Applicability of Machine Learning Methods to Multi-Label Medical Text Classification

Iuliia Lenivtceva (✉), Evgenia Slasten, Mariya Kashina, Georgy Kopanitsa

ITMO University, 49 Kronverkskiy prospect, 197101 Saint Petersburg, Russian Federation

lenivezzki@gmail.com, slastenevgenia@gmail.com,
k.mariya1997@gmail.com, georgy.kopanitsa@gmail.com

Abstract. Structuring medical text using international standards allows to improve interoperability and quality of predictive modelling. Medical text classification task facilitates information extraction. In this work we investigate the applicability of several machine learning models and classifier chains (CC) to medical unstructured text classification. The experimental study was performed on a corpus of 11671 manually labeled Russian medical notes. The results showed that using CC strategy allows to improve classification performance. Ensemble of classifier chains based on linear SVC showed the best result: 0.924 micro F-measure, 0.872 micro precision and 0.927 micro recall.

Keywords: multi-label learning, medical text classification, interoperability, FHIR, data structuring

1 Introduction

Medical data standardization is crucial in terms of data exchange and integration as data formats vary greatly from one healthcare provider to another. Many international standards for terminologies (SNOMED CT [1], LOINC [2]) and data exchange (openEHR [3], ISO13606 [4], HL7 standards [5]) are successfully implemented and perform well in practice. The most developing and perspective standard for medical information today is FHIR-HL7 [6].

The data are usually stored in structured, semi-structured or unstructured form in medical databases. Structured and semi-structured data can be mapped to standards with minimum losses of information [7]. However, a big part of Electronic Health Record (EHR) is in free text [8]. Unstructured medical records are more complicated to process, however, they usually contain detailed information on patients which is valuable in modeling and research [9].

The extraction of useful knowledge becomes more challenging as medical databases become more available and contain a wide range of texts [10]. Sorting documents and searching concepts and entities in texts manually is time-consuming. Text classification is an important task which aims to sort documents or notes according to the predefined classes [11] which facilitates entities extraction such as symptoms [12], drug names

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50423-6_38

[13], dosage [14], drug reactions [15], etc. The task of information extraction (IE) is domain specific and requires considering its specificity in practice. Thus, high performance in IE can be achieved through free text classification to a particular domain [16].

The developed applications and methods for processing free texts are language specific [17]. Russian medical free text processing is challenging mostly because there is no open source medical corpora [18]. Moreover, each medical team develops their own storage format, which makes it difficult to standardize, exchange and integrate Russian medical data.

Our long-term goal is to develop methods for data extraction from Russian unstructured clinical notes and mapping these data on FHIR for better interoperability and personalized medicine. The purpose of the article is to investigate the applicability of machine learning algorithms to classify Russian unstructured and semi-structured allergy anamnesis to facilitate entities extraction.

2 Related work

Studies on text classification using machine learning methods are widely represented in literature.

A. Jain et al [16] describes classifiers based on Multinomial Naïve Bayes (MNB), k-Nearest Neighbors (k-NN) and Support Vector Machine (SVM) as the most popular models for multi-label classification. Logistic regression (LR) is also a widespread model for the task [19].

Binary relevance (BR) approach suggests to train N independent binary classifiers for multi-label classification with N labels. This approach has a linear complexity; however, it does not consider interdependences between labels [19]. Classifier Chains (CC) is a popular and representative algorithm for multi-label classification. CC suggests to link N binary classifiers in a chain with random ordering as it shows better predictive performance of the classification. The set of predicted labels is treated as extra features for the next classifiers in a chain. CC and ensembles [20] are known to solve overfitting problem. CC are more computationally demanding than simple binary classifiers [21].

The performance metrics of multi-label classifiers applied to medical text are represented in table 1. The literature review showed that there is no a single concept on which metrics to use when evaluating multi-label classifiers.

