Improving Automatic Categorization of Technical vs. Laymen Medical Words using FastText Word Embeddings

Hanna Pylieva, Artem Chernodub, Natalia Grabar, Thierry Hamon

Improving Automatic Categorization of Technical vs. Laymen Medical Words using FastText Word Embeddings

Hanna Pylieva[1], Artem Chernodub[1,2,] Natalia Grabar[3] and Thierry Hamon[4,5]

[1] Ukrainian Catholic University, Faculty of Applied Sciences,
Kozelnytska st. 2a, Lviv, Ukraine
[2] Institute of Mathematical Machines and Systems Problems NASU, Neurotechnologies Dept.,
Glushkova 42 ave., Kyiv, Ukraine
[3] CNRS, Univ. Lille, UMR 8163 - STL - Savoirs Textes Langage, F-59000, Lille, France
[4] LIMSI, CNRS, Université Paris-Saclay, F-91405 Orsay, France
[5] Université Paris 13, Sorbonne Paris Cité, F-93430, Villetaneuse, France

Abstract. Detection of difficult for understanding words is a crucial task for ensuring the proper understanding of medical texts such as diagnoses and drug instructions. In this paper, we study usage of recently developed word embeddings, which contain context information for words together with other linguistic and non-linguistic features, for improving the detection of difficult medical words. We propose new cross-validation scenarios in order to test the generalization ability of the medical words difficulty detection from different perspectives and provide the experimental study of previously used methods for feature extraction together with recently proposed FastText embeddings. We found that for known words and unknown users FastText embeddings surely improves the detection of word understandability reaching 85.9 F-score (up to 2.9 F-score improvement).

Keywords: text simplification, difficulty detection, word embeddings

1 Introduction

Specialized areas, such as medical area, convey and use technical words, or terms, which are typically related to knowledge developed within these areas. In the medical area, this specific knowledge often corresponds to fundamental medical notions related to disorders, procedures, treatments, human anatomy, etc. For instance, technical terms like *blepharospasm* (abnormal contraction or twitch of the eyelid), *alexithymia* (inability to identify and describe emotions in the self), *appendicectomy* (surgical removal of the vermiform appendix from intestine), or *lombalgia* (low back pain) are frequently used in the medical area texts.

As in any specialized areas, two main kinds of users exist in the medical area:

- medical doctors, both researchers of practitioners, are experts of the domain. They contribute to the creation and development of biomedical knowledge and its exploitation for the healthcare process of patients;

- patients and their relatives are consumers of the healthcare process. Usually, they do not have expert knowledge, while it is important that they understand the purpose and issues of their healthcare process.

If the understanding of technical medical terms is easy for the medical staff, patients and their relatives may present some difficulties in the understanding and using of such terms: they show indeed poor *health literacy*.

Hence, the existing literature provides several studies dedicated to the understanding of medical notions and terms by non-expert users, and on their impact on a successful healthcare process [1-2]. Yet, it is not uncommon that patients and their relatives must face very technical health documents and information. Examples of this kind are frequent and usually the non-expert users are at loss in such situations:

- understanding of information on drug intake [3-4], such as instructions related to the description and specification of steps necessary for the preparation and intake of drugs,
- understanding of clinical documents [5], which contain important information on the healthcare process of patients,
- understanding of clinical brochures or informed consents [6], which are specifically created for patients and which are typically read by patients during their clinical pathway,
- more generally, understanding of information provided for patients by different websites [7-8] in different languages (English, Spanish, French) and different medical specialties,
- for the same reasons, communication between patients and medical staff [9-10] remains complicated.

These various observations provide the main motivation to our work. We propose to address the needs of non-specialized users in the medical domain. As we noticed, the main need is related to the understanding of medical and health information. In what follows, we first present some related work (Section 2). We then introduce the material used (Section 3) and the proposed method (Section 4). Our results and their discussion are presented in Section 5. Finally, we conclude with some directions for future work in Section 6.

2 Related work

Related work is globally related to the detection of technical contents in documents and to their adaptation. Here, we are interested by the first aspect: detection and diagnosis of technical medical contents.

