Indonesian Journal of Electrical Engineering and Computer Science
Vol. 18, No. 2, May 2020, pp. 946~952
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v18.i2.pp946-952

A framework for named entity recognition of clinical data

Ravikumar J[1], Ramakanth Kumar P[2]
[1]Department of CSE, Dr. Ambedkar Institute of Technology, India
[2]Department of CSE, R.V. College of Engineering, India

Article Info

Article history:

Received Aug 30, 2019
Revised Oct 20, 2019
Accepted Nov 6, 2019

Keywords:

Information extraction
Named entity recognition
Surgical data

ABSTRACT

With emergence of technologies like big data, the healthcare services are also being explored to apply this technology and reap benefits. Big Data analytics can be implemented as a part of e-health which involves the extrapolation of actionable insights from sources like health knowledge base and health information systems. Present day medical data creates a lot of information consistently. At present, Hospital Information System is a quickly developing innovation. This data is a major asset for getting data from gathering of gigantic measures of surgical information by forcing a few questions and watchwords. Be that as it may, there is issue of getting data precisely what the client need, because Hospital Information System contains more than one archive identified with a specific thing, individual or episode and so on. Information extraction is one of information mining systems used to concentrate models portraying essential information classes. The proposed work will work for the most part concentrating on accomplishing great execution in Medical Domain. Fundamentally this had two primary purposes one was separating significant information from patient content record and second one labelling name substance, for example, individual, association, area, malady name and symptoms. Improve survival rates and tweak care conventions and review inquiries to better deal with any interminable consideration populace. Lower costs by decreasing pointless hospitalizations. Abbreviate length of stay when confirmation is fundamental.

Corresponding Author:

Ravikumar J,
Department of Computer Science and Engineering,
Dr. Ambedkar Institute of Technology, Bengaluru, India.
Email: ravij041@dr-ait.org

1. INTRODUCTION

To understand relevant parts present in the text and to gather information from several pieces of text, Information Extraction (IE) systems are used. Using IE systems, the relevant information can be produced and presented in a structured format like relations (in the database sense). This is also known as knowledge base. The primary goal of the IE systems is organizing the information in a way that is useful to people. These systems also put information in a semantically acceptable form so that further inferences can be made by computer algorithms. Now a days, enormous amount of clinical data is being generated by health organizations. Hence, extraction of information from these clinical notes will enable the improvement of clinical wards works. It also helps to get good understanding of patient care and also the disease progression. The key point in unlocking the information present in the clinical text is recognizing named entities. In the information extraction process, an important sub task is named entity recognition. The process involves named entity phrases recognition and classifying them into particular categories. In the medical domain, the important categories are procedures, clinical findings and drugs. In the literature, named entity recognition is one of the widely studied area [1-4].

The clinicians written notes are less structured and are usually in minimal grammatical form with cryptic short hand. This poses challenges in named entity recognition. Principally, NER systems are either pattern based or rule based. Because of the diverse writing style of individual clinicians, the patterns and rules are not generalizable. Technologies like machine learning are also not fully advanced in NER because of lack of available training data. NERs are categorized into three approaches. Rule-based approaches, machine learning based approaches and dictionary based approaches. The machine learning based systems focus on choosing effective features for building classifiers. For experimentation, several machine learners have been used. Support Vector Machines (SVMs) [5] and Hidden Markov Model (HMM) [6] are examples of such learners. Natural language processing focuses on development of models to understand natural language [7]. The NLP framework includes modules for syntactic processing like tokenization, detection of sentences and parts-of-speech tagging. The NLP system also include modules for semantic processing like named entity recognition tagging, identification of concept, extraction of relation and anaphoric resolution. An Information Extraction system is an NLP system that has semantic processing models to extract information that are pre-defined. In the medical domain, NLP systems are used by researchers to identify clinical syndromes and biomedical concepts from the reports of radiology [8] and discharge summaries [9]. NER process includes tasks such as finding, storing and sorting content into categories such as the person's names, locations, organizations, expression of times, monetary values, quantities and percentages. NER system extracts data directly from sentences of plain English. NER is also known as identification of entity, entity chunking and extraction of entity. It is an intelligence system that is state-of-the-art that works equivalent to human brain in terms of efficiency. From raw data, NER system finds entity elements and determines the category to which the element belongs. It reads and highlights the important entity elements in the text. Depending on the project, NER might be given separate sensitive entities. The system used in one project may not be suitable for another project. The system also faces numerous challenges such as correct information extraction for specific but closely related categories. The process of NER happens in several steps. First, the Knowledge base is to be built which consists of known Named Entities. Then, linking of entity to a knowledge base should happen. This process consists of components such as extractors, searchers and disambiguators. Extraction process involves identifying and preparing named entity mentions. The tasks such as parts of speech tagging, tokenization, detection of sentence boundary, capitalization rules and in-document co-reference. In-document co-reference is used to find more specific search items. In the Search phase, titles, disambiguous pages can be leveraged to capture synonyms. In order to reduce the computation, a searcher should balance precision and recall for capturing of correct entity. In this paper we present a named entity recognizer using a classifiers to find entities.

