EXPERIMENTAL RESEARCH

TOWARDS AUTOMATED META-ANALYSIS OF BIOMEDICAL TEXTS IN THE FIELD OF CELL-BASED IMMUNOTHERAPY

D.A. Devyatkin[1*], A.I. Molodchenkov[1], A.V. Lukin[1], Y.S. Kim[2], A.A. Boyko[3], P.A. Karalkin[4], J.-H. Chiang[5], G.D. Volkova[6], A.Yu. Lupatov[2]

[1]Federal Research Centre «Computer Science and Control» RAS,
9 60-let Oktyabrya av., Moscow, 119333 Russia; *e-mail: devyatkin@isa.ru
[2]Institute of Biomedical Chemistry, 10 Pogodinskaya str., Moscow, 119121 Russia
[3]Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, 16/10 Miklukho-Maklaya str., Moscow, 117997 Russia
[4]Hertsen Moscow Oncology Research Center – branch of National Medical Research Radiological Center,
3 2-nd Botkinsky proezd, Moscow, 125284 Russia
[5]National Cheng Kung University, Tainan City, Taiwan
[6]Moscow State Technological University «STANKIN», 1 Vadkovsky lane, Moscow, 127994 Russia

Cell-based immunotherapy is a promising approach for the treatment of chronic infections, autoimmune disorders, and malignant tumors. There are many strategies of cell-based immunotherapy of cancer; these include injection of various immune effector cells, propagated and «trained» in a cell culture. Alternatively, cells presenting tumor antigens on their surface in a form recognized by the immune system can be used to achieve a therapeutic effect. The research results in this field are presented in thousands of texts, and their manual analysis is very complicated. We have developed an approach for automated text analysis in this area of biomedical science. Here we present the first results of the automated analysis of the data extracted from abstracts of scientific articles available in PubMed. These results demonstrate the associations between types of tumors and the most commonly used methods of their cell-based immunotherapy.

Key words: cancer; cell-based immunotherapy; text mining; automated meta-analysis

DOI: 10.18097/BMCRM00109

Abbreviations: ADCC – antibody dependent cell-mediated cytotoxicity; APC – antigen presenting cells; AP B-cells – antigen-presenting B-cells; CAR – chimeric antigen receptor; CIK – cytokine-induced killers; DC – dendritic cells; LAK – lymphokine activated killers; MHC – main histocompatibility complex; NER – named entities; NK-cells – natural killer cells; NLP – natural language processing; TCR – T-cell receptor; TIL – tumor-infiltrating lymphocytes

INTRODUCTION

Cell-based immunotherapy is a rapidly growing area of biomedical science aimed at cell transplantation to achieve a therapeutic effect. Cells exhibiting immunomodulating properties can be used to suppress pathological immune responses to autoantigens, as well as to stimulate immunity to fight against tumor or chronic infections. To treat cancer by administration of immune cells two main approaches are actively investigated in research laboratories and clinics. The first approach employs active immunotherapy with a cell vaccine, which can present tumor antigens to lymphocytes and induce an immune response. It is based on the use of antigen presenting cells (APC) or inactivated tumor cells. The most effective APC are dendritic cells (DC) that can present tumor antigens in the context of their own main histocompatibility complex (MHC) molecules to lymphocytes [1]. The usual procedure of the immunotherapy incudes isolation of monocytes from patients' blood, their *ex vivo* differentiation to DC, loading DC with a tumor antigen followed by DC growth and maturation in vitro, administration of antigen loaded DC to the patient. Alternatively, DC can be fused *in vitro* with tumor cells to produce dendritomas. The second approach is based on a passive (adoptive) immunization [2]. It includes injections of various immune effector cells, usually lymphocytes, capable of lysing tumor cells and isolated from blood or tumor (tumor-infiltrating lymphocytes (TIL)). The lymphocytes must be propagated and «trained» in a culture to achieve the therapeutic effect. To direct lymphocytes against a tumor, various cytokines (lymphokine activated killer (LAK), cytokine-induced killer (CIK)) or a genetic modification of antigen-recognizing receptors (redirected TCR T-cells (reTCR T-cells), chimeric antigen receptor (CAR) T-cells) can be used. Usually any procedure for conducting cell-based immunotherapy includes isolation of immune cells from a patient, their modification and propagation in a cell culture, and subsequent administration back to the patient. However, it is rather difficult to track development and evaluate effectiveness of cell-based immunotherapy methods by analyzing scientific literature. There are a large number of cell populations with immunomodulatory properties and they can be cultured under various conditions to improve their therapeutic characteristics. In addition, the characteristics of the patient and patient's disease can significantly affect outcome of the treatment. The problem is complicated by the fact that there are thousands of texts related to this topic. In this regard, development of a tool

© 2019 by the authors. Licensee IBMC, Moscow, Russia. This article is an open access and distributed under the terms and conditionsof the Creative Commons Attribution (CC BY-SA 4.0) license (http://creativecommons.org/licenses/by-sa/4.0/).