Table 1. Performance of medical multi-label classifiers

Classifier	#labels	Data and tools	F1		PRC		REC		Citation
			micro	macro	micro	macro	micro	macro	
BR	10	Real data	0.78		0.84		0.80		R.-W. Zhao et al [22]
CC			0.79		0.89		0.75		
Binary	45	Open dataset Medical	-	0.38	-	-	-	-	J. Read et al [23]
CC			-	0.39	-	-	-	-	
kNN			-	-	-	-	-	-	

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50423-6_38

LR		WEKA	-	-	-	-	-	-	
Rule-based	7	Real data	0.95		0.96		0.94		Y. Baghdadi et al [24]
SVM			0.99		0.97		0.98		
SVM	6	Open dataset cTAKES	0.83		-		0.934		W.-H. Weng et al [25]
NB	8	Real data WEKA	0.82		0.77		0.89		S. Spat et al [26]
1-NN			0.86		0.87		0.86		
J48			0.88		0.90		0.87		
SVM	45	Real data Manual labeling	0.823	-	0.823	-	0.831	-	A. A. Argaw et al [10]
SVM	2618	Real data	0.683	0.652	-	0.535	-	0.868	L. V. Lita et al [27]
SVM	78	Open dataset	0.530	-	-	-	-	-	T. Baumel et al [28]
BR	420	Real data	0.720	0.706	0.818	0.812	0.643	0.659	R. Kaur et al [8]

3 Methods

3.1 Data Description

Clinical documents (written in Russian) of more than 250 thousand patients were provided by Almazov National Medical Research Centre (St. Petersburg, Russia) for the research. The patients' personal information was discarded. We searched for different forms of the words «allergy» and «(in)tolerance» (Russian equivalents «аллергия», «(не)переносимость») using regular expressions to find all the notes containing any information on allergy and intolerances. The corpus of 269 thousand notes was created after the search and duplicates removal. We classified allergy notes according to four labels which are described in table 2.

Table 2. Classes description

Label	Classes description	Example in Russian	Example in English
AL	A note contains information about allergen or intolerance. It might be the name of a drug or a drug's group (nitrates). A note also might only mention that allergy or intolerance takes place.	Аллергологический анамнез аллергия на укус насекомых. Назначена терапия метотрексан 10 мг, отменена в связи с плохой переносимостью препарата. Аллергологический анамнез аллергия на не помнит	Allergy anamnesis allergy to a bite of an insect. Methotrexate 10 mg treatment was started, but due to the poor tolerance the drug was canceled. Allergy does not remember exactly.

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50423-6_38

R	A note contains information about the reaction to some allergen. The allergen might be specified or not.	Аллергологический анамнез аллергия на атопический дерматит. Аллергия на медикаменты пенициллин крапивница йод нет.	Allergy anamnesis allergy atopic dermatitis. Allergy to medications penicillin urticaria, iodine no.
NN	A note declares that there is no allergy or intolerance.	Аллергия нет.	No allergy.
N	A note does not contain information about allergy or intolerance.	План лечения введение препаратов переносит удовлетворительно.	Treatment plan drug administration tolerates satisfactorily.

Two experts assigned an appropriate label to each note. In case of disagreement the decision was made by consensus.

The final corpus contains 11671 labeled notes.

3.2 Task Description

AllergyIntorence is one of the FHIR resources, it contains structured information on patient's allergies, intolerances and symptoms. The task of mapping this data to FHIR involves machine learning methods as it is stored in unstructured form. Fig. 1 represents the main blocks of information that can be mapped to FHIR. Bold blocks denote information that is mentioned in the processed corpus.

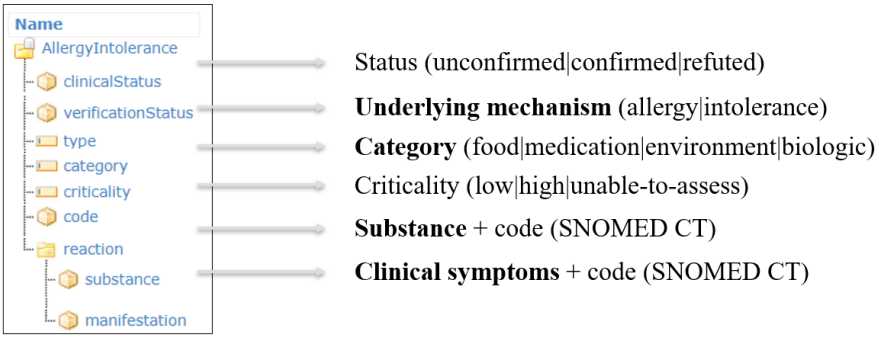

Fig. 1. Blocks of information to be mapped to FHIR

Underlying mechanism can be extracted by searching keywords «allergy» and «intolerance» in the corpus. Category refers to an exact substance type. The most sophisticated task is to extract exact substances and clinical symptoms written in Russian and to bind corresponding codes from international terminological systems to ensure interoperability. To facilitate this task classification of multi-topic clinical notes is required.

