and transformed grammar.

Lee M. Christensen *et al.* [3] develop MPLUS (A Probabilistic Medical Language Understanding System) NLP system which is an analysis tool for medical text and based on Bayesian n/w. This system defines relations with in a medical domain. Bayesian n/w is same as semantic n/w in respect of structure. MPLUS is differ from SymText in terms of size and modularity. It is based on bottom up Chart parser and coded in LISP with some C routines for BN's access.

Brian Hazlehurst *et al.* [4] describe MediClass NLP system which process free text and coded data for automatic detection of clinical events in EMR (Electronic Medical Records). This system is used for clinical operations and healthcare services. It is an open source system which use informatic techniques. It is also used for research projects based on NLP techniques for automatic indexing and coding of clinical contents. It uses both NLP techniques and Knowledge based techniques. This system applied on patient safety, disease surveillance and care quality.

S. Goryachev *et al.* [5] develop HITEx (Health Information Text Extraction) open source NLP system and based on data driven approach. It is used for medical and non-medical domain. It builds from General Architecture for Text Engineering framework (GATE). It accesses clinical information like diagnosis, discharge medication, smoking status, etc.

Guergana K Savova *et al.* [6] build cTAKES (clinical Text Analysis and Knowledge Extraction System) open source NLP system for extract information from EMR. It is built from open source technique (UIMA) and open NLP process toolkit. The goal of developing this system is to develop heterogeneous clinical domain system. cTAKES include sentence boundary, detector, tokenizer, normalizer, POS, shallow parser, named entity recognization. It integrates both rule based and machine learning techniques to annote negation context, named entity and syntactic context. It accepts both plain text and clinical documents.

Alan R. Aronson and François-Michel Lang [7] gives MetaMap NLP system which is used for biomedical text retrieval. The limitation of this system is that it only apply on English text and not use for real time. It reduces accuracy in presence of ambiguity. It also used for text mining, classification, Knowledge discovery etc.

Hua Xu *et al.* [8] build MedEx (Medical Information Extraction System) NLP system which is used for discharge summary. There searcher collects 50 discharge summary with f- measure of 93.2% and 95% over 25 visit notes. It is based on semantic approach. The researcher achieves f-measure of 90% by integrating semantic tagger and chart parser. It is based on algorithm which is used to find medical name and signature. The researcher used new method of parsing by integrate sequential tagger and combined tagger. Sequential tagger is used for combination of regular expression, look up and rule based disambiguation.

Vijay Garla *et al.* [9] develop YTEX (Yale cTAKES) NLP system which is an extension of cTAKES NLP system. It uses regular expression based named entity recognition. The researcher randomly selects 236 patient forms. It simplifies feature extraction and develop machine learning, rule based classifiers.

Yu-Kai Lin *et al.* [10] build MedTime NLP system which is based on rule based and machine learning approach. It is a temporal information extraction system which is an important area for biomedical and health information research.

Sunghwan Sohn *et al.* [11] develop MedXN (Medical Extraction and Normalization) NLP system which is used to extract and normalize medical information by using RxNorm. It is based on UIMA framework. The researcher collects 397 medication and achieve f-score of 97.8%. The researcher used 59 clinical notes from Mayo Clinic and contained 659 manually annotated clinical mentions. It uses decomposition and composition strategy for extraction of clinical information after this MEDXN convert this information in RxCUI.

Foster R. Goss *et al.* [12] build MTERMS (Medical Text Extraction Reasoning Mapping System) NLP system to extract and encode allergy information. It converts free text into XML output by using pipeline approach. It has f-measure of 87.6%, recall 91% and precision 84.45%. The lexicon used in this system have multiple standard technologies (RxNorm, UNII, SNOMED CT).

III. Comparison

Comparison of medical tools in terms of applications and programming language are shown in Table I "Clinical NLP Systems".

