

# A Year of Papers Using Biomedical Texts: Findings from the Section on Natural Language Processing of the IMIA Yearbook

Natalia Grabar<sup>1,2</sup>, Cyril Grouin<sup>1</sup>, Section Editors for the IMIA Yearbook Section on Natural Language Processing

<sup>1</sup> LIMSI, CNRS, Université Paris-Saclay, Orsay, France

<sup>2</sup> STL, CNRS, Université de Lille, Villeneuve-d'Ascq, France

## Summary

**Objectives:** To analyze the content of publications within the medical Natural Language Processing (NLP) domain in 2018.

**Methods:** Automatic and manual pre-selection of publications to be reviewed, and selection of the best NLP papers of the year. Analysis of the important issues.

**Results:** Two best papers have been selected this year. One dedicated to the generation of multi- documents summaries and another dedicated to the generation of imaging reports. We also proposed an analysis of the content of main research trends of NLP publications in 2018.

**Conclusions:** The year 2018 is very rich with regard to NLP issues and topics addressed. It shows the will of researchers to go towards robust and reproducible results. Researchers also prove to be creative for original issues and approaches.

## Keywords

Natural Language Processing; semi-automatic selection of publication; topics; issues

Yearb Med Inform 2019;218-23

<http://dx.doi.org/10.1055/s-0039-1677937>

## Introduction

Natural Language Processing (NLP) aims at providing methods, tools, and resources designed in order to mine textual and narrative documents, and to make it possible to access the information they convey [1, 2]. We consider as relevant, any research based on NLP, either using new or current methods, or producing new results. This year, we also focused on NLP methods used in applications, such as those described in the two best papers we selected: the automatic summarization of several documents using clustering of sentences [3], and the automatic generation of medical imaging reports [4].

Among the languages processed, the main place is occupied by the English language, certainly because the resources (corpora, reference-annotated data, lexica, terminologies, etc.) are mainly available for this language, and also because it is often easier to publish the research work done on the English data. This language is indeed managed and understood by all the research community. Yet, works on data from other languages can also be found. Among the pre-selected publications, we can for instance mention: Chinese, addressed through the manually annotated dataset containing 540 breast radiology reports [5]; Italian, addressed through 5,432 non-annotated medical reports belonging to patients with rare arrhythmias and the manually curated hospital database [6]; French, addressed for detecting medical events as an epidemiological purpose [7], or through the clinical data warehouse for thousands of patients [8];

Japanese, addressed through the data set of 5,000 patient pharmacovigilance complaints [9]; and Korean, addressed through 30 rheumatic patients discharge summaries annotated with temporal information [10]. Besides, non-clinical data have been used in Spanish and English for the building of vocabulary useful for patients to better understand the medical and health information in these languages [11]. In almost all these works, the corpora built and annotated are proprietary and can only be used by the teams that built them in collaboration with their clinical colleagues.

As can be seen from this rapid survey, the published research work may be limited by the availability or non-availability of resources in a given language. This is certainly the main reason that motivated works which purpose is to make available health and medical data close to clinical data. We can mention two kinds of works, done in German and French:

- The exploitation of clinical cases written by medical professionals for educational purposes and published in medical scientific literature [12, 13]. In this way, patients privacy concerns are removed since no real, concrete individuals are addressed;
- The software for trustful reconstruction of corpus copies [12]. This may be achieved by extracting well-specified text fragments from e-books and assembling, on demand, identical copies of the same text corpus. In this way, the risk of ethical violation is removed since no physical corpus is distributed.

## About the Selection Process

In order to retrieve all papers published in 2018 in the field of Natural Language Processing, we queried two databases: Medline, specifically dedicated to the biomedical domain, and the ACL (Association for Computational Linguistics) anthology, a database that brings together the major NLP conferences (ACL, International Conference on Computational Linguistics (COLING), European Chapter of the ACL (EACL), Empirical Methods in NLP (EMNLP), International Joint Conference on NLP (IJCNLP), Language Resources and Evaluation (LREC), North American Chapter of the ACL (NAACL), etc.) and journals (CL, Transactions of the ACL).

We applied a basic query on Medline (Figure 1) to target all journal papers published in English in 2018, having abstract, and composed of sequences “*clinical language processing*” or “*medical language processing*” or “*natural language processing*”. As of 2019, February 1<sup>st</sup>, we collected 435 entries. We applied a similar query on the ACL anthology database and collected 130 entries. In order to process those 565 papers, since some of these papers are not related to NLP despite the use of one of the three previously defined sequences, we considered positive and negative filters.