2. RELATED WORK

Named entity extraction is a type of information retrieval which focuses on identifying instances i.e., names of various types of entities. For example, cancer would be an instance of disease; swelling would be an instance of symptoms and so on. One of the earliest NER models was based on decision tree [10]. Sekine developed a system was developed for Japanese. The author used features viz. POS (part-of-speech) tags extracted by a morphological analyzer, information based on character and specialized dictionary. The researcher presented the algorithm which included two phases one for decision tree creation from training data and the other for generating the tagged output that is based on the decision tree.

Another early work was done by Bikel, Schwartz and Weischedel [11]. Authors used Hidden Markov Model (HMM) to identify named entity. Primary features like bi-gram and orthographic features like word case, word shape etc. were used. The authors evaluated the model in English, Spanish and on speech input. To quantify the performance on data available to the community (MUC-6 and MET-1), results are reported on standard materials only. The results obtained have been found better consistently than any other learning algorithm. Borthwick [12] in his PhD thesis used maximum entropy algorithm. This thesis explains a statistical named-entity recognition system known as MENE (Maximum Entropy Named Entity).It utilizes a very flexible object-based architecture which allows it to make use of a broad range of knowledge sources in making its tagging decisions. McCallum and Li [13] developed Conditional Random Fields based algorithm to extract NER in coNLL-2003 shared task competition. The work described WebListing which is a method to obtain seeds for the lexicons from the labeled data. It then uses Web and HTML formatting regularities, service of search engine for augmentation of those lexicons.

Sarawagi and Cohen [14] propose a semi Markov CRF (Conditional Random Field) algorithm for named entity extraction. Semi-CRFs offer much of the power of higher-order models. The major advantage is that it allows features which measures properties of segments than individual elements. These features can be quite natural for applications like NER. The researchers extended their work with the use of dictionary and notion of similarity function [15]. Naidu and Sekine [16] provide wide overall survey of NER research. The researchers presented a survey of 15 years of research in the NERC field from 1991 to 2006. Handcrafted rule-based algorithms were used by early systems. Machine learning techniques are being used in the recent systems.

The survey of the techniques was conducted as well as other critical aspects of NERC such as features and evaluation methods. Aronson [17] developed MetaMap to map bio-medical concepts from Unified Medical Language System (UMLS). MetaMap is a program developed at the NLM (National Library of Medicine) that can be used to map biomedical text to the Metathesaurus. MetaMap uses a knowledge intensive approach based on symbolic, natural language processing (NLP) and computational linguistic techniques.

The researchers developed MetaMap based NLP system in [18] that extracts various entities like temporal information, corresponding codes from clinical notes by matching with Unified Medical Language System (UMLS). In this method, matching of MedLEE generated structured output that consists of findings and modifiers to get most specific code. Recall and precision applied to Unified Medical Language System (UMLS) coding were evaluated. The results were compared with reference standard determined in the manual method by seven experts.