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50423-6_38

3.3 Preprocessing

The steps of preprocessing are:

1. Clean medical notes from symbols and extra spaces. Full stops are left as they play an important role in sentence tokenization.
2. Reduce notes to minimize noise during classification as the original note might contain up to 9239 words. Only 2 meaningful sentences before and after regular expression («аллергия», «(не)переносимость») are left.
3. Correct syntactic, case and spaces errors using regular expressions.
4. Dictionary-based spelling correction with Levenshtein distance calculation.
5. Tokenize and normalize words.
6. Train-test split, training set contains 7819 notes and test set – 3852.
7. Vectorize both train and test sets using Bag of Words (BOW) representation. The dictionary size for BOW is 8000 words.

3.4 Classification

We applied four shallow machine learning models: MNB, LR, SVM, k-NN and two ensembles of classifier chains: ECCLR, ECCSVM. The optimal parameters of the shallow models were adjusted by grid search. Optimal parameters of the models are introduced in table 3.

Table 3. Parameters of classifiers

Model	Parameters
Shallow classifiers	
MNB	Alpha: 0.5
LR	Solver: saga, penalty: l2, C=3, max_iter=4000
Linear SVM	Loss: squared hinge, penalty: l2, max_iter=4000, C=1.3684
k-NN	Algorithm: brute, n_neighbors=1, weights: uniform
Ensembles of Classifier Chains	
ECCLR	Ensemble of 10 logistic regression classifier chains with random ordering of labels
ECCSVM	Ensemble of 10 linear SVM classifier chains with random ordering of labels

The pipeline was built using python version 3.7.1. For lexical normalization «pymorphy2» was used. All the preprocessing steps were realized with custom skripts. «scikit-learn» package was used to implement supervised learning algorithms, evaluate models and to perform t-SNE. «Bokeh», «matplotlib» and «plotly» were used for visualization.

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50423-6_38

3.5 Evaluation metrics

According to [21] macro and micro averaging precision, recall and F-measure are often used to evaluate multi-label classification performance. So, we used these metrics to evaluate the performance of the classification.
Micro-averaging:

$$B_{micro}(h) = B\left(\sum_{j=1}^{q} TP_j, \sum_{j=1}^{q} FP_j, \sum_{j=1}^{q} TN_j, \sum_{j=1}^{q} FN_j\right) \quad (1)$$

Macro-averaging:

$$B_{macro}(h) = \frac{1}{q}\sum_{j=1}^{q} B(TP_j, FP_j, TN_j, FN_j) \quad (2)$$

B∈ {Precision, Recall, F^{β}}, q – number of class labels.

Precision (positive predictive value) is the fraction of correctly identified examples of the class among all the examples identified as this class.

$$Precision(TP_j, FP_j, TN_j, FN_j) = \frac{TP_j}{TP_j + FP_j} \quad (3)$$

Recall evaluates the fraction of identified examples from the class among all the examples of this class.

$$Recall(TP_j, FP_j, TN_j, FN_j) = \frac{TP_j}{TP_j + FN_j} \quad (4)$$

F-measure is harmonic mean (β=1) of precision and recall.

$$F^{\beta}(TP_j, FP_j, TN_j, FN_j) = \frac{(1+\beta^2)TP_j}{(1+\beta^2)TP_j + FP_j + \beta^2 FN_j} \quad (5)$$

TP – true positive examples, TN – true negative examples, FP – false positive examples, FN – false negative examples, β=1.

t-SNE was performed using predicted probabilities for each label. The perplexity equals 30 according to recommendations of G.E. van der Maaten et al [29].

4 Results

After text cleaning still there were notes which contained neither allergies nor intolerances.

Fig. 2 illustrates the distribution of classes in the corpus. The classes are imbalanced.

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50423-6_38

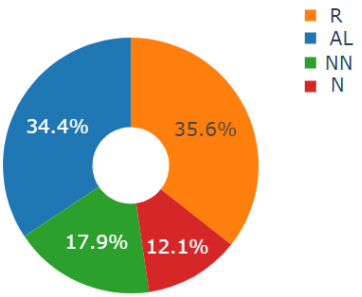

Fig. 2. Classes distribution in the corpus

Performances of different classifiers are represented in table 4. LR and linear SVM showed the best results among shallow classifiers. However, the use of CC with LR and linear SVM as base classifiers improved performance metrics and showed best results.