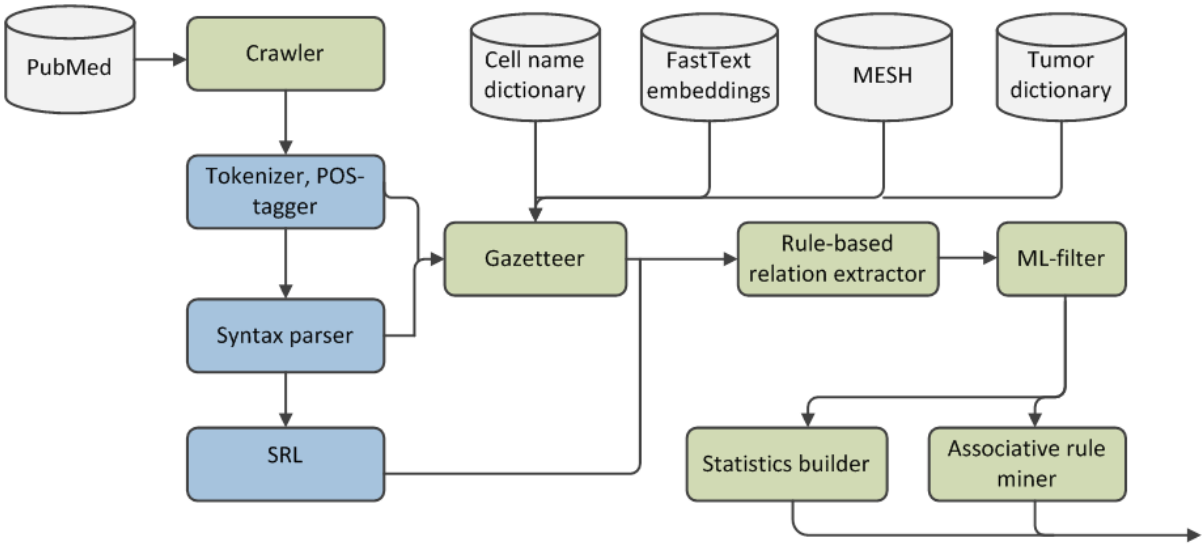

Figure 1. The framework for meta-analysis of biomedical texts.

for automated meta-analysis of biomedical texts in the field of cell-based immunotherapy is a challenge facing cell biology and computer science.

The automated meta-analysis consists in retrieving of required research papers, accurate extraction of particular named entities (NER), revealing their roles and relations between them. Complex linguistic features are usually required to tackle with these problems. Krallinger with colleagues highlighted the natural language processing (NLP) methods to retrieve some of these features [3]. There are specific NLP components for biomedical texts such as POS-taggers [4] and syntax parsers [5], which were trained on specific domain models to return syntactic dependency relations between tokens in a clause. The approaches for bio-NER include dictionary-lookup [6], rule-based [7], and machine-learning-based sequence labeling [8]. Nevertheless, the quality of all these techniques depends on the entity type, its linguistic features and availability of dictionaries. It is worth to note the following supervised machine learning models applied to detect biomedical entities: SVM [9], conditional random field (CRF) [10], and recurrent neural networks [11]. Besides good quality of extraction these models require large corpora to train reliably; however, this represents a serious problem due to limitation in the the existing data for information extraction.

In the case of cell-based therapy, we should extract mentions of specific cell types, their roles in the text and particular types of relationships between them. At the same time, we do not have any labeled corpora applicable to training end-to-end machine learning approaches. Considering this situation, a hybrid approach has been proposed. Our solution consists in applying gazetteers and rule-based analyzer together to detect candidates for the entities and then to filter them with a pre-trained machine learning model. We also consider all available external linguistic resources, such as MESH (https://meshb.nlm.nih.gov/search), pre-trained FastText models [12], cell and tumor dictionaries to tackle with the small corpora problem. After that syntax and semantic features are used to reveal relations between the entities. A similar approach was employed by Shelmanov and colleagues [13].

In this paper, we have proposed a framework, which provides solutions for NLP tasks that arise in meta-analysis: from web-crawling to named entities recognition and estimation of statistic scores. This approach has been applied to identify associations between types of tumors and the most commonly used methods of their cell-based immunotherapy.

METHODS

As a basic method for the analysis, we propose a framework that contains several steps (Fig. 1).