A first set of filters was applied to names of journal, based on both full name and concepts found in the name: a positive score was given to main journals where biomedical NLP work is published (*Biomedical informatics insights*, *IJMI*, *JAMIA*, *JB*, *BMJ bioinformatics*) while a negative score was given to journals that do not concern NLP but are mainly related to cognitive studies or communication disorders (*Human brain mapping*, *Frontiers in neuroscience*, *communication disorders*, etc.).

A second set of filters was used for concepts mainly found in both title and abstract of papers. A positive score was given to concepts generally found in papers related to NLP; those concepts may concern objectives, resources, tools, methods, and evaluation metrics (*named entity recognition*, *part of speech*, *tagged words*, *EHR*, *Pubmed*, *Social Media*, *CRF*, *F1-score*, etc.). A negative score was given to concepts used in studies about disorders involving anatomical parts or

language abilities (*language production or comprehension*, *cortex*, *chemical fragment*, *pMMRs*, etc.) or to papers claiming at using NLP while it was limited to the analysis of tools results rather than improvements made to NLP methods.

For each of the 565 papers, the final score ranked from 0.05 to 0.95 (cf. Figure 2).

We used this score as a meta-element in the best papers selection. Hence, both section editors independently browsed abstracts, keywords, and automatic scores, and gave a Yes / Maybe / No score for each paper. As a result, 112 candidates were kept (19%). We then performed an adjudication process focusing on the geographic origin of papers so as to provide a diversity: out of the 15 selected best papers, eight came from the USA, two from Italy, and one from France, Iran, China, South Korea and Japan.

## Results

The issue on the robustness of methods and results has become increasingly important in the NLP domain. This concern can be observed through several facts: making available the medical and health corpora as described in previous section [12, 13]; and working with different methods for a given task, and in cross-domain or cross-language contexts, as will be presented in this section.

Hence, in one work on Chinese, the researchers worked with breast radiology reports [5]. The purpose was to extract BI-RADS finding categories from these reports. The researchers developed and compared three different types of NLP approaches, including a rule-based method, a traditional machine learning-based method using the Conditional Random Fields (CRF) algorithm, and deep learning-based approaches. On a manually annotated dataset with 540 reports, the evaluation shows that the deep learning-based method achieved the best F1-score of 0.904, when compared with rule-based and CRF-based approaches (0.848 and 0.881, respectively). Similar issue was addressed for the drug safety surveillance in electronic health records in English through the use of classical learning (Support Vector Machines (SVMs)) and deep learning [14]. On the expert-annotated corpus with 791 Electronic Health Record (EHR) notes, the SVM model achieved the best average F1-score of 89.1% on test data, outperforming the long short-term memory (LSTM) model with attention (F1-score of 65.72%) as well as the rule induction baseline system (F1-score of 7.47%).

From another point of view, the researchers studied the problem of cross-domain and cross-language validity of methods and models. Since medical records are written by clinicians from different specialties, the named entity recognition (NER) may be

```
(English[LA] AND journal article[PT] AND 2018[dp] AND ((medical OR
clinical OR natural) AND "language processing"))
```

Fig. 1 Query used for collecting candidate publications for review

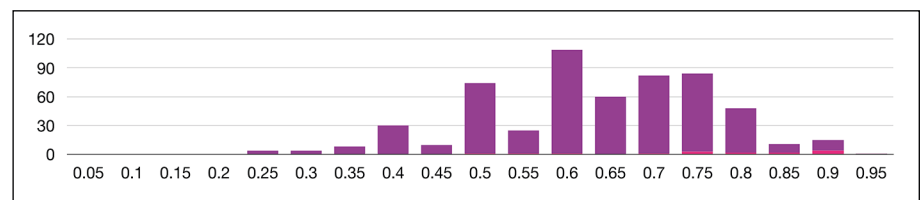


Fig. 2 Distribution of papers according to the filter scores

complicated. Hence, the authors proposed a label-aware double transfer-learning framework (La-DTL) for cross-specialty NER, so that a medical NER system designed for one specialty could be conveniently applied to another one with minimal annotation efforts [15]. The transferability is guaranteed by two components: (i) label-aware Maximum Mean Discrepancy (MMD) for feature representation transfer, and (ii) parameter transfer with a theoretical upper bound which is also label aware. The results demonstrated that La-DTL provides consistent accuracy improvement over strong baselines. Besides, in a cross-language context, the researchers developed an ontology-based approach for the identification of events and their attributes in medical reports written in Italian [6], and they tested the approach on English documents. It showed above 90% precision on Italian and promising results in English.