Minard et al. [19] proved that the developed hybrid approach that is based on both domain knowledge and machine learning gives better performance. The work is compared with multiple approaches based on domain-knowledge and machine-learning techniques to Medical Entity Recognition.The approaches rely on machine learning and rule-based methods. To extract features from the input texts, NLP is used. Later they are fed to machine learning methods. For concept extraction, the researcher used Conditional Random Fields, and Support Vector Machines for assertion and relation annotation. The authors tested various combinations of rule-based and machine-learning methods depending on the task.

Li, Schuler and Savova [20] have used both CRF and SVM based for model extraction of disorder in clinical text. The authors presented a comparative analysis between support vector machines and Conditional Random Fields for clinical named entity recognition. The authors also explored the applicability of these methods to clinical domain. The outcome of the comparison indicates that CRFs perform better than SVMs when evaluated against a set of gold standard named entities. The best F-score with CRFs is 0.86 and for the SVMs is 0.64 as compared to a baseline of 0.60. Various techniques have been proposed for named entity recognition [21-25]. But, there exist scope for improvement.

3. DATA COLLECTION

The data set used for training the model is given in the Table 1.

Table 1. Data Set Details

Data set	Purpose	Notes	Entities	Entity types
i2b2 2010	Training	349	27,837	Problem, Treatment
	Test	477	45,009	Test

4. THE PROPOSED MODEL

Patients health status, tests conducted, diseases and response to the treatments are stored in clinical records. Analysis of such information provides immense value for clinical practice, organization and management of healthcare services. Concept extraction (CE) activity aims to recognize mentions to medical concepts like problems, diagnosis data (tests) and treatments mentioned in the clinical records. (e.g., progress reports and discharge summaries). Further these identified concepts are classified into predefined categories. The concept in clinical data is usually mentioned in text format. Hence, it is a challenging task for Natural Language Processing systems to extract these concepts automatically.

In the proposed model, machine learning model has been used to recognize and extract concepts from clinical data. This work employs a new approach known as bidirectional Long Short Tem Memory – Conditional Random Field (LSTM-CRF) initialized with general-purpose, off-the-shelf word embeddings. Figure 1 depicts the training for named entity detection. Figure 2 indicates the training for named entity boundary determination.

He was admitted to ICU for meningitis.

Positive example
Can only mean disorder

He was continued on Acyclovir

Negative example
Cannot mean disorder

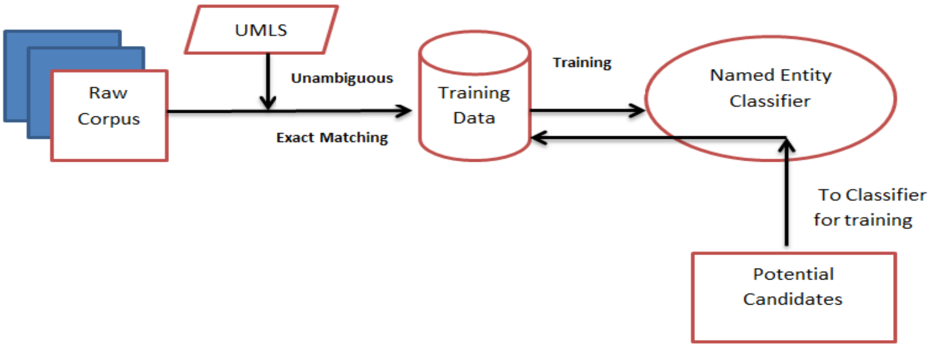


Figure 1. Training for named entity detection

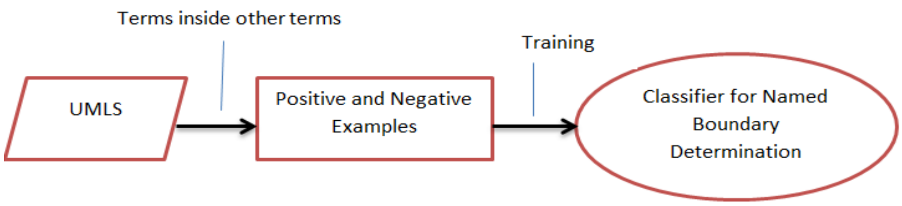


Figure 2. Training for named entity boundary determination

The input sequence is x=(x1,x2,x3,…xm), i.e., the words of a sentence and a sequence of output states S=(s1,s2,s3,…sm), i.e. the named entity tags. In conditional random fields the conditional probability P(s1,s2,s3,…sm| x1,x2,x3,…xm) is modeled as the output state sequence. It has been done by defining a feature map.

