Proceedings - 2017 Ivannikov ISPRAS Open Conference, ISPRAS 2017, 2018, vol.2018-January, pages 64-69

A machine learning approach to classification of drug reviews in Russian

Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

© 2017 IEEE. The automatic extraction of drug side effects from social media has gained popularity in pharmacovigilance. Information extraction methods tailored to medical subjects are essential for the task of drug repurposing and finding drug reactions. In this article, we focus on extracting information about side effects and symptoms in users' reviews about medications in Russian. We manually develop a real-world dataset by crawling user reviews from a health-related website and annotate a set of reviews on a sentence level. The paper addresses the classification problem with more than two classes, comparing a simple bag-of-words baseline and a feature-rich machine learning approach.

http://dx.doi.org/10.1109/ISPRAS.2017.00018

Keywords

Drug side effects, Machine learning, Russian, Social media, Text mining

References

[1] T. Oprea and J. Mestres, "Drug repurposing: far beyond new targets for old drugs, " The AAPS journal, vol. 14, no. 4, pp. 759-763, 2012.

[2] J. L. Medina-Franco, M. A. Giulianotti, G. S. Welmaker, and R. A. Houghten, "Shifting from the single to the multitarget paradigm in drug discovery, " Drug discovery today, vol. 18, no. 9, pp. 495-501, 2013.

[3] H.-M. Lee and Y. Kim, "Drug repurposing is a new opportunity for developing drugs against neuropsychiatric disorders, " Schizophrenia research and treatment, vol. 2016, 2016.

[4] G. Jin and S. T. Wong, "Toward better drug repositioning: prioritizing and integrating existing methods into efficient pipelines, " Drug discovery today, vol. 19, no. 5, pp. 637-644, 2014.

[5] R. Sloane, O. Osanlou, D. Lewis, D. Bollegala, S. Maskell, and M. Pirmohamed, "Social media and pharmacovigilance: a review of the opportunities and challenges, " British journal of clinical pharmacology, vol. 80, no. 4, pp. 910-920, 2015.

[6] A. Benton, L. Ungar, S. Hill, S. Hennessy, J. Mao, A. Chung, C. E. Leonard, and J. H. Holmes, "Identifying potential adverse effects using the web: A new approach to medical hypothesis generation, " Journal of biomedical informatics, vol. 44, no. 6, pp. 989-996, 2011.

[7] C. C. Yang, H. Yang, L. Jiang, and M. Zhang, "Social media mining for drug safety signal detection, " in Proceedings of the 2012 international workshop on Smart health and wellbeing. ACM, 2012, pp. 33-40.

[8] X. Liu and H. Chen, "Azdrugminer: an information extraction system for mining patient-reported adverse drug events in online patient forums, " in International Conference on Smart Health. Springer, 2013, pp. 134-150.

[9] K. O'Connor, P. Pimpalkhute, A. Nikfarjam, R. Ginn, K. L. Smith, and G. Gonzalez, "Pharmacovigilance on twitter? mining tweets for adverse drug reactions, " in AMIA annual symposium proceedings, vol. 2014. American Medical Informatics Association, 2014, p. 924.

[10] C. C. Yang, H. Yang, and L. Jiang, "Postmarketing drug safety surveillance using publicly available health-consumer-contributed content in social media, " ACM Transactions on Management Information Systems (TMIS), vol. 5, no. 1, p. 2, 2014.

[11] E. Aramaki, Y. Miura, M. Tonoike, T. Ohkuma, H. Masuichi, K. Waki, and K. Ohe, "Extraction of adverse drug effects from clinical records." in MedInfo, 2010, pp. 739-743.

[12] M. Rastegar-Mojarad, R. K. Elayavilli, Y. Yu, and H. Liu, "Detecting signals in noisy data-can ensemble classifiers help identify adverse drug reaction in tweets, " in Proceedings of the Social Media Mining Shared Task Workshop at the Pacific Symposium on Biocomputing, 2016.

[13] T. Huynh, Y. He, A. Willis, and S. Rüger, "Adverse drug reaction classification with deep neural networks, " in Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, 2016, pp. 877-887.

[14] R. Leaman, L. Wojtulewicz, R. Sullivan, A. Skariah, J. Yang, and G. Gonzalez, "Towards internet-age pharmacovigilance: extracting adverse drug reactions from user posts to health-related social networks, " in Proceedings of the 2010 workshop on biomedical natural language processing. Association for Computational Linguistics, 2010, pp. 117-125.