In the NLP (Natural Language Processing) area, work related to the diagnosis of technical medical documents is quite frequent. Traditionally, researchers exploit the readability measures. Among these measures, it is possible to distinguish classical readability measures and computational readability measures [11]. Classical measures

usually rely on number of letters and/or of syllables a word contains and on linear regression models [12], while computational readability measures may involve vector models and a great variability of features, among which the following have been used for processing the biomedical documents: combination of classical readability formulas with medical terminologies [13]; n-grams of characters [14], manually [15] or automatically [16] defined weights of terms, stylistic [17] or discursive [18] features, lexicon [19], morphological features [20], combinations of different features [21].

At a more fine-grained level, detection and diagnosis of technical medical words has been addressed much less frequently. In the general language, some research actions are often performed as part of the NLP challenges, such as the SemEval NLP challenge[1] held in 2012. This challenge proposed the following task: for a short text and a target word, several possible substitutions satisfying the context have also been proposed. The objective was to rate and to order the substitutions according to their degree of simplicity [22]. The participants applied rule-based and/or machine learning systems. Combinations of various features, designed to detect the simplicity of words, have been used, such as: lexicon from spoken corpus and from Wikipedia, Google n-grams, WordNet [23]; word length, number of syllables, latent semantic analysis, mutual information and word frequency[24]; Wikipedia frequency, word length, n-grams of characters and of words, random indexing and syntactic complexity of documents [25]; n-grams and frequency from Wikipedia, Google n-grams [26]; WordNet and word frequency [27]. The best systems reached up to 0.60 Top-rank and 0.575 Recall.

Another work has been done on scholar texts in French written for children with the purpose to differentiate between the texts from various scholar levels and to test various features suitable for that [28]. This system reached up to 0.62 classification accuracy.

In the medical area, we can mention three experiments: manual rating of medical words [15], automatic rating of medical words on the basis of their presence in different vocabularies [16], and exploitation of machine learning approach with various features [30]. This last experiment achieved up to 0.85 F-measure on individual annotations.

The purpose of the current work is to propose novel machine learning approaches for a more efficient distinction of technical medical words which may present understanding difficulties to non-experts users. The medical data processed are in French.

3 Dataset description

3.1 Linguistic data description

For the text classification task aimed the data was collected and annotated as described in [30]. The source terms are obtained from the medical terminology Snomed International [29] in French, available from the ASIP SANTE website[2]. The purpose of this terminology is to provide an extensive description of the medical field. Snomed contains 151,104 medical terms organized into eleven semantic axes such as disorders and

[1] http://www.cs.york.ac.uk/semeval-2012. Accessed 30 October 2018.

[2] http://esante.gouv.fr/services/referentiels/referentiels-d-interoperabilite/snomed-35vf. Accessed 30 October 2018.

abnormalities, procedures, chemical products, living organisms, anatomy, social status, etc. For the purpose of the task, we chose five axes related to the main medical notions: disorders, abnormalities, procedures, functions, and anatomy. Our assumption is that terms in these categories are familiar to a layman, in contrast to contents of such specific groups as chemical products (*hydrogen sulfide*) and living organisms (*Sapromyces, Acholeplasma laidlaw*ii).

The 104,649 selected terms are lemmatized and tokenized into words (or tokens) resulting in 29,641 unique words such that '*trisulfure d'hydrogène'* provides three words (*trisulfure, de, hydrogène*).

The dataset contains three morphological groups of words:

- compound words which contain several bases: abdominoplastie (abdominoplasty), dermabrasion (dermabrasion);
- constructed words which contain one base and at least one affix: cardiaque (cardiac), acineux (acinic), lipoïde (lipoid);
- simple words which contain one base, no affixes and possibly infections (when the lemmatization fails): acné (acne), fragment (fragment).

3.2 Annotation process

The set of 29,641 unique words was annotated by three French speakers, 25-40-year-old, without medical training, without specific medical problems, but with the linguistic background. The annotators are expected to represent the average knowledge of medical words among the population as a whole. The annotators are presented with a list of terms and asked to assign each word to one of the three categories:

- I can understand the word;
- I am not sure about the meaning of the word;
- I cannot understand the word.

The assumption is that the words, which are not understandable by the annotators, are also difficult to understand by patients. The annotators were asked not to use dictionaries during the annotation process. The annotation results are represented in Table 1.