Table I: Clinical NLP Systems

Tools	Programming Language Used	Description
MedLEE	Prolog	Used for chest X-ray reports, mammography reports, discharge summaries. Encode and extract clinical information from free text patient reports.
MPLUS/SymText	LISP,C++	Based on Radiology reports, define relations within a medical domain.
MediClass	Java	Process free text and coded data for automatic detection of clinical events in EMR (Electronic Medical Records).
HITEx	Java	Smoking status, negation, temporal concepts, co-morbidity.
cTAKES	Java	Develop heterogeneous clinical domain system, reterive clinical information from radiology reports.
MetaMap	Perl,C,Java,Prolog	Extracting Medical problems.
MedEx	Java,	Used for discharge summary.
YTEX	Java, Matlab	Simplify feature extraction and develop machine learning, rule based classifiers.
MedTime	Java	Temporal information extraction system.
MedXN	Java	Used to extract and normalize medical information by using RxNorm.
MTERMS	XML	Extract and encode allergy information.

IV. Conclusion

This paper contains the review of number of publications related with natural language system used for clinical domain. The review includes programming languages in which these systems written, clinical reports on which these system belongs and framework. From this paper we conclude about the Medical tools used for various clinical domain and specific feature of each medical tool.

References

[1] C. Friedman, "Towards a comprehensive medical language processing system: Methods and issues," *Proceedings of the AMIA Annual Fall Symposium: American Medical Informatics Association*, pp. 595-599, 1997.

[2] Koehler, and B. Spencer, "SymText: A natural language understanding system for encoding free text medical data," Doctor Dissertation, University of Utah, 1998.

[3] L. M. Christensen, P. J. Haug, and M. Fiszman, "MPLUS: A probabilistic medical language understanding system," *Proceedings of the Workshop on Natural Language Processing in the Biomedical Domain: Association for Computational Linguistics*, pp. 29-36, 2002.

[4] B. Hazlehurst, H. R. Frost, D. F. Sittig, and V. J. Stevens, "MediClass: A system for detecting and classifying encounter-based clinical events in any electronic medical record," *Journal of the American Medical Informatics Association*, vol. 12, no. 5, pp. 517-529, 2005.

[5] S. Goryachev, M. Sordo, and Q. T. Zeng, "A suite of natural language processing tools developed for the I2B2 project," *AMIA Annual Symposium Proceedings*, p. 931, 2006.

[6] G. K. Savova, J. J. Masanz, P. V. Ogren, J. Zheng, S. Sohn, K. C. Kipper-Schuler, and C. G. Chute, "Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): Architecture, component evaluation and applications," *Journal of the American Medical Informatics Association*, vol. 17, no. 5, pp. 507-513, 2010.

[7] A. R. Aronson, and F.-M. Lang, "An overview of MetaMap: Historical perspective and recent advances," *Journal of the American Medical Informatics Association*, vol. 17, no. 3, pp. 229-236, 2010.

[8] H. Xu, S. P. Stenner, S. Doan, K. B. Johnson, L. R. Waitman, and J. C. Denny, "MedEx: A medication information extraction system for clinical narratives," *Journal of the American Medical Informatics Association*, vol. 17, no. 1, pp. 19-24, 2010.

[9] V. Garla, V. L. Re, Z. Dorey-Stein, F. Kidwai, M. Scotch, J. Womack, A. Justice, and C. Brandt, "The Yale cTAKES extensions for document classification: Architecture and application," *Journal of the American Medical Informatics Association*, vol. 18, no. 5, pp. 614-620, 2011.

[10] Y. K. Lin, H. Chen, and R. A. Brown, "MedTime: A temporal information extraction system for clinical narratives," *Journal of Biomedical Informatics*, pp. 20-28, 2013.

[11] S. Sohn, C. Clark, S. R. Halgrim, S. P. Murphy, C. G. Chute, and H. Liu, "MedXN: An open source medication extraction and normalization tool for clinical text," *Journal of the American Medical Informatics Association*, vol. 21, no. 5, pp. 858-865, 2014.

[12] F. R. Goss, J. M. Plasek, J. J. Lau, D. L. Seger, F. Y. Chang, and L. Zhou, "An evaluation of a natural language processing tool for identifying and encoding allergy information in emergency department clinical notes," *AMIA Annual Symposium Proceedings*, pp. 580-588, 2014.