$$\Phi(X1,\ldots,Xm,S1,\ldots,Sm) \in Rd \tag{1}$$

that maps an entire input sequence paired with an entire state sequence to some dimensional feature vector. The probability as a log-linear model with the parameter vector has been modeled as

$$\omega \varepsilon R^{d} \tag{2}$$

$$P(s|x;w) = \frac{\exp(\omega.\Phi(x,s))}{\sum sf \exp(\omega.\Phi(x,sf))} \tag{3}$$

where s’ ranges over all possible output sequences. The expression w . ϕ(x,s) = score $_{crf}$ (x,s) can be viewed as a scoring how well the state sequence fits the given input sequence. The idea is now, to replace the linear scoring function by a non-linear neural network. Hence score can be defined as,

$$\text{score}_{\text{lstm-crf}}(x,s) = \sum_{i=0}^{n} W_{s_{i-1},s_i}.\text{LSTM}(x)_i + b_{s_{i-1},s_i} \tag{4}$$

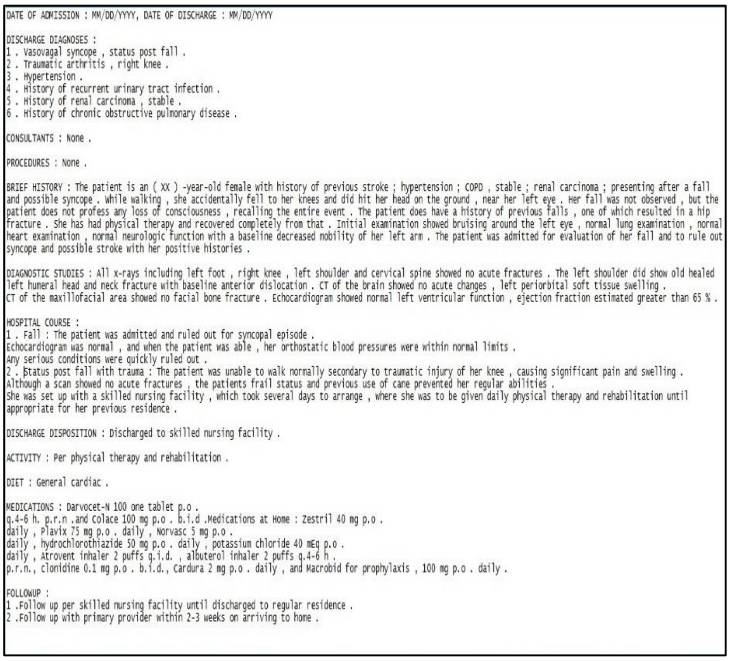
where Ws_{i-1}, s_i and b are the weight vector and the bias corresponding to the transition from s_{i-1} to s_{i1}espectively. The score functions are also called *potential functions.* After constructing this score function, the conditional probability p(S|x :W,b) can be optimized as in the usual CRF and propagating back through the network.

In the input sentence, each word is first mapped to random vector or a vector from a word embedding. Word embeddings are vector representation of words of natural language that preserve the syntactic and semantic similarities between them. The vector representations are generated by count-based approaches such as or trained models. In its embedded representation, each word in a text is represented by a real-valued vector, x, of arbitrary dimensionality, d. Figure 3 indicates the algorithm used for named entity recognition.

To conduct experiments, more than some concept-annotated reports are taken for training; testing and results are examined after execution through proposed methodology. Figure 4 and Figure 6 indicate the example for clinical records and Figure 5 and Figure 7 are present the output of the proposed method.