Table 1. Number (and percentage) of words assigned to reference categories by three annotators (A1, A2, and A3)

Categories	*A1 (%)*	*A2 (%)*	*A3 (%)*
1. I can understand	8,099 (28%)	8,625 (29%)	7,529 (25%)
2. I am not sure	1,895 (6%)	1,062 (4%)	1,431 (5%)
3. I cannot understand	19,647 (66%)	19,954 (67%)	20,681 (70%)
Total annotations	29,641	29,641	29,641

4 Machine learning-based categorization

We propose to tackle the problem through the supervised categorization: the purpose is to categorize words, or terms, according to whether they can be understood or not by non-specialized people. The manual annotations of these words provide the reference data. The categorization pipeline is the following. First, for all words in the dataset Natural Language Processing (NLP) features were calculated. Then they were used for training the classifiers. Finally, the quality of the trained classifiers was evaluated using the cross-validation.

4.1 Standard NLP features for words

We will refer to previously used NLP features described in [30] as "*standard features"* (opposed to *"embeddings"* described in the next subsection). They include 24 linguistic and extra-linguistic features related to general and specialized languages. The features are computed automatically and can be grouped into ten classes:

- syntactic categories;
- presence of words in reference lexica;
- the frequency of words through a non-specialized search engine;
- the frequency of words in the medical terminology;
- number and types of semantic categories associated with words;
- length of words as a number of their characters and syllables;
- number of word's bases and affixes;
- initial and final substrings of the words;
- number and percentage of consonants, vowels and other characters;
- classical readability scores.

4.2 Proposed usage of FastText word embeddings

Currently, *word embedding vectors* [32] (or *word vector representations*) are used in the most of state-of-the-art methods for various NLP tasks [31]. Usually, word embeddings are pre-trained on the giant corpora of natural texts such as Google News, Wikipedia texts in an unsupervised manner to predict the context of the target words. They exploit the distributional hypothesis that semantically close words are next to each other in the sentence and that semantically close words share similar contexts.

FastText word embeddings [33] is a good candidate as features for words difficulty detection task because they are able to use words' morphological information and generalize over it. The fact that word embeddings capture context and morphological information leads to the hypothesis that incorporating this information as features will improve classification accuracy for our specific problem. FastText embedding vectors are the sum of character n-gram representations, so that they could be generated even for unknown words. Nevertheless, being trained on Wikipedia texts the portion of known words from our dataset for current FastText embeddings is quite big. According to our analysis, 44.26% (13,118 out of 29,641) medical words in the dataset and 56.00%

(16,598 out of 29,641) lowercased medical words in the dataset were used for training of the currently published French FastText[3] model.

5 Experiments

In [30] different algorithms of supervised classification methods were trained using standard NLP features (section 4.1) to detect the words' understandability. The success of the applied classification algorithms was evaluated within the three accuracy measures: *accuracy (A)*, *precision (P)*, *recall (R)* and *F1-measure (F)*. These scores are weighted average for 1-vs-rest binary classifiers for each of three classes. They allow evaluation of the suitability of the methodology to the difference between understandable and non-understandable words and the relevance of the chosen features to the target problem.

5.1 Experiments on the reproduction of previous results

In order to ensure the consistency of the experiments, first we reproduced the WEKA results using pre-computed standard set of features from [30]. Second, we developed a solution based on decision tree (DT) classifier from well-known scikit-learn library[4]. Here we got 0.85-1.41 lower *F* scores for scikit-learn compared to our own reproduction in WEKA (Table 2).

Table 2. Comparison of various implementations for decision tree classifier on three datasets (A1, A2, A3) in user-in vocabulary-out cross-validation

	Results from [30]	*Our implementation, WEKA J48 DT*	*Our implementation, scikit-learn DT*
A1	80.6	80.5	79.8
A2	81.4	80.9	80.0
A3	84.5	84.5	83.2

Since the input features were identical for WEKA[5] and scikit-learn frameworks, we decided that this small degradation of quality is caused by the different implementations of decision tree classifiers in these frameworks. Nevertheless, in all subsequent experiments we will use the scikit-learn implementation because of it's ease of use for experiments.

5.2 Experiments on user-in vocabulary-out cross-validation

These experiments also follows the scenario from [30]. The cross-validation is done on each dataset (i.e. each user's annotation) separately. The goal of these experiments

[3] https://fasttext.cc. Accessed 30 Oct. 2018.
[4] http://scikit-learn.org. Accessed 30 Oct. 2018.
[5] https://www.cs.waikato.ac.nz/ml/weka. Accessed 30 Oct. 2018.

is to measure the ability of the method to generalize class recognition on the *known user* and his known manner to annotate words (that is, his understanding of the meaning of medical words) for *unknown words*.