## Natural Language Processing and Application Contexts

We observed that in several publications, NLP methods and tools were used to investigate the issues relevant for other domains and applications, such as information retrieval, speech pathology, or generation of imaging reports. We assume that this is a very interesting observation because it indicates that NLP methods and tools are becoming mature and usable in different application contexts.

The already mentioned work by Viani *et al.*, [6] aims to identify events and their attributes from episodes of care in medical reports written in Italian. The information retrieval approach exploits a non-annotated corpus of medical reports of patients with rare arrhythmias, a domain-specific ontology that includes the events and attributes to be extracted, and a rule-based NLP system. The evaluation is performed on an independent test set containing manually curated hospital database, which stores most of the information written in reports. The proposed approach shows above 90% precision for most considered clinical events.

In another work, NLP tools and resources are exploited as early, non-invasive biomarkers for the identification of “pre-clinical” Alzheimer’s disease and other dementias

**Table 1** Best paper selection of articles for the IMIA Yearbook of Medical Informatics 2019 in the section ‘Natural Language Processing’. The articles are listed in alphabetical order of the first author’s surname.

Section
Natural Language Processing
<ul style="list-style-type: none"> <li>▪ Jing B, Xie P, Xing E. On the automatic generation of medical imaging reports. <i>Proc of ACL</i> 2018. Melbourne, Australia; 2018. p. 2577–86.</li> <li>▪ Moradi M. CIBS: A biomedical text summarizer using topic-based sentence clustering. <i>J Biomed Inform</i> 2018;88:53–61.</li> </ul>

[16]. Indeed, recent studies suggested that speech alterations might be one of the earliest signs of cognitive decline. Since traditional neuropsychological language tests provide ambiguous results, the authors propose to analyse the spoken language productions using NLP techniques and, in this way, pinpoint language modifications of patients. Ninety-six participants were enrolled (48 healthy controls and 48 cognitively impaired participants). Each subject underwent a brief neuropsychological screening. The spontaneous speech during three tasks (describing a complex picture, a typical working day, and recalling a last remembered dream) was then recorded, transcribed, and annotated at various linguistic levels. A multidimensional parameter computation was performed by a quantitative analysis of spoken texts, computing rhythmic, acoustic, lexical, morpho-syntactic, and syntactic features. The results showed significant differences between controls and impaired participants: in the linguistic experiments, a number of features regarding lexical, acoustic, and syntactic aspects were significant. The authors concluded that this was a promising issue for the identification of pre-clinical stages of dementia.

## Original Issues

The 2018 pre-selection of publications shows several original issues addressed by the researchers. One such issue is dedicated to the health literacy of people and the simplification of texts in order to help patients maintain good health and manage their diseases. Yet, medical educational texts are often written beyond the reading level of the average individual. One work proposes

to isolate particularly difficult terms within documents and to replace them with easier synonyms or explanations in plain language [11]. The main accent is put on the automatic generation of explanations for difficult terms in English and Spanish. The proposed algorithm (SubSimplify) uses word-level parsing techniques and specialized medical affix dictionaries to identify the morphological units of terms and then source their definitions. For the evaluation, the authors extracted 400 difficult terms. For English terms, they compared SubSimplify explanations with the explanations from the Consumer Health Vocabulary, WordNet Synonyms and Summaries, as well as Word Embedding Vector (WEV) synonyms. For Spanish terms, they compared the explanations to WordNet Summaries and WEV synonyms. The vocabulary was also evaluated for quality, coverage, and usefulness of the simplification, for which the proposed resource outperforms all existing written resources. In another work, researchers propose to link medical terms to lay definitions [17]. The proposed system NoteAid incorporates two core components: CoDeMed, a lexical resource of lay definitions for medical terms, and MedLink, a computational unit that links medical terms to lay definitions. Ten physicians evaluated the user interface and the content quality of the system. The results indicate that the system is easy to use, has good visual display, has satisfactory system speed, and proposes adequate lay definitions.