Step 1	**Start** Input: Clinical records having patients health status, tests conducted, diseases and response to the treatments. The input sequence is x=(x1,x2,x3,…xm).
Step 2	**Development of Classification Model** Identified concepts like problems, diagnosis data (tests) and treatments mentioned in the clinical records are classified into predefined categories.
Step 3	**Feed the training data to model** The records are divided into training data and testing data. 70% of data is used as training data and it is fed to the model.
Step 4	**Feed the test data to model** 30 % of the data is used as testing data. The data consisting of patients information are fed to the model to test the accuracy of the model
Step 5	**Feed the clinical records to model** Once the model is built, the real data (clinical records) are fed to the pre developed model. This generates output
Step 6	**Obtain output** The output is the document that include list of words that indicate problem diagnosed, test conducted or treatment given
Step 7	**End**

Figure 3. Algorithm steps


DATE OF ADMISSION : MM/DD/YYYY, DATE OF DISCHARGE : MM/DD/YYYY

DISCHARGE DIAGNOSES :
1 . Vasovagal syncope , status post fall .
2 . Traumatic arthritis , right knee .
3 . Hypertension .
4 . History of recurrent urinary tract infection .
5 . History of renal carcinoma , stable .
6 . History of chronic obstructive pulmonary disease .

CONSULTANTS : None .

PROCEDURES : None .

BRIEF HISTORY : The patient is an (XX) -year-old female with history of previous stroke ; hypertension ; COPD , stable ; renal carcinoma ; presenting after a fall and possible syncope . While walking , she accidentally fell to her knees and did hit her head on the ground , near her left eye . Her fall was not observed , but the patient does not profess any loss of consciousness , recalling the entire event . The patient does have a history of previous falls , one of which resulted in a hip fracture . She has had physical therapy and recovered completely from that . Initial examination showed bruising around the left eye , normal lung examination , normal heart examination , normal neurologic function with a baseline decreased mobility of her left arm . The patient was admitted for evaluation of her fall and to rule out syncope and possible stroke with her positive histories .

DIAGNOSTIC STUDIES : All x-rays including left foot , right knee , left shoulder and cervical spine showed no acute fractures . The left shoulder did show old healed left humeral head and neck fracture with baseline anterior dislocation . CT of the brain showed no acute changes , left periorbital soft tissue swelling .
CT of the maxillofacial area showed no facial bone fracture . Echocardiogram showed normal left ventricular function , ejection fraction estimated greater than 65 % .

HOSPITAL COURSE :
1 . Fall : The patient was admitted and ruled out for syncopal episode .
Echocardiogram was normal , and when the patient was able , her orthostatic blood pressures were within normal limits .
Any serious conditions were quickly ruled out .
2 . Status post fall with trauma : The patient was unable to walk normally secondary to traumatic injury of her knee , causing significant pain and swelling .
Although a scan showed no acute fractures , the patients frail status and previous use of cane prevented her regular abilities .
She was set up with a skilled nursing facility , which took several days to arrange , where she was to be given daily physical therapy and rehabilitation until appropriate for her previous residence .

DISCHARGE DISPOSITION : Discharged to skilled nursing facility .

ACTIVITY : Per physical therapy and rehabilitation .

DIET : General cardiac .

MEDICATIONS : Darvocet-N 100 one tablet p.o .
q.4-6 h. p.r.n .and Colace 100 mg p.o . b.i.d .Medications at Home : Zestril 40 mg p.o .
daily , Plavix 75 mg p.o . daily , Norvasc 5 mg p.o .
daily , hydrochlorothiazide 50 mg p.o . daily , potassium chloride 40 mEq p.o .
daily , Atrovent inhaler 2 puffs q.i.d. , albuterol inhaler 2 puffs q.4-6 h .
p.r.n., clonidine 0.1 mg p.o . b.i.d., Cardura 2 mg p.o . daily , and Macrobid for prophylaxis , 100 mg p.o . daily .