Table 4. Performance of the applied classifiers

Model	Precision		Recall		F-measure	
	Micro	Macro	Micro	Macro	Micro	Macro
Shallow classifiers						
MNB	0.781	0.764	0.864	0.873	0.864	0.852
LR	**0.866**	**0.850**	**0.920**	0.915	**0.920**	**0.910**
Linear SVM	**0.865**	**0.849**	**0.919**	0.916	**0.919**	**0.909**
k-NN	0.694	0.715	0.803	0.827	0.803	0.809
Ensembles of Classifier Chains						
ECCLR	**0.867**	**0.852**	**0.925**	**0.921**	**0.922**	**0.912**
ECCSVM	**0.872**	**0.855**	**0.927**	**0.922**	**0.924**	**0.914**

Classification report for the best classifier is represented in table 5.

Table 5. Classification report for ECCSVM

	precision	recall	F1-score	support
AL	0.93	0.94	0.94	1317
R	0.95	0.92	0.93	1388
NN	0.92	0.93	0.93	690
N	0.83	0.89	0.86	457
micro avg	0.92	0.93	0.92	3852
macro avg	0.91	0.92	0.91	3852
weighted avg	0.92	0.93	0.92	3852
samples avg	0.92	0.93	0.92	3852

Fig. 3 illustrates t-SNE representation classes.

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50423-6_38

Fig.4, Fig.5, Fig.6, Fig.7 represent 10 most important keywords in the corpus which indicate that the note belongs to the corresponding class. The diagrams show how often each word can be met in the corpus (word counts) and how important this word is for classification (weights of classifier). The diagram is plotted using LR weights.