We carried out the experiments using (i) the standard features only, (ii) the FastText word embeddings only and (iii) their combination. Experiments with isolated FastText word embeddings as features and the data from three annotators resulted in poor F-scores (Table 3), that can be treated that contextual information which is dominant in the word embeddings is not enough to define the word understandability. Adding the FastText word embeddings to the standard feature set resulted in up to 1% higher F-score due to higher Precision (up to 1.8%), meaning that contextual information slightly impacts on the understandability of a word by a given person.

Table 3. Experiments on user-in vocabulary-out cross-validation

Train user	*Test user*	*Standard features only*				*Embeddings only*				*Standard features + embeddings*			
		A	*P*	*R*	*F*	*A*	*P*	*R*	*F*	*A*	*P*	*R*	*F*
A1	*A1*	**82.5**	77.2	**82.5**	79.8	72.5	67	72.5	69.3	82.4	**79**	82.4	**80.2**
A2	*A2*	**82**	78.9	**82**	80	73.5	69.9	73.5	71.3	81.9	**79.5**	81.9	**80.3**
A3	*A3*	85.5	81.2	85.5	83.2	74.9	70.4	74.9	72.3	**85.9**	**83**	**85.9**	**84.2**

5.3 Experiments on user-out vocabulary-in cross-validation

In this experiment, we learn from all the annotations of one user and then test the model on annotations of another user. In this setting, we measure the ability of the classifier to generalize on all known words, but for unknown users (Table 4). This scenario is plausible to a real-world situation, where it is possible to obtain annotations from a couple of users but not from all the possible users, while it is necessary to predict the familiarity of medical words for all the potential users.

Table 4. Experiments on user-out vocabulary-in cross-validation

Train user	*Test user*	*Standard features only*				*Embeddings only*				*Standard features + embeddings*			
		A	*P*	*R*	*F*	*A*	*P*	*R*	*F*	*A*	*P*	*R*	*F*
A1	*A2*	81.7	78.6	81.7	80.1	74	70.3	74	71.2	**84.2**	**82**	**84.2**	**82.8**
A1	*A3*	85	81.2	85	83	75.4	70.7	75.4	72.6	**87.6**	**84.9**	**87.6**	**85.9**
A2	*A1*	82.2	77	82.2	79.1	72.8	67.3	72.8	69.6	**83.9**	**80.2**	**83.9**	**81.1**
A2	*A3*	85.4	81.1	85.4	83	75.3	71.1	75.3	73	**86.8**	**83.5**	**86.8**	**84.7**
A3	*A1*	82.8	77.4	82.8	79.7	72.7	67.1	72.7	69.4	**84.9**	**81.3**	**84.9**	**82.4**
A3	*A2*	82.2	79	82.2	80.2	74.1	70.4	74.1	71.6	**84.2**	**82.1**	**84.2**	**82.8**

In these experiments we got a significant improvement of combined features in comparison to the standard features. When knowledge of words understandability of one user is used to predict it for another user, adding the FastText word embeddings provides up to 2.9 better F-score. Notice that used separately, standard features and embeddings shows similar performance as in user-in vocabulary-out cross-validation (Table 3). Our hypothesis is that there exists a robust nonlinear dependency between some subsets of standard features and subword-level components of FastText word embeddings. Testing this hypothesis is the topic of our further research.

5.4 Experiments on user-out vocabulary-out cross-validation

In this experiment, we take (k-1) folds of data from one user for training and use k-th fold for testing from the remaining user. In this case, we measure the ability of the method to generalize both on *unknown users* and *unknown vocabulary*.

The cross-validation setting is now the most strict and knowledge of words understandability of one user is used to predict whether another user will understand other medical words. In these experiments, embeddings provide approximately 0.5% higher F-score in case of learning on users A1 and A3 (Table 5). When learning on user A2, embeddings decrease F by 0.5, which means that annotations and health literacy of user A2 are different from users A1 and A3. It seems that adding embeddings makes overfitting of machine learning model to the dataset. As a result, tests on other "kind of word understandability" and on combined features are less successful compared to using standard features only for learning. This may be due to the lack of systematicity in annotations of A2.