1. Crawling abstracts from Pubmed. We applied Scrapy-based web-crawler with a manually created rule set to download the search output from PubMed [14].
2. Rich linguistic features extraction. In this step the Isanlp framework is applied to obtain morphology, syntax parsing, and semantic role labeling features [15].
3. Combining tumor and cell dictionaries (provided by biologists) and morphology-based rules to extract entity candidates from the abstracts. We use syntax relations and build all their possible combinations. Regular expression-based features were also considered. This is an adaptation of the linguistically motivated approach for mapping terms from medical texts to concepts in UMLS Metathesaurus, which was proposed in https://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/index.html and called MetaMap [7]. We also applied the Fasttext model, pre-trained on biomedical texts, to catch synonyms.
4. Syntax and semantic features analysis. We use syntax relations and semantic roles to reveal links between the entities and their roles (target cell, active cell) in the sentence.
5. Applying a pre-trained sequence-labeling machine learning model to filter uninformative entity candidates. This step has not been implemented yet, now we use dictionary-based filter instead.
6. Coocurence statistics calculating and associative rule mining for the extracted entities [16] to obtain stable combinations of tumors, therapy and cell types. We used the Eclat algorithm [17] because of its scalability.

We also created a lexis only approach for the steps 3–4, which do not consider any syntax or semantic features. This approach is used as a baseline in the experiments on text mining.

Table 1. Categories, queries and counts of the extracted documents

Category	Query	Documents
All	*((((((cancer immunotherapy) OR cancer vaccine) OR ((adoptive immunotherapy) AND cancer)))) NOT (((((cancer immunotherapy) OR cancer vaccine) OR ((adoptive immunotherapy) AND cancer))) AND Review[ptyp]))*	96160
Humans (including clinical trials)	*(((((((cancer immunotherapy) OR cancer vaccine) OR ((adoptive immunotherapy) AND cancer)))) NOT (((((cancer immunotherapy) OR cancer vaccine) OR ((adoptive immunotherapy) AND cancer))) AND Review[ptyp]))) AND Humans[Mesh])*	62448
Clinical trials	*(((((((cancer immunotherapy) R cancer vaccine) OR ((adoptive immunotherapy) AND cancer)))) NOT (((((cancer immunotherapy) OR cancer vaccine) OR ((adoptive immunotherapy) AND cancer))) AND Review[ptyp]))) AND Clinical Trial[ptyp])*	7585
Other animals	*(((((((cancer immunotherapy) OR cancer vaccine) OR ((adoptive immunotherapy) AND cancer)))) NOT (((((cancer immunotherapy) OR cancer vaccine) OR ((adoptive immunotherapy) AND cancer))) AND Review[ptyp]))) AND Animals[Mesh:noexp])*	33712

Table 2. Extracted documents

	Effector or APC	Count of extracted documents			
		Total	Humans	Other animals	Clinical trials
Active immunotherapy (cell vaccine)	Antigen loaded dendritic cells (DC vaccine)	1978	1200	778	298
	Dendritomas	23	13	10	2
	Antigen-presenting B-cells (AP B-cells)	1	1	0	0
	Inactivated tumor cells	61	54	7	23
Passive (adoptive) immunotherapy	Tumor-infiltrating T-cells (TIL)	297	217	80	44
	Chimeric antigen receptor T-cells (CAR-T-cells)	136	92	44	14
	Redirected TCR T-cells (reTCR-T-cells)	11	10	1	0
	Lymphokine activated killers (LAK)	126	92	34	18
	Cytokine-induced killers (CIK)	106	78	28	17
	Antibody dependent cell-mediated cytotoxicity (ADCC)	7	6	1	1
	Chimeric antigen receptor NK-cells (CAR-NK-cells)	1	1	0	0

RESULTS AND DISCUSSION

In the first instance, we extracted 96160 non-labeled abstracts related to cell-based cancer immunotherapy from PubMed (Table 1). The total corpus of the texts did not include annotations of reviews and was divided into three groups for separate analysis. The first group contained experimental research texts on a human material including established cell lines and patients' tissue samples. The second group integrated similar animal studies and preclinical trials. The third group corresponded to clinical trials conducted according to the relevant regulations. Separate analysis was useful in assessing the maturity of immunotherapeutic approaches.

Based on the thesaurus of cell-based immunotherapy methods compiled by specialists and applying proposed framework we extracted the data on various immunotherapy methods from the primary text corpus (Table 2). The data demonstrate that DC vaccines are the most common method of immunotherapy for both basic researchers and clinical trials. Despite rather low efficiency [1], high impact of DC vaccines is apparently due to the simplicity of their preparation and the absence of side effects in patients. Obviously, the efforts of researchers in this direction is mainly aimed at increasing the effectiveness of DC vaccination. Dynamic analysis shows that although there is growth in the number of studies on DC vaccines (Fig. 2), their percentage among other methods is gradually decreasing (Fig. 3). More and more methods of passive immunotherapy, including genetically modified effector cells, are entering the stage of clinical trials. Interestingly, cell-based immunotherapy approaches with a long history, such as TIL or CIK, are still involved in clinical trials. Perhaps this is because of the recent opportunity to significantly increase the effectiveness of passive cell-based immunotherapy by applying lymphodepleting chemotherapy before the cell administration [18]. In addition, the development of methods for propagating *in vitro* antigen-specific lymphocyte clones has improved the clinical outcome of traditional approaches [19].