FOLLOWUP :
1 .Follow up per skilled nursing facility until discharged to regular residence .
2 .Follow up with primary provider within 2-3 weeks on arriving to home .

Figure 4. Clinical record 1

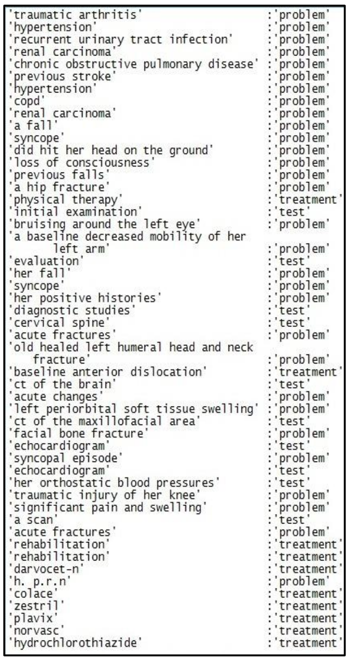

```
'traumatic arthritis'                        :'problem'
'hypertension'                               :'problem'
'recurrent urinary tract infection'          :'problem'
'renal carcinoma'                            :'problem'
'chronic obstructive pulmonary disease'      :'problem'
'previous stroke'                            :'problem'
'hypertension'                               :'problem'
'copd'                                       :'problem'
'renal carcinoma'                            :'problem'
'a fall'                                     :'problem'
'syncope'                                    :'problem'
'did hit her head on the ground'             :'problem'
'loss of consciousness'                      :'problem'
'previous falls'                             :'problem'
'a hip fracture'                             :'problem'
'physical therapy'                           :'treatment'
'initial examination'                        :'test'
'bruising around the left eye'               :'problem'
'a baseline decreased mobility of her
      left arm'                              :'problem'
'evaluation'                                 :'test'
'her fall'                                   :'problem'
'syncope'                                    :'problem'
'her positive histories'                     :'problem'
'diagnostic studies'                         :'test'
'cervical spine'                             :'test'
'acute fractures'                            :'problem'
'old healed left humeral head and neck
   fracture'                                 :'problem'
'baseline anterior dislocation'              :'treatment'
'ct of the brain'                            :'test'
'acute changes'                              :'problem'
'left periorbital soft tissue swelling'      :'problem'
'ct of the maxillofacial area'               :'test'
'facial bone fracture'                       :'problem'
'echocardiogram'                             :'test'
'syncopal episode'                           :'problem'
'echocardiogram'                             :'test'
'her orthostatic blood pressures'            :'test'
'traumatic injury of her knee'               :'problem'
'significant pain and swelling'              :'problem'
'a scan'                                     :'test'
'acute fractures'                            :'problem'
'rehabilitation'                             :'treatment'
'rehabilitation'                             :'treatment'
'darvocet-n'                                 :'treatment'
'h. p.r.n'                                   :'problem'
'colace'                                     :'treatment'
'zestril'                                    :'treatment'
'plavix'                                     :'treatment'
'norvasc'                                    :'treatment'
'hydrochlorothiazide'                        :'treatment'
```

Figure 5. Output of the algorithm for clinical record 1

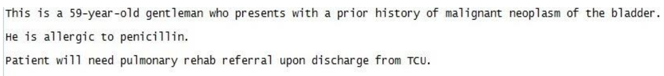

This is a 59-year-old gentleman who presents with a prior history of malignant neoplasm of the bladder.
He is allergic to penicillin.
Patient will need pulmonary rehab referral upon discharge from TCU.

Figure 6. Clinical record 2

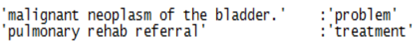

```
'malignant neoplasm of the bladder.'    :'problem'
'pulmonary rehab referral'              :'treatment'
```

Figure 7. Output of the algorithm for clinical record 2

5. RESULTS AND DISCUSSIONS

The performance comparison between the proposed method and existing state of the art CE systems indicate that the bidirectional LSTM-CRF initialized with GloVe outperforms all recent approaches. These results are not directly comparable since the some of the previous works have used the original dataset which has a significantly larger number of training samples. In general, the results achieved with the proposed approach are close and in many cases above the results achieved by systems based on hand-engineered features. Figure 8 indicates the prediction accuracy. Result comparison of proposed method with other state-of-the-art techniques shown in Table 2.