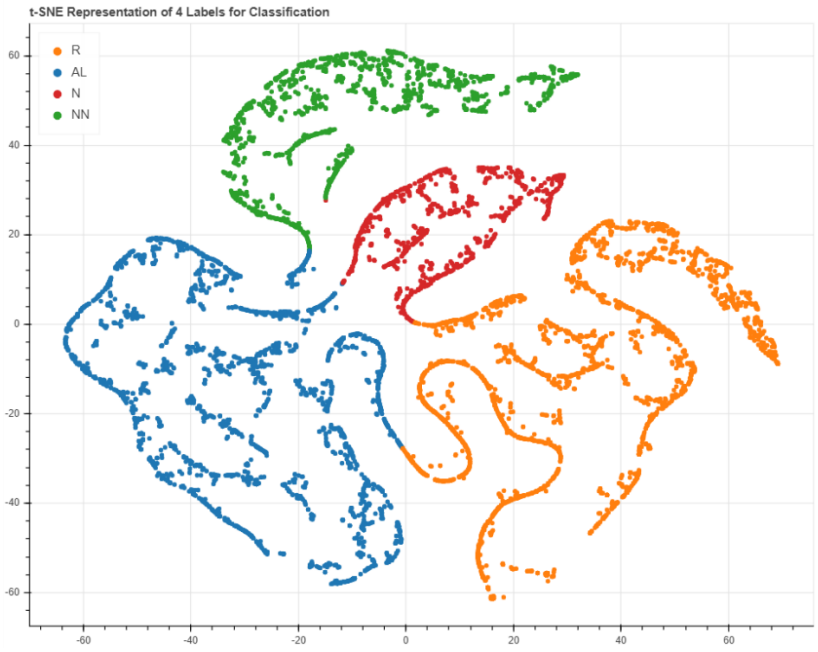

Fig. 3. t-SNE representation of classes

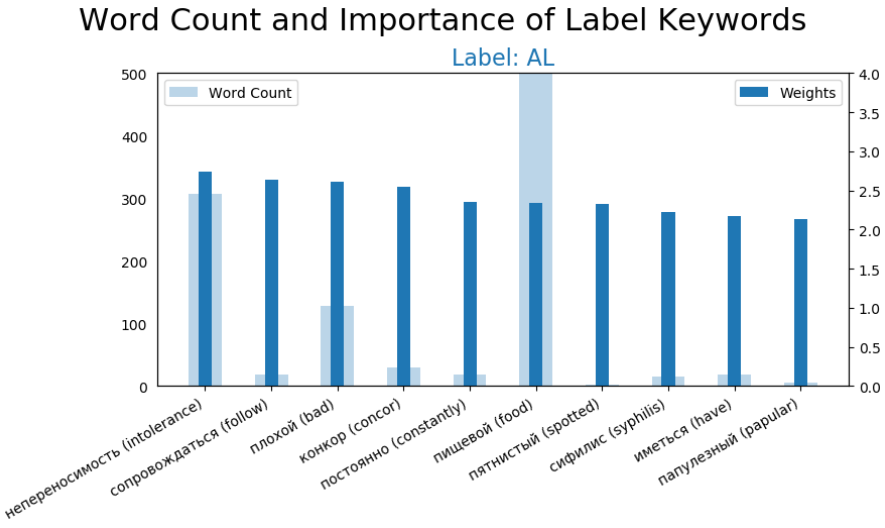

Fig. 4. Top 10 positive keywords for label AL

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50423-6_38

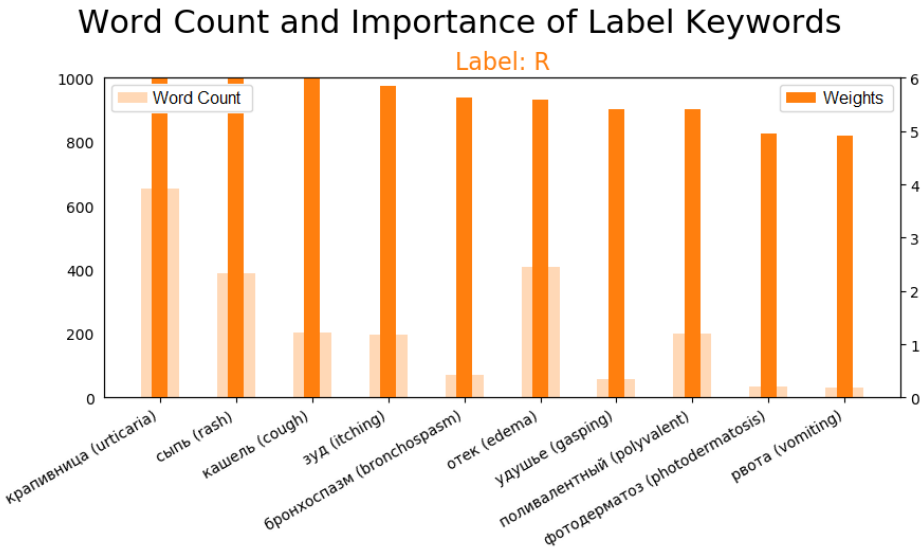

Fig. 5. Top 10 positive keywords for label R

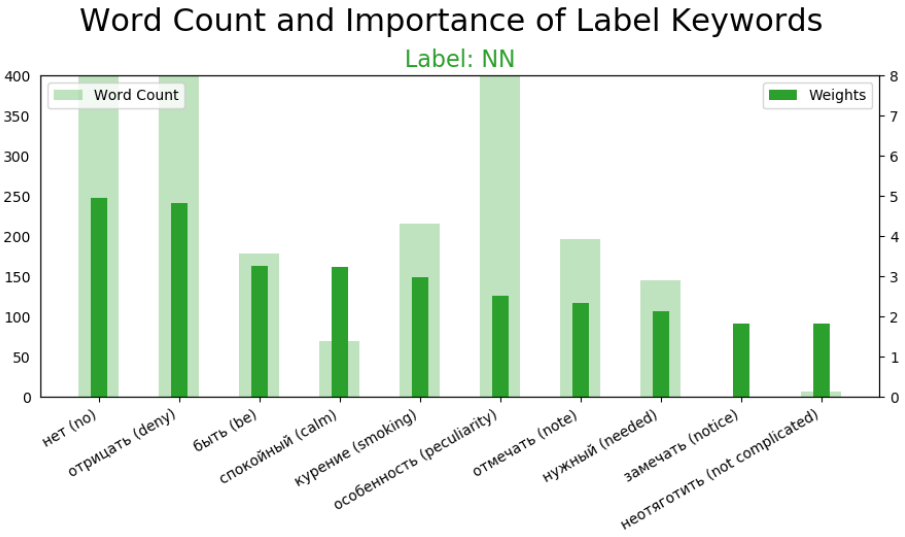

Fig. 6. Top 10 positive keywords for label NN

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50423-6_38

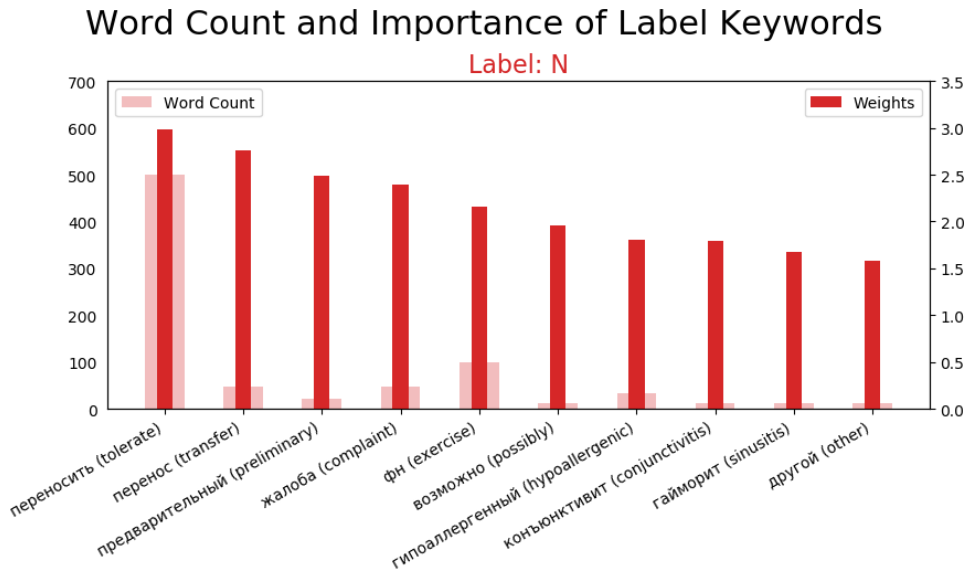

Fig. 7. Top 10 positive keywords for label N

5 Discussion

Regarding previous studies on multi-label medical text classification many authors use applications for entities extraction and algorithms implementation (table 1). However, there is no open source applications for medical purposes developed for the Russian case such as MetaMap [30], for instance. Thus, all the steps were realized manually and with custom scripts.

In the medical text multi-label classification task with limited labeled data we concentrated on improving F-measure as it enforces a better balance between performing on relevant and irrelevant labels and, thus, suitable for multi-label task evaluation [31]. Also, precision, recall and F-measure are not sensitive to classes imbalance.

Two of the proposed shallow classifiers LR and linear SVM performed well on real unstructured labeled data. Using CC strategy allowed to improve the results of basic classifiers and the best performance was shown by ensemble of classifier chains based on linear SVC. Classification report for this classifier (table 5) has shown that three most important labels for mapping AL, R and NN are well separated from each other and from the fourth class N. The fourth class showed lower performance which can be caused by the least number of labeled data in the corpus and the variety of topics covered in it.