Another issue addresses the generation of misspellings for a better mining of noisy health-related texts such as those provided by social media [18]. Health-related terms are indeed often misspelled there, resulting in the exclusion of relevant data for studies. The authors propose a system that automatically

generates common misspellings for complex health-related terms. The spelling variant generator relies on a dense vector model learned from large, unlabeled text, which is used to find semantically close terms to the original keyword, followed by the filtering of terms that are lexically dissimilar beyond a given threshold. This system outperforms the current state-of-the-art medication name variant generator with best F1-score of 0.69 and F14-score of 0.78. Extrinsic evaluation of the system on a set of cancer-related terms demonstrated an increase of over 67% in retrieval rate from Twitter posts when the generated variants were included.

Yet another interesting issue proposes to detect novel and emerging drug terms in social media [19]. This work is motivated by the rapid development of new psychoactive substances and changes in the use of more traditional drugs. Researchers propose to use distributed word-vector embeddings trained on social media data to uncover drug terms previously unknown to researchers. For instance, for the term marijuana a list of 200 candidate terms for the target substance was produced. Among them, 115 were considered to be related to marijuana (65 terms for the substance itself, 50 terms for paraphernalia), and 30 terms were fully new to the experts.

## Papers with an Emphasis on Methods

Several works design less common methods and approaches. In one work, the graph representation of information and data may bring new insights into already known facts. It can help in the identification of previously unknown treatment and of causal relations between biomedical entities [20]. The method is based on the Unified Medical Language System (UMLS) relations (7,000 treats and 2,918 causes relations), the graph pattern features (paths) extracted from the SemMedDB graph, and logistic regression and decision tree models. The two models predict treats and causes relations with high F-scores of 99% and 90%, respectively, while the logistic regression model coefficients help to identify highly discriminative patterns. In the second work, co-occurrence

graphs are exploited for word sense disambiguation [21]. The knowledge comes from the context of the ambiguous terms and not from the UMLS. The authors work with PubMed abstracts and the personalized PageRank algorithm. Their system outperforms state-of-the-art knowledge-based systems, obtaining more than 10% of accuracy improvement in some cases.

The word embeddings permit to generate semantic representations and vectors for terms. A comparison between the embeddings trained on various resources (clinical notes, biomedical publications, Wikipedia, and news) is proposed [22]. The qualitative evaluation is done manually analyzing the five most similar words computed by embeddings for each term, while the quantitative evaluation measures the capacity to capture the semantic similarity between medical terms, and to improve the results of several biomedical NLP applications. Most of the evaluation shows the efficiency of the word embeddings trained on data from EHRs. Since medical terms are usually composed of several words, it also becomes important to propose and evaluate aggregation methods for multi-word terms, such as summation of component word vectors, mean of component word vectors, direct construction of compound term vectors using the compoundify tool, and direct construction of concept vectors using the MetaMap tool [23]. This work is positioned within the task of semantic similarity and relatedness in the biomedical domain. Besides, word embeddings are exploited in a great variety of works, such as for the detection of novel and emerging drug terms [19] or for the simplification [11]. It was also noted that the quality of word embeddings can be improved through the combination of corpora and knowledge bases [24].

When addressing more standard issues, methods involved often show mature results and permit to go further in the investigation of research questions. The clinical documents, mainly in English, continue to be widely exploited when looking for various types of information, such as the identification of health outcomes [25], adverse drug events [9, 26, 27], diagnostic criteria for autism spectrum disorders [28], sentiment analysis through the subjective expressions made by clinicians [29], similar clinical notes

[30], or temporal segmentation in patient histories [10]. According to the tasks aimed, different approaches are exploited (machine learning and deep learning, rule-based). On the other side, patient-generated texts from social media are used for managing patient expression and for concept normalization [31], for mining suicide risk [32], for detecting novel and emerging drug terms [19], for mining rare health-related events related to birth defects [33], or for detecting the topics addressed by patients in breast cancer forum [34]. Different approaches are exploited but rule-based approaches remain frequent for huge amount of data.

## Concluding Remarks

An analysis of 2018 publications related with medical NLP points out a rich and outstanding research year. Research works addressed indeed important topics related to the robustness and reproducibility of methods and results, and they provided several datasets which can be usable in works to come. Some NLP methods prove to be mature and useful for other applications (information retrieval, diagnosis of speech pathologies and dementias, generation of imaging reports, etc.). As the research advances, we expect that other application domains may become concerned. In 2018, researchers addressed some novel issues and used original approaches. However, more classical issues and approaches cover a wide range of research questions and provide interesting and exploitable results.