An important part of the study is the identification of associative relationships between cell-based immunotherapy methods and the tumors for which the treatment is intended. We applied a dictionary of tumor names to both the whole corpus of documents and the clinical trial documents subset. DC vaccines is mentioned by researchers in the context of almost all of tumor types (Fig. 4). In clinical studies, DC loaded with tumor antigens

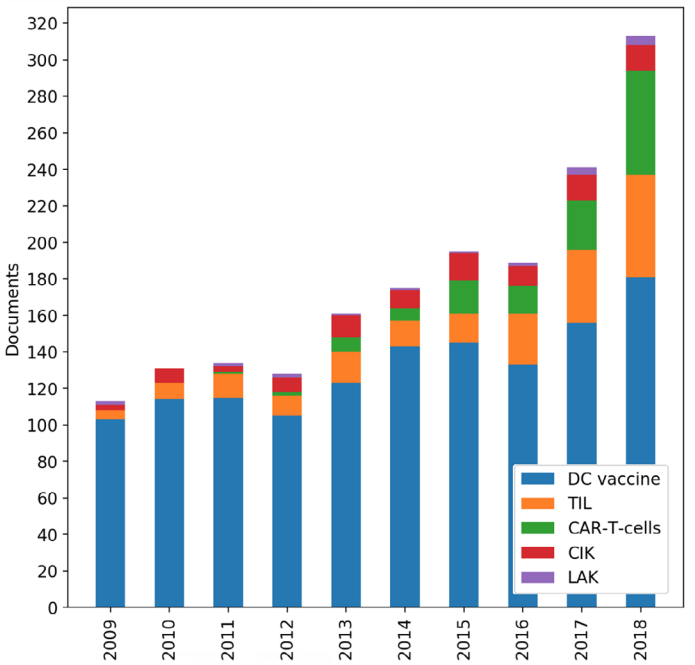

Figure 2. Dynamic distribution of the documents devoted to particular cell-based immunotherapies.

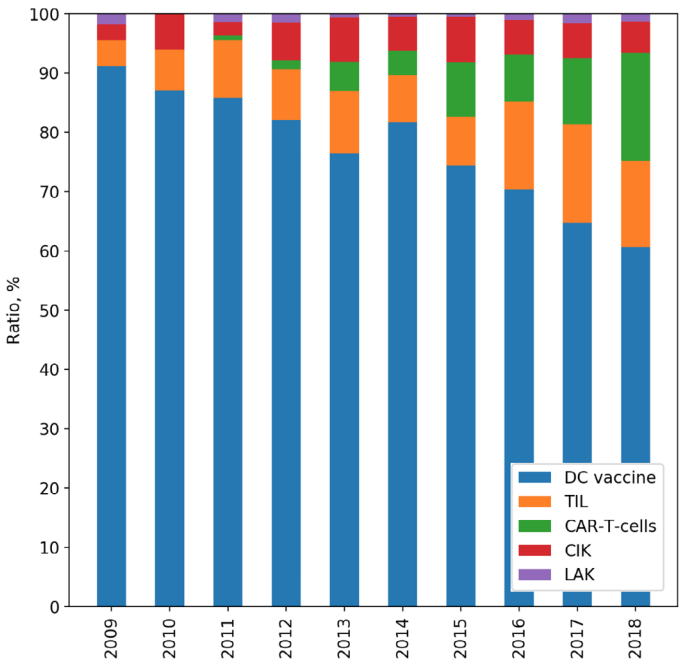

Figure 3. The ratios of the documents devoted to particular cell-based immunotherapies.