Table 5. Experiments on user-out vocabulary-out cross-validation

Train user	*Test user*	*Standard features only*				*Embeddings only*				*Standard features + embeddings*			
		A	*P*	*R*	*F*	*A*	*P*	*R*	*F*	*A*	*P*	*R*	*F*
A1	*A2*	81.7	78.6	81.7	80.1	73.6	69.9	73.6	71.3	**81.8**	**79.8**	**81.8**	**80.6**
A1	*A3*	**85**	81.2	**85**	83	74.8	70.4	74.8	72.4	84.9	**82.2**	84.9	**83.4**
A2	*A1*	**82.2**	76.9	**82.2**	**79.1**	72.5	66.9	72.5	69.3	81.7	**77.5**	81.7	**79.1**
A2	*A3*	**85.3**	81	**85.3**	**83**	75.1	70.7	75.1	72.7	84.4	**81.3**	84.4	82.5
A3	*A2*	**82.7**	77.3	**82.7**	79.7	72.5	66.9	72.5	69.2	82.6	**78.9**	82.6	**80.2**
A3	*A3*	82.1	79	82.1	80.1	73.8	70.2	73.8	71.4	**82.2**	**80**	**82.2**	**80.7**

6 Conclusions

We proposed to address the detection of medical words which understanding may be difficult for non-specialized users of the medical area. We exploit for this machine learning algorithms and several sets of NLP features: standard features (syntactic information, reference lexica, frequency, etc.), distributional features (such as provided by word embeddings), and their combination.

Our results indicate that adding FastText word embeddings provides a significant improvement of the performance for the generalization for unknown users (up to 2.9 F-score) but provides a slight increase (0.5 - 1 F-score) or even decrease (-0.5 F-score) of performance for unknown words. We consider this positive issue because it is important to be able to generalize annotations provided by a set of users on the whole population.

We have several directions for future work. For instance, we will try to understand the reasons of the decrease of performances with word embeddings, which may help to rectify the results. Besides, we currently use existing pre-trained word embeddings. Yet, we assume that their training on medical data may improve their impact on the categorization results. We also plan to implement and test other deep learning/neural networks/NLP methods which use the morphological information of words, such as character-level recurrent neural networks and character embeddings together with 1D convolutions. Indeed, when language data present stable patterns, which is the case in

the medical field, processing of subword strings may help for the generalization over new and unseen words. As we presented above, this is one of the current limitations of our work.

References

1. Mcgray, A.: Promoting health literacy. J of Am Med Infor Ass 12, 152-163 (2005).
2. Eysenbach, G.: Poverty, human development, and the role of ehealth. J Med InternetRes 9(4), 34-4 (2007).
3. Vander Stichele, R.: Promises for a measurement breakthrough. In: Sons, J.W. (ed.) Drug regimen compliance. Issues in clinical trials and patient management, pp. 71-83. JM Metry and UA Meyer (1999).
4. Patel, V., Branch, T., Arocha, J.: Errors in interpreting quantities as procedures: The case of pharmaceutical labels. Int Journ Med Inform 65(3), 193{211 (2002).
5. Zeng-Treiler, Q., Kim, H., Goryachev, S., Keselman, A., Slaugther, L., Smith, C.: Text characteristics of clinical reports and their implications for the readability of personal health records. In: MEDINFO. pp. 1117{1121. Brisbane, Australia (2007).
6. Williams, M., Parker, R., Baker, D., Parikh, N., Pitkin, K., Coates, W., Nurss, J.: Inadequate functional health literacy among patients at two public hospitals. JAMA 274(21), 1677{1682 (1995).
7. Oregon Practice Center: Barriers and drivers of health information technology use for the elderly, chronically ill, and underserved. Tech. rep., Agency for healthcare research and quality. Oregon Evidence-based Practice Center (2008).
8. Brigo, F., Otte, M., Igwe, S., Tezzon, F., Nardone, R.: Clearly written, easily comprehended? The readability of websites providing information on epilepsy. Epilepsy & Behavior 44, 35-39 (2015).
9. Jucks, R., Bromme, R.: Choice of words in doctor-patient communication: an analysis of health-related internet sites. Health Commun 21(3), 267-77 (2007).
10. Tran, T., Chekroud, H., Thiery, P., Julienne, A.: Internet et soins: un tiers invisible dans la relation medecin/patient. In: Ethica Clinica 53, 34-43 (2009).
11. François T., Fairon C. An AI readability formula for French as a foreign language //Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. – Association for Computational Linguistics, 2012. – C. 466-477 (2011).
12. Gunning, R.: The art of clear writing. McGraw Hill, New York, NY (1973).
13. Kokkinakis, D., Toporowska Gronostaj, M.: Comparing lay and professional language in cardiovascular disorders corpora. In: Pham T., James Cook University, A. (ed.) WSEAS Transactions on BIOLOGY and BIOMEDICINE. pp. 429-437 (2006).
14. Poprat, M., Marko, K., Hahn, U.: A language classier that automatically divides medical documents for experts and health care consumers. In: Int Congress of the European Federation for Medical Informatics. pp. 503{508. Maastricht (2006).
15. Zheng, W., Milios, E.,Watters, C.: Filtering for medical news items using a machine learning approach. In: Ann Symp Am Med Inform Assoc (AMIA). pp. 949-953, (2002).
16. Borst, A., Gaudinat, A., Boyer, C., Grabar, N.: Lexically based distinction of readability levels of health documents. In: MIE 2008 (2008), poster.
17. Grabar, N., Krivine, S., Jaulent, M.: Classication of health webpages as expertand non expert with a reduced set of cross-language features. In: Ann Symp AmMed Inform Assoc (AMIA). pp. 284-288 (2007).