Recall is higher than precision for all classifiers and for both averaging strategies. It means that classifiers are good at identifying classes and differentiating them from each other. The number of false negatives is low, which means that classifiers do not intend to lose important notes. This result is satisfying from the point of mapping task as it is important to find as many class representatives as possible.

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50423-6_38

The obtained result of 0.924 micro F-measure, 0.872 micro Precision and 0.927 micro Recall by ECCSVC outperformed almost all the represented in table 1 results. Y. Baghdadi et al [24] reported high overall performance of implemented classifiers and the data were previously standardized. W.-H. Weng et al [25] used additional tools for clinical text processing and information extraction. The closest task was solved by A. A. Argaw et al [10] in terms of real data manual labeling. All the obtained metrics of our ECCSVC are higher, however, the number of labels in the classification task is lower.

t-SNE representation shows that classes are well separated.

Fig. 4 shows 10 most important words associated with allergens and substances. The list of keywords for this task contain such entities as «intolerance» which indicates the presence of patient's intolerance in the text of anamnesis; «food» which is associated with the category of allergy in the FHIR resource; medications such as «concor» which might be associated with a substance in the FHIR resource; number of verbs indicating the presence of allergy such as «follow», «have». The words «intolerance» and «food» are also most frequent words of this class in a corpus.