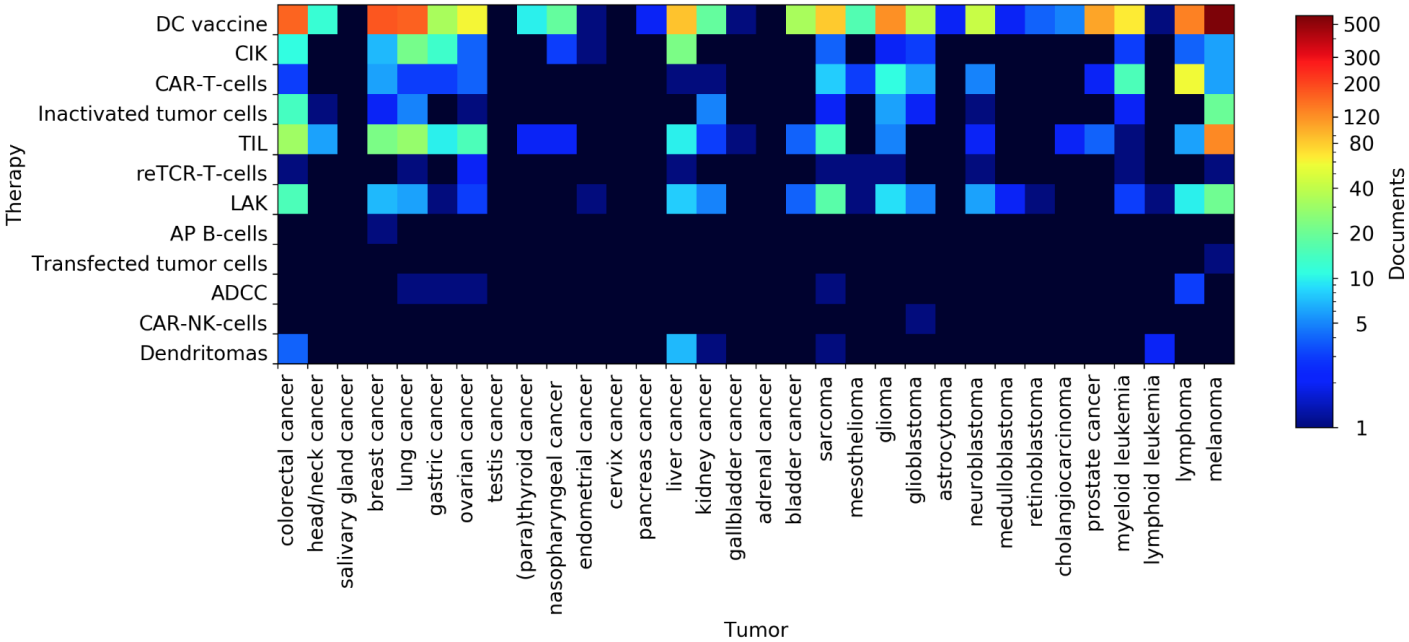

Figure 4. Distribution map of the documents, devoted to particular therapies and tumors (all considered documents).

are most commonly used to treat melanoma, prostate cancer, brain tumors, colorectal cancer, lung cancer, breast cancer, and lymphomas (Fig. 5). Quite interesting data are related to the immunotherapy with CAR-T-cells. A feature of such effector cells is that their chimeric antigen-recognizing receptor interacts with a tumor antigen that is not associated with MHC on the tumor surface. Since the presence of such kind of antigens is mostly a characteristic feauture of hematopoietic neoplasms, B-cell lymphomas expressing CD19 antigen have always been a common target for CAR-T-cells [20]. Our analysis has shown that lymphomas are the main target for CAR-T-cells based therapy (Fig. 5). However, there are many mentions of this effector in the context of solid tumors (Fig. 4) and even clinical trials, mainly aimed at treating sarcomas and glioblastomas (Fig. 5). It indicates the development of CAR-T-cells based therapy towards the treatment of solid tumors. Table 3 presents the most confident associations between the methods of immunotherapy and tumor types mined from analyzed abstracts.

To evaluate the effectiveness of our approach to automated meta-analysis we have tested the proposed framework against baseline approach on a small corpus, which contains 16 manually labeled abstracts from PubMed Central (Table 4). Through the experiment, we estimated precision (P), recall (R) and F1 scores as the most important characteristics [21]. The results show that the proposed framework significantly over performs the baseline approach, especially in terms of precision. It means that syntax and semantic features are crucial for this task. However, the lower recall scores display that an additional extension of the dictionaries is still needed.