Table 2. Result Comparison of Proposed Method with Other State-of-The-Art Techniques

Method	F-score
Proposed Method	0.291
MetaMap[31]	0.113
ZE’13[13]	0.265

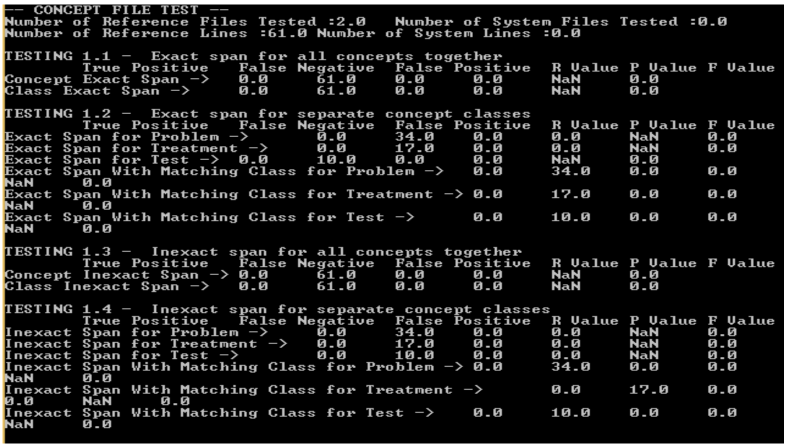


Figure 8. Prediction accuracy

6. CONCLUSION

In this work, a new approach has been proposed for extraction of named entities and classifying them. This is of immense use to practitioners as well as management of the hospitals. The key finding of this work is its ability to provide end-to-end recognition using general-purpose, off-the-shelf word embeddings. This avoids additional efforts from time-consuming feature construction. This work can be a good contribution to a research in the area of NER extraction in clinical data.

REFERENCES

[1] Z. Liu, X. Wang, Q. Chen and B. Tang, "*Chinese Clinical Entity Recognition via Attention-Based CNN-LSTM-CRF,*" IEEE International Conference on Healthcare Informatics Workshop (ICHI-W), New York, NY, 2018, pp. 68-69, 2018.

[2] Yefeng Wang, Jon Patrick, “*Cascading Classifiers for Named Entity Recognition in Clinical Notes*”, Workshop Biomedical Information Extraction-Borovets, Bulgaria, pp. 42-49, 2009.

[3] Zengjian Liu, Ming Yang, Xiaolong Wang, “Entity re cognition from clinical texts via recurrent neural network”, *BMC Medical Informatics and Decision Making,* 17(S2).

[4] Maximilian Hofer, Andrey Kormilitzin, Paul Goldberg, Alejo Nevado-Holgado, “Few-shot Learning for Named Entity Recognition in Medical Text”, arXiv:1811.05468,pp:1-10,NOV 2018.

[5] T. Joachims, C. Nedellec, and C. Rouveirol. “*Text categorization with support vector machines: learning with many relevant. In Machine Learning*”: ECML-98 10th European Conference on Machine Learning, Chemnitz, Germany Springer, pp. 137-142, 1998.

[6] L. Rabiner et al. “*A tutorial on hidden Markov models and selected applications in speech recognition*”.Proceedings of the IEEE, vol. 77(2), pp. 257-286, 1989.

[7] S.M. Meystre, G.K. Savova, K.C. Kipper-Schuler, J.F. Hurdle, “Extracting informa-tion from textual documents in the electronic health record: A Review Of Recentresearch”, *Yearb Med. Inform.* 35, pp. 128-144, 2008.

[8] R.W.V. Flynn, T.M. Macdonald, N. Schembri, G.D. Murray, A.S.F. Doney, “Automated data capture from free-text radiology reports to enhance accuracy of hospital inpatient stroke codes”, Pharmacoepidemiol, Drug Saf. 19.