Fig. 5 shows 10 most important words associated with clinical symptoms in FHIR resources and reactions. All the most frequent keywords of this class are symptoms.

Fig. 6 shows 10 most important words associated with the situation when no allergy was detected. This class keywords contain many negative words such as «no», «deny», «not complicated» and general purpose normalized words, which are usually met in calm allergy anamnesis: «calm», «be», «notice». The keywords of this group are not frequent in a corpus because of low number of labeled notes for this class. The NN notes would be marked as «no allergy» and would not be considered during information extraction and mappings.

Fig. 7 shows 10 most important words associated with class N, which indicates that the exact note is not connected with allergy or intolerance. The most important and frequently met keyword in this class is «tolerate (переносить)». This word has one root with the word «intolerance (непереносимость)». Thus, this word frequent due to the initial mechanism of search. Other keywords represent different topics not connected with allergy and intolerance. Thus, the notes from this class would not be considered during information extraction and mappings.

6 Conclusion

In this study we investigated the applicability of several classifiers to the task of clinical free-text allergy anamnesis classification for filtering multi-topic data.

The research showed that LR, linear SVC, ECCLR and ECCSVC performed well and can be applied to the task of clinical free-text allergy anamnesis classification. The use of chaining strategy improved the performance of shallow classifiers.

In the future we plan to apply a model for Named Entity Recognition (NER) to extract named entities such as allergies and symptoms from medical free text and map them to FHIR. Also, we plan to develop a model to ICD-10 Russian codes and terms identification in medical free-text allergy anamnesis.

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50423-6_38

Acknowledgements. This work financially supported by the government of the Russian Federation through the ITMO fellowship and professorship program. This work was supported by a Russian Fund for Basic research 18-37-20002. This work is financially supported by National Center for Cognitive Research of ITMO University.

References

1. Fung KW, Xu J, Rosenbloom ST, Campbell JR (2019) Using SNOMED CT-encoded problems to improve ICD-10-CM coding—A randomized controlled experiment. Int J Med Inform 126:19–25. https://doi.org/10.1016/j.ijmedinf.2019.03.002
2. Fiebeck J, Gietzelt M, Ballout S, et al (2019) Implementing LOINC: Current status and ongoing work at the Hannover Medical School. In: Studies in Health Technology and Informatics. IOS Press, pp 247–248
3. Mascia C, Uva P, Leo S, Zanetti G (2018) OpenEHR modeling for genomics in clinical practice. Int J Med Inform 120:147–156. https://doi.org/10.1016/j.ijmedinf.2018.10.007
4. Santos MR, Bax MP, Kalra D (2010) Building a logical EHR architecture based on ISO 13606 standard and semantic web technologies. In: Studies in Health Technology and Informatics
5. Ulrich H, Kock AK, Duhm-Harbeck P, et al (2017) Metadata repository for improved data sharing and reuse based on HL7 FHIR. In: Studies in Health Technology and Informatics
6. Hong N, Wen A, Mojarad MR, et al (2018) Standardizing Heterogeneous Annotation Corpora Using HL7 FHIR for Facilitating their Reuse and Integration in Clinical NLP. AMIA . Annu Symp proceedings AMIA Symp 2018:574–583
7. Lenivtseva Y, Kopanitsa G (2019) Investigation of Content Overlap in Proprietary Medical Mappings. Stud Health Technol Inform 258:41–45. https://doi.org/10.3233/978-1-61499-959-1-41
8. Kaur R, Ginige JA (2019) Analysing Effectiveness of Multi-Label Classification in Clinical Coding. In: ACM International Conference Proceeding Series. Association for Computing Machinery
9. Wang Y, Wang L, Rastegar-Mojarad M, et al (2018) Clinical information extraction applications: A literature review. J. Biomed. Inform. 77:34–49
10. Alemu A, Hulth A, Megyesi B (2007) General-Purpose Text Categorization Applied to the Medical Domain. Comput Sci 16:
11. Onan A, Korukoğlu S, Bulut H (2016) Ensemble of keyword extraction methods and classifiers in text classification. Expert Syst Appl 57:232–247. https://doi.org/10.1016/j.eswa.2016.03.045
12. Métivier JP, Serrano L, Charnois T, et al (2015) Automatic symptom extraction from texts to enhance knowledge discovery on rare diseases. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag, pp 249–254
13. Levin MA, Krol M, Doshi AM, Reich DL (2007) Extraction and mapping of drug names from free text to a standardized nomenclature. AMIA Annu Symp Proc 438–442
14. Xu H, Jiang M, Oetjens M, et al (2011) Facilitating pharmacogenetic studies using electronic health records and natural-language processing: a case study of warfarin. J Am Med Informatics Assoc 18:387–391. https://doi.org/10.1136/amiajnl-2011-000208
15. Wang X, Hripcsak G, Markatou M, Friedman C (2009) Active Computerized Pharmacovigilance Using Natural Language Processing, Statistics, and Electronic Health