The main feature of our solution is considering all available

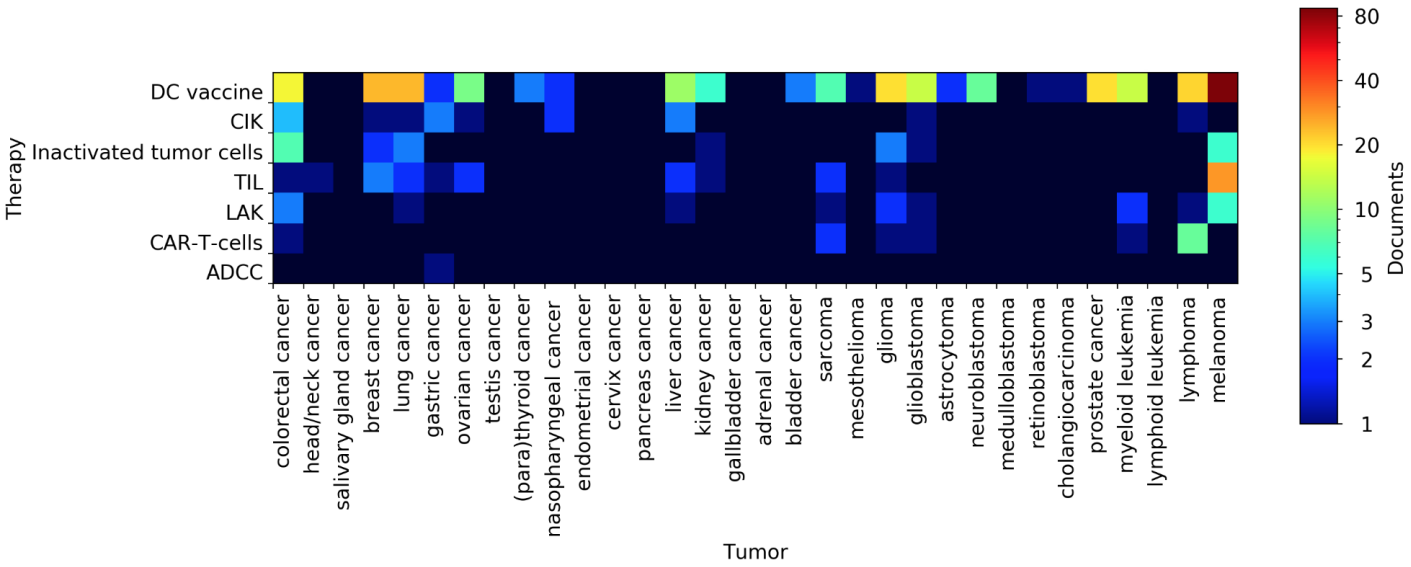

Figure 5. Distribution map of the documents, devoted to particular therapies and tumors (clinical trials).

Table 3. Associative patterns for cell-based immunotherapy of cancer

Tumor	Therapy	Effector	Conf
colorectal cancer	Active immunotheray (cell vaccine)	Antigen loaded dendritic cells (DC vaccine)	0.76
melanoma			0.70
lung cancer			0.64
breast cancer			0.53
sarcoma			0.54
	Passive (adoptive) immunotherapy	Chimeric antigen receptor T-cells (CAR-T-cells)	0.36
lymphoma			0.69
lung cancer		Cytokine-induced killers (CIK)	0.36

Table 4. Evaluation results for the proposed framework

	Proposed framework			Baseline		
Module	P*	R**	F1-score	P*	R**	F1-score
Disease identification	0.95	0.73	0.82	0.9	0.42	0.60
Cell name extraction	0.96	0.52	0.67	0.17	0.54	0.25
Cell role identification	0.50	0.77	0.61	0.38	0.77	0.25
Tumor name identification	0.95	0.38	0.55	0.13	0.40	0.19

Note: *precision; **recall.

external linguistic resources, such as cell and tumor dictionaries, pre-trained FastText embeddings and ontologies, together with the results of full linguistic analysis. This allows significantly reducing demand on the size of training corpora. In contrast, the recent deep neural-network-based language models, like a BERT, are usually pre-trained to extract complex features in an unsupervised manner and then fine-tuned in large labeled corpora to solve a particular problem; therefore external resources are hardly considered. Although these models demonstrate astonishing results in the biomedical domain [22], the demand on the size of labeled data remains high for these models. For example, it is demonstrated that about 2 000 texts is required to train such model even with transfer learning [23]. Unfortunately, the existing large corpora often do not meet the aim of a particular meta-analysis; they usually focus on general tasks. Hence, this approach is not applicable in our case.

Another feature of the proposed approach is that it provides an end-to-end solution for the meta-analysis problem. In spite of plenty of studies, which are related to particular information extraction tasks, there is a lack of full information extraction frameworks for biomedical purposes. It is worth noticing that the only similar framework has been proposed by Hakala with colleagues [24].

Thus, in this article, we have shown that the proposed framework can be used as a tool for automated meta-analysis of biomedical texts in the field of cell-based immunotherapy. Moreover, if the modules for assessment of clinical outcomes [25, 26] were added to this framework, it would be useful for the generalizing meta-analysis of clinical trials effectiveness of cell-based immunotherapy.

COMPLIANCE WITH ETHICAL STANDARDS

We did not use biological samples from humans or other animals in this work. The analysis of clinical trial did not include patient personal data.

FUNDING

This study was supported by the Russian Foundation for Basic Research, grants no. 16-29-07246 and 16-29-07210.