[9] H. Yang, I. Spasic, J.A. Keane, G. Nenadic, “A text mining approach to the pre-diction of disease status from clinical discharge summaries”, *J. Am. Med. Inform. Assoc,* pp. 596-600, 2009.

[10] Sekine, S. 1998. “*Nyu: Description of the Japanese NE System Used For Met-2*”. In Proc. Message Understanding Conference, 1998.

[11] Bikel, D. M., Schwartz, R., & Weischedel, R. M, “An algorithm that learns what's in a name”. *Machine Learning*, vol.34, pp. 211-231, 1999.

[12] Borthwick, A. “A maximum entropy approach to named entity recognition”. *PhD diss.*, New York University, 1999.

[13] McCallum, A & Wei L. “*Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons*”. In Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4, pp. 188-191,2003.

[14] Sarawagi, S. & Cohen, W. W. "Semi-markov conditional random fields for information extraction". In *Advances in Neural Information Processing Systems*, pp. 1185-1192. 2004.
[15] Cohen, W. W., &Sarawagi, S. "*Exploiting dictionaries in named entity extraction: combining semimarkov extraction processes and data integration methods*". In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pp: 89-98.ACM, 2004.
[16] Nadeau, D. &Sekine, S. "A survey of named entity recognition and classification. Lingvisticae Investigationes" pp. 3-26, 2007.
[17] Aronson AR. "Effective mapping of biomedical text to the UMLS metathesaurus": *The Meta Mapprogram*, 2001.
[18] Friedman, C., Shagina, L., Lussier, Y., &Hripcsak, G. (2004). "Automated Encoding of Clinical Documents Based on Natural Language Processing". *Journal of the American Medical Informatics Association (JAMIA*), 11(5), pp. 392-402, 2004.
[19] Minard AL, Ligozat AL, Ben Abacha A, et al. "Hybrid methods for improving information access in clinical documents: concept, assertion, and relation identification". *J Am Med Inform Assoc*. 2011;pp. 588-93, 2011.
[20] Li, D., Kipper-Schuler, K., &Savova, G. "*Conditional random fields and support vector machines for disorder named entity recognition in clinical texts*". In Proceedings of the workshop on current trends in biomedical natural language processing, pp. 94-95. Association for Computational Linguistics, 2008.
[21] O. Ghiasvand and R. J. Kate, "*Biomedical Named Entity Recognition with less Supervision*". International Conference on Healthcare Informatics, Dallas, TX, pp. 495-495, 2015.
[22] H. T. Le, R. C. Sam, H. C. Nguyen and T. T. Nguyen, "*Named entity recognition in vietnamese text using label propagation*". International Conference on Soft Computing and Pattern Recognition (SoCPaR), Hanoi, pp. 366-370, 2013.
[23] S. S. Sonawane and P. A. Kulkarni, "*Entity based co-reference resolution with name entity recognition using hierarchical classification*". Annual IEEE India Conference (INDICON), New Delhi, pp. 1-6, 2015.
[24] S. P. Tripathi and H. Rai, "*SimNER-An Accurate and Faster Algorithm for Named Entity Recognition*", Second International Conference on Advances in Computing, Control and Communication Technology (IAC3T), Allahabad, India, pp. 115-119, 2018.
[25] F. Su and C. Xiao, "*A Complex Named Entities Extraction Model Oriented Web Video*," Second International Workshop on Education Technology and Computer Science, Wuhan, pp. 276-279, 2010.

BIOGRAPHIES OF AUTHORS

Mr. Ravikumar J Working as Assistant professor at Dr. Ambedkar Institute of Technology, Bengaluru, having about 8 years of Teaching and 1 year of industry Experience and area of Interest is Digital Image processing, computer networks and IOT

Dr Ramakanth Kumar P Working as Professor & HoD Department of Computer Science and Engineering, having 25 years of Teaching and 14 years of R&D Experience and area of interest is Digital Image Processing, Pattern Recognition, Natural Language processing etc