## References

1. Friedman C, Hripisak G. Natural Language Processing and its future in medicine. *Acad Med* 1999 Aug;74(8):890-5.
2. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural Language Processing: an introduction. *J Am Med Inform Assoc* 2011;18:544-51.
3. Moradi M. CIBS: A biomedical text summarizer using topic-based sentence clustering. *J Biomed Inform* 2018 Dec;88:53-61.
4. Jing B, Xie P, Xing E. On the automatic generation of medical imaging reports. *Proc of ACL* 2018. Melbourne, Australia; 2018. p. 2577-86.
5. Miao S, Xu T, Wu Y, Xie H, Wang J, Jing S, et al. Extraction of bi-rads findings from breast ultrasound reports in Chinese using deep learning



- approaches. *Int J Med Inform* 2018;119:17–21.
6. Viani N, Larizza C, Tibollo V, Napolitano C, Priori S, Bellazzi R, et al. Information extraction from Italian medical reports: An ontology-driven approach. *Int J Med Inform* 2018;111:140–8.
  7. Tvardik N, Kergourlay I, Bittar A, Segond F, Darmoni S, Metzger MH. Accuracy of using natural language processing methods for identifying healthcare-associated infections. *Int J Med Inform* 2018 Sep;117:96–102.
  8. Garcelon N, Neuraz A, Salomon R, Bahi-Buisson N, Amiel J, Picard C, et al. Next generation phenotyping using narrative reports in a rare disease clinical data warehouse. *Orphanet J Rare Dis* 2018;13(1):85.
  9. Usui M, Aramaki E, Iwao T, Wakammiya S, Sakamoto T, Mochizuki M. Extraction and standardization of patient complaints from electronic medication histories for pharmacovigilance: Natural language processing analysis in Japanese. *JMIR Med Inform* 2018;6(3):e11021.
  10. Lee W, Choi J. Temporal segmentation for capturing snapshots of patient histories in Korean clinical narrative. *Health Inform Res* 2018;24(3):179–86.
  11. Kloehn N, Leroy G, Kauchak D, Gu Y, Colina S, Yuan N, et al. Improving consumer understanding of medical text: Development and validation of a new subsimplify algorithm to automatically generate term explanations in English and Spanish. *J Med Internet Res* 2018;20(8):e10779.
  12. Lohr C, Buechel S, Hahn U. Sharing copies of synthetic clinical corpora without physical distribution - a case study to get around IPRs and privacy constraints featuring the German JSYNCC corpus. In: *Proc of LREC*; 2018. p. 1259–66.
  13. Grabar N, Claveau V, Dalloux C. Cas: French corpus with clinical cases. In: *Proc of LOUHI* 2018. p. 1–12.
  14. Munkhdalai T, Liu F, Yu H. Clinical relation extraction toward drug safety surveillance using electronic health record narratives: Classical learning versus deep learning. *JMIR Public Health Surveill* 2018;4(2):29–40.
  15. Wang Z, Qu Y, Chen L, Shen J, Zhang W, Zhang S, et al. Label-aware double transfer learning for cross-specialty medical named entity recognition. In: *Proc of NAACL-HLT*; 2018. p. 1–15.
  16. Beltrami D, Gagliardi G, Favretti RR, Ghidoni E, Tamburini F, Calzà L. Speech analysis by natural language processing techniques: A possible tool for very early detection of cognitive decline? *Front Aging Neurosci* 2018;10:369.
  17. Chen J, Druhl E, Ramesh BP, Houston TK, Brandt CA, Zulman DM, et al. A natural language processing system that links medical terms in electronic health record notes to lay definitions: System development using physician reviews. *J Med Internet Res* 2018;20(1):1–15.
  18. Sarker A, Gonzalez-Hernandez G. An unsupervised and customizable misspelling generator for mining noisy health-related text sources. *J Biomed Inform* 2018;88:98–107.
  19. Simpson S, Adams N, Brugman C, Connors T. Detecting novel and emerging drug terms using natural language processing: A social media corpus study. *JMIR Public Health Surveill* 2018;4(1):e2.
  20. Bakal G, Talari P, Kakani E, Kavuluru R. Exploiting semantic patterns over biomedical knowledge graphs for predicting treatment and causative relations. *J Biomed Inform* 2018;82:189–99.
  21. Duque A, Stevenson M, Martinez-Romo J, Araujo L. Co-occurrence graphs for word sense disambiguation in the biomedical domain. *Artif Intell Med* 2018;87:9–19.
  22. Wang Y, Liu S, Afzal N, Rastegar-Mojarad M, Wang L, Shen F, et al. A comparison of word embeddings for the biomedical natural language processing. *J Biomed Inform* 2018;87:12–20.
  23. Henry S, Cuffy C, Mc Innes B. (2018) Vector representations of multi-word terms for semantic relatedness. *J Biomed Inform* 2018;77:111–19.
  24. Alsuhailani M, Bollegala D, Machara T, Kawarabayashi K. Jointly learning word embeddings using a corpus and a knowledge base. *PLoS One* 2018;13(3):e0193094.
  25. Wong J, Horwitz M, Zhou L, Toh S. Using machine learning to identify health outcomes from electronic health record data. *Curr Epidemiol Rep* 2018;5(4):331–42.
  26. Li F, Liu W, Yu H. Extraction of information related to adverse drug events from electronic health record notes: Design of an end-to-end model based on deep learning. *JMIR Med Inform* 2018;6(4):12159–70.
  27. Combi C, Zorzi M, Pozzani G, Moretti U, Arzenton E. From narrative descriptions to meddra: automatically encoding adverse drug reactions. *J Biomed Inform* 2018;84:184–99.
  28. Leroy G, Gu Y, Pettygrove S, Galindo M, Arora A, Kurzius-Spencer M. Automated extraction of diagnostic criteria from electronic health records for autism spectrum disorders: Development, evaluation, and application. *J Med Internet Res* 2018;20(11):10497.
  29. Weissman G, Ungar L, Harhay M, Courtright K, Halpern S. Construct validity of six sentiment analysis methods in the text of encounter notes of patients with critical illness. *J Biomed Inform* 2019;89:114–21.
  30. Gabriel R, Kuo T, McAuley J, Hsu C. Identifying and characterizing highly similar notes in big clinical note datasets. *J Biomed Inform* 2018;82:63–9.
  31. Tutubalina E, Miftahudinov Z, Nikolenko S, Malykh V. Medical concept normalization in social media posts with recurrent neural networks. *J Biomed Inform* 2018;84:93–102.
  32. Coppersmith G, Leary R, Crutcheley P, Fine A. Natural language processing of social media as screening for suicide risk. *Biomed Inform Insights* 2018;10:1178222618792860.
  33. Klein A, Sarker A, Cai H, Weissenbacher D, Gonzalez-Hernandez G. Social media mining for birth defects research: A rule-based, bootstrapping approach to collecting data for rare health-related events on Twitter. *J Biomed Inform* 2018;87:68–78.
  34. Jones J, Pradhan M, Hosseini M, Kulanthaivel A, Hosseini M. Novel approach to cluster patient-generated data into actionable topics: Case study of a web-based breast cancer forum. *JMIR Med Inform* 2018;6(4):e45.