REFERENCES

1. *Palucka, K., & Banchereau, J.* (2013). Dendritic-cell-based therapeutic cancer vaccines. Immunity, **39**(1), 38-48. DOI: 10.1016/j.immuni.2013.07.004
2. *Lupatov, A. Yu., Karalkin, P. A., Boyko, A. A., & Yarygin, K. N.* (2018). Autotransplantation of T-lymphocytes as a tool for antigen-specific immunotherapy of oncological diseases. Vestnik Transplantologii i Iskusstvennykh Organov, **20**(3), 95-104. DOI: 10.15825/1995-1191-2018-3-95-104
3. *Krallinger, M., Rabal, O., Lourenço, A., Oyarzabal, J., & Valencia, A.* (2017). Information retrieval and text mining technologies for chemistry. Chemicalreviews, **117**(12), 7673-7761. DOI: 10.1021/acs.chemrev.6b00851
4. *Tsuruoka, Y., Tateishi, Y., Kim, J.-D., Ohta, T., McNaught, J., Ananiadou, S., & Tsujii, J.* (2005). Developing a robust part-of-speech tagger for biomedical text. Advances in Informatics, **3746**, 382−392.
5. *Miyao, Y., & Tsujii, J.* (2008). Feature forest models for probabilistic HPSG parsing. computational linguistics, **34**(1), 35−80. DOI:10.1162/coli.2008.34.1.35
6. *Hina, S., Atwell, E., & Johnson, O.* (2010). Secure information extraction from clinical documents using snomed ct gazetteer and natural language processing . International conference for internet technology and secured transactions. IEEE, 1-5.
7. *Aronson, A. R., & Lang, F. M.* (2010). An overview of MetaMap: historical perspective and recent advances. Journal of the American Medical Informatics Association, **17**(3), 229–236. DOI:10.1136/jamia.2009.002733
8. *Jagannatha, A. N., & Yu, H.* (2016). Structured prediction models for RNN based sequence labeling in clinical text. Proceedings of the conference on empirical methods in natural language processing. NIH Public Access, **2016**, 856–865.
9. *Mika, S., & Rost, B.* (2004). Protein names precisely peeled off free text. Bioinformatics, **20**(1), i241−i247. DOI:10.1093/bioinformatics/bth904
10. *McDonald, R., & Pereira, F.* (2005). Identifying gene and protein mentions in text using conditional random fields. BMC Bioinformatics, **6**(1), S6. DOI: 10.1186/1471-2105-6-S1-S6
11. *Zeng, D., Sun, D., Lin, L., & Liu, B.* (2017). LSTM-CRF for drug-named entity recognition. Entropy, **19** (6), 283. DOI: 10.3390/e19060283
12. *Wang, Y., Liu, S., Afzal, N., Rastegar-Mojarad, M., Wang, L., Shen, F., Kingsbury, P., & Liu, H.* (2018). A comparison of word embeddings for the biomedical natural language processing. Journal of Biomedical Informatics, **87**, 12-20. DOI: 10.1016/j.jbi.2018.09.008
13. *Shelmanov, A. O., Smirnov, I. V., & Vishneva, E. A.* (2015). Information extraction from clinical texts in Russian. Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue", **14**(21), 537-549.
14. Yadav, *M., & Goyal, N.* (2015). Comparison of open source crawlers – a review. International Journal of Scientific and Engineering Research, **2229**(5518), 1544-1551.
15. *Larionov, D., Shelmanov, A., Chistova, E., & Smirnov, I.* (2019). Semantic role labeling with pretrained language models for known and unknown predicates. Proceedings of Recent Advances of Natural Language Processing, 620-630. https://github.com/IINemo/isanlp
16. *Agrawal, R., & Srikant, R.* (1994). Fast algorithms for mining association rules. Proceedings of the 20th Very Large Data Base Conference, 487-499.
17. *Zaki, M. J.* (2000). Scalable algorithms for association mining. IEEE Transactions on Knowledge and Data Engineering, **12**(3), 372–390. DOI:10.1109/69.846291
18. *Rosenberg, S. A., Yang, J. C., Sherry, R. M., Kammula, U. S., Hughes, M. S., Phan, G. Q., & Dudley M. E.* (2011). Durable complete responses in heavily pretreated patients with metastatic melanoma using T-cell transfer immunotherapy. Clinical Cancer Research, **17**(13), 4550-4557. DOI: 10.1158/1078-0432.CCR-11-0116
19. *Radvanyi, L.G., Bernatchez, C., Zhang, M., Fox, P.S., Miller, P., Chacon, J., & Hwu, P.* (2012). Specific lymphocyte subsets predict response to adoptive cell therapy using expanded autologous tumor-infiltrating lymphocytes in metastatic melanoma patients. Clinical Cancer Research, **18**(24), 6758-6770. DOI: 10.1158/1078-0432.CCR-12-1177
20. *Kochenderfer, J. N., Wilson, W. H., Janik, J. E., Dudley, M. E., Stetler-Stevenson, M., Feldman, S. A, & Rosenberg, S. A.* (2010). Eradication of B-lineage cells and regression of lymphoma in a patient treated with autologous T cells genetically engineered to recognize CD19. Blood, **116**(20), 4099-4102. DOI: 10.1182/blood-2010-04-281931
21. *Flach, P.* (2012). Machine learning: the art and science of algorithms that make sense of data. Book, Cambridge University Press.
22. *Lin, C., Miller, T., Dligach, D., Bethard, S., & Savova, G.* (2019) A BERT-based universal model for both within-and cross-sentence clinical temporal relation extraction. Proceedings of the 2nd Clinical Natural Language Processing Workshop, 65-71. https://www.aclweb.org/ anthology/W19-1908
23. *Pang, N., Qianm L., Lyu, W., & Yang, J-D.* (2019) Transfer learning for scientific data chain extraction in small chemical corpus with BERT-CRF model. arXiv preprint arXiv:1905.05615
24. *Hakala, K., Kaewphan, S., Salakoski, T., & Ginter, F.* (2016) Syntactic analyses and named entity recognition for PubMed and PubMed Central—up-to-the-minute. Proceedings of the 15th Workshop on Biomedical Natural Language Processing, 102-107.
25. *Lupatov, A. Y., Yarygin, K. N., Panov, A. I., Suvorov, R. E., Shvets, A. V., Volkova, G. D.* (2015). Assessment of dendritic cell therapy effectiveness based on the feature extraction from scientific publications. Proceedings of the International Conference on Pattern Recognition Applications and Methods (ICPRAM), **2**, 270-276. DOI:10.5220/0005248802700276
26. *Boyko, A. A., Kaidina, A. M., Kim, Y. C., Lupatov, A. Yu., Panov, A. I., Suvorov, R. E., Shvets, A. V.* (2016). A framework for automated meta-analysis: dendritic cell therapy case study. 8th International Conference on Intelligent Systems (IEEE), **8**, 160-166. DOI: 10.1109/IS.2016.7737416