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50423-6_38

Records: A Feasibility Study. J Am Med Informatics Assoc 16:328–337. https://doi.org/10.1197/jamia.M3028
16. Jain A, Mandowara J (2016) Text Classification by Combining Text Classifiers to Improve the Efficiency of Classification. Int J Comput Appl 6:2250–1797
17. Ali AR, Ijaz M (2009) Urdu text classification. In: Proceedings of the 6th International Conference on Frontiers of Information Technology, FIT '09
18. Toldova S, Lyashevskaya O, Bonch-Osmolovskaya A, Ionov M (2015) Evaluation for morphologically rich language: Russian NLP. In: Proceedings on the International Conference on Artificial Intelligence (ICAI). CSREA Press, Las Vegas, pp 300–306
19. Cheng W, Hüllermeier E (2009) Combining instance-based learning and logistic regression for multilabel classification. In: Machine Learning. pp 211–225
20. Tahir MA, Kittler J, Bouridane A (2012) Multilabel classification using heterogeneous ensemble of multi-label classifiers. Pattern Recognit Lett 33:513–523. https://doi.org/10.1016/j.patrec.2011.10.019
21. Zhang ML, Zhou ZH (2014) A review on multi-label learning algorithms. IEEE Trans. Knowl. Data Eng. 26:1819–1837
22. Zhao RW, Li GZ, Liu JM, Wang X (2013) Clinical multi-label free text classification by exploiting disease label relation. In: Proceedings - 2013 IEEE International Conference on Bioinformatics and Biomedicine, IEEE BIBM 2013. pp 311–315
23. Read J, Pfahringer B, Holmes G, Frank E (2009) Classifier chains for multi-label classification. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp 254–269
24. Baghdadi Y, Bourrée A, Robert A, et al (2019) Automatic classification of free-text medical causes from death certificates for reactive mortality surveillance in France. Int J Med Inform 131:. https://doi.org/10.1016/j.ijmedinf.2019.06.022
25. Weng W-H, Wagholikar KB, McCray AT, et al (2017) Medical subdomain classification of clinical notes using a machine learning-based natural language processing approach. BMC Med Inform Decis Mak 17:155. https://doi.org/10.1186/s12911-017-0556-8
26. Stephan Spat, Bruno Cadonna, Ivo Rakovac, Christian Gutl, Hubert Leitner, Günther Stark, Thomas R. Pieber PB (2011) Multi-label Classification of Clinical Text Documents considering the Impact of Text Pre-processing and Training size. In: 23rd International Conference of the European Federation for Medical Informatics
27. Lita LV, Yu S, Niculescu S, Bi J (2008) Large Scale Diagnostic Code Classification for Medical Patient Records. IJCNLP 877–882
28. Baumel T, Nassour-Kassis J, Cohen R, et al (2017) Multi-Label Classification of Patient Notes a Case Study on ICD Code Assignment. In: AAAI Conference on Artificial Intelligence. pp 409–416
29. van der Maaten, L. J. P., & Hinton GE (2008) Visualizing High-Dimensional Data Using t-SNE. J Mach Learn Res 9:2579–2605
30. Aronson AR, Lang FM (2010) An overview of MetaMap: Historical perspective and recent advances. J Am Med Informatics Assoc 17:229–236. https://doi.org/10.1136/jamia.2009.002733
31. Krzysztof Dembczynski, Arkadiusz Jachnik, Wojciech Kotłowski, et al (2013) Optimizing the F-measure in multi-label classification: plug-in rule approach versus structured loss minimization. In: ICML'13: Proceedings of the 30th International Conference on International Conference on Machine Learning. pp 1130–1138

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50423-6_38