18. Goeuriot, Lorraine, Natalia Grabar, and Béatrice Daille. "Caractérisation des discours scientifique et vulgarisé en français, japonais et russe." Poster at TALN (2007).
19. Miller, Trudi, et al. "A Classifier to Evaluate Language Specificity of Medical Documents." 2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07).
20. Chmielik, Jolanta, and Natalia Grabar. "Détection de la spécialisation scientifique et technique des documents biomédicaux grâce aux informations morphologiques." TAL 51.2 (2011): 151-179.
21. Zeng-Treiler, Q., Kim, H., Goryachev, S., Keselman, A., Slaugther, L., Smith, C.: Text characteristics of clinical reports and their implications for the readability of personal health records. In: MEDINFO. pp. 1117-1121. Brisbane, Australia (2007).
22. Specia, L., Jauhar, S., Mihalcea, R.: Semeval-2012 task 1: English lexical simplification. In: *SEM 2012. pp. 347-355 (2012).
23. Sinha, R.: Untsimprank: Systems for lexical simplification ranking. In: *SEM 2012. pp. 493-496 (2012)
24. Jauhar, S., Specia, L.: UOW-SHEF: SimpLex - lexical simplicity ranking based on contextual and psycholinguistic features. In: *SEM 2012. pp. 477-481. Montreal, Canada (2012), http://www.aclweb.org/anthology/S12-1066
25. Johannsen, A., Marttinez, H., Klerke, S., Sogaard, A.: Emnlp@cph: Is frequency all there is to simplicity? In: *SEM 2012. pp. 408-412. Montreal, Canada (2012), http://www.aclweb.org/anthology/S12-1054
26. Ligozat, A., Grouin, C., Garcia-Fernandez, A., Bernhard, D.: Annlor: A naive notation-system for lexical outputs ranking. In: *SEM 2012. pp. 487-492 (2012)
27. Amoia, M., Romanelli, M.: SB: mmSystem - using decompositional semantics for lexical simplification. In: *SEM 2012. pp. 482-486. Montreal, Canada (2012), http: //www.aclweb.org/anthology/S12-1067.
28. Gala, N., Franscois, T., Fairon, C.: Towards a french lexicon with difficulty measures: NLP helping to bridge the gap between traditional dictionaries and specialized lexicons. In: eLEX-2013 (2013).
29. Côté, Roger A. "Répertoire d'anatomopathologie de la SNOMED internationale, v3. 4." Université de Sherbrooke, Sherbrooke, Québec (1996).
30. Grabar N., Hamon T., Amiot D. Automatic diagnosis of understanding of medical words. In: Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR), pp. 11–20, (2014).
31. Repository to track the progress in Natural Language Processing (NLP). https://nlpprogress.com, last accessed 30 October 2018.
32. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems, pp. 3111–3119 (2013).
33. Bojanowski P., Grave E., Joulin A., Mikolov T. Enriching word vectors with subword information. arXiv preprint arXiv:1607.04606 (2016).