Received: 20.07.2019
Revised 30.08.2019
Accepted: 16.09.2019

НА ПУТИ К АВТОМАТИЗИРОВАННОМУ МЕТА-АНАЛИЗУ БИОМЕДИЦИНСКИХ ТЕКСТОВ В ОБЛАСТИ КЛЕТОЧНОЙ ИММУНОТЕРАПИИ

Д. А. Девяткин[1*], А.И. Молодченков[1], А.В. Лукин[1], Я.С. Ким[2], А.А. Бойко[3], П.А. Каралкин[4], J.-H. Chiang[5], Г.Д. Волкова[6], А.Ю. Лупатов[2]

[1]Федеральный исследовательский центр «Информатика и управление» РАН,
119333, Москва, пр. 60-лет Октября, 9; *e-mail: devyatkin@isa.ru
[2]Институт биомедицинской химии, 119121, Москва, ул. Погодинская, 10
[3]Институт биоорганической химии им. Шемякина и Овчинникова, 117997, Москва, ул. Миклухо-Маклая, 16/10
[4]Московский научно-исследовательский онкологический институт имени П.А. Герцена –
филиал «НМИЦ радиологии» Минздрава, 125284, Москва, 2-й Боткинский проезд, 3
[5]National Cheng Kung University, Tainan City, Taiwan
[6]Московский государственный технологический университет «СТАНКИН», 127994, Москва, Вадковский пер., 1

Клеточная иммунотерапия это перспективный подход к лечению хронических инфекций, аутоиммунных нарушений и злокачественных опухолей. Существует множество стратегий иммунотерапии рака, включая инъекции различных иммунных эффекторных клеток, размноженных и «обученных» в клеточной культуре. В качестве альтернативы для достижения терапевтического эффекта могут быть использованы клетки, представляющие опухолевый антиген на своей поверхности в «понятном» для иммунной системы виде. Результаты исследований в этой области представлены в тысячах текстов, ручной анализ которых затруднен. Мы разработали подход для автоматического анализа текстов в этой области биомедицинской науки. В данной работе мы представляем первые результаты автоматического анализа данных, извлеченных из абстрактов научных статей, доступных в PubMed. На корпусе извлеченных текстов мы демонстрируем ассоциации между типами опухолей и наиболее часто используемыми способами клеточной терапии.

Ключевые слова: злокачественные опухоли, клеточная иммунотерапия, анализ текстов, автоматизированный мета-анализ

ФИНАНСИРОВАНИЕ

Работа выполнена при финансовой поддержке РФФИ, гранты16-29-07246 и 16-29-07210.

Поступила: 20.07.2019, после доработки: 30.08.2019, принята к публикации: 16.09.2019