## Correspondence to:

Natalia Grabar  
STL, CNRS, Université de Lille  
Domaine du Pont-de-bois  
59653 Villeneuve-d'Ascq cedex  
France  
E-mail: natalia.grabar@univ-lille.fr

## Appendix: Content Summaries of Best Papers for the Natural Language Processing Section of the 2019 IMIA Yearbook

Jing B, Xie P, Xing E

On the automatic generation of medical imaging reports

Proc of ACL 2018. Melbourne, Australia; 2018. p. 2577–86

This paper presents a system for the automatic generation of medical imaging reports, which may help physicians create high-quality reports, reduce time, and assist less experi-

enced physicians. To address these issues, the authors propose to focus on the detection of multiple heterogeneous forms of information including findings and tags, and the detection of abnormal regions in medical images. To cope with these challenges, the authors (i) build a multi-task learning framework which jointly performs the prediction of tags and the generation of paragraphs; (ii) propose a co-attention mechanism to localize regions containing abnormalities and generate narrations for them; and (iii) develop a hierarchical LSTM (Long short-term memory) model to generate long paragraphs. They demonstrate the effectiveness of the proposed methods on two publicly available datasets.

Moradi M

CIBS: A biomedical text summarizer using

topic-based sentence clustering

J Biomed Inform 2018;88:53–61

This paper addresses the creation of multi-document summaries using topic-based sentence clustering. The proposed summarizer extracts biomedical concepts from the input documents and employs an Itemset mining algorithm to discover main topics. It then applies a clustering algorithm to group the sentences into clusters. Finally, the selection of sentences from all the clusters produces a summary that covers a wide range of topics of the input text. The results show that the CIBS method can improve the performance of single- and multi-document biomedical text summarization.