[15] A. Cassels, M. A. Hughes, C. Cole, B. Mintzes, J. Lexchin, and J. P. McCormack, "Drugs in the news: an analysis of canadian newspaper coverage of new prescription drugs, " Canadian Medical Association Journal, vol. 168, no. 9, pp. 1133-1137, 2003.

[16] R. Harpaz, A. Callahan, S. Tamang, Y. Low, D. Odgers, S. Finlayson, K. Jung, P. LePendu, and N. H. Shah, "Text mining for adverse drug events: the promise, challenges, and state of the art, " Drug safety, vol. 37, no. 10, pp. 777-790, 2014.

[17] A. Sarker, R. Ginn, A. Nikfarjam, K. O'Connor, K. Smith, S. Jayaraman, T. Upadhaya, and G. Gonzalez, "Utilizing social media data for pharmacovigilance: A review, " Journal of biomedical informatics, vol. 54, pp. 202-212, 2015.

[18] J. Lardon, R. Abdellaoui, F. Bellet, H. Asfari, J. Souvignet, N. Texier, M.-C. Jaulent, M.-N. Beyens, A. Burgun, and C. Bousquet, "Adverse drug reaction identification and extraction in social media: a scoping review, " Journal of medical Internet research, vol. 17, no. 7, 2015.

[19] S. Yeleswarapu, A. Rao, T. Joseph, V. G. Saipradeep, and R. Srinivasan, "A pipeline to extract drug-adverse event pairs from multiple data sources, " BMC medical informatics and decision making, vol. 14, no. 1, p. 13, 2014.

[20] C. C. Freifeld, J. S. Brownstein, C. M. Menone, W. Bao, R. Filice, T. Kass-Hout, and N. Dasgupta, "Digital drug safety surveillance: monitoring pharmaceutical products in twitter, " Drug safety, vol. 37, no. 5, pp. 343-350, 2014.

[21] A. Nikfarjam and G. H. Gonzalez, "Pattern mining for extraction of mentions of adverse drug reactions from user comments, " in AMIA Annual Symposium Proceedings, vol. 2011. American Medical Informatics Association, 2011, p. 1019.

[22] J.-C. Na, W. Y. M. Kyaing, C. S. Khoo, S. Foo, Y.-K. Chang, and Y.-L. Theng, "Sentiment classification of drug reviews using a rulebased linguistic approach, " in International Conference on Asian Digital Libraries. Springer, 2012, pp. 189-198.

[23] Y. Niu, X. Zhu, J. Li, and G. Hirst, "Analysis of polarity information in medical text, " in AMIA annual symposium proceedings, vol. 2005. American Medical Informatics Association, 2005, p. 570.

[24] B. W. Chee, R. Berlin, and B. Schatz, "Predicting adverse drug events from personal health messages, " in AMIA Annual Symposium Proceedings, vol. 2011. American Medical Informatics Association, 2011, p. 217.

[25] J. Bian, U. Topaloglu, and F. Yu, "Towards large-scale twitter mining for drug-related adverse events, " in Proceedings of the 2012 international workshop on Smart health and wellbeing. ACM, 2012, pp. 25-32.

[26] M. Yang, X.Wang, and M. Y. Kiang, "Identification of consumer adverse drug reaction messages on social media." in PACIS, 2013, p. 193.

[27] I. Alimova and E. Tutubalina, "Automated detection of adverse drug reactions from social media posts with machine learning, " in Proceedings of International Conference on Analysis of Images, Social Networks and Texts, 2017, pp. 1-12.

[28] Z. S. Miftahutdinov, E. Tutubalina, and A. Tropsha, "Identifying diseaserelated expressions in reviews using conditional random fields, " in Proceedings of International Conference on Computational Linguistics and Intellectual Technologies Dialog, vol. 1, 2017, pp. 155-167.

[29] A. Sarker and G. Gonzalez, "Portable automatic text classification for adverse drug reaction detection via multi-corpus training, " Journal of biomedical informatics, vol. 53, pp. 196-207, 2015.

[30] E. Tutubalina and S. Nikolenko, "Combination of deep recurrent neural networks and conditional random fields for extracting adverse drug reactions from user reviews, " Journal of Healthcare Engineering, vol. 2017, 2017.

[31] Z. Miftakhutdinov and E. Tutubalina, "Kfu at clef ehealth 2017 task 1: Icd-10 coding of english death certificates with recurrent neural networks." CLEF, 2017.

[32] E. Tutubalina and S. Nikolenko, "Automated prediction of demographic information from medical user reviews, " in International Conference on Mining Intelligence and Knowledge Exploration. Springer, 2016, pp. 174-184.

[33] -, "Exploring convolutional neural networks and topic models for user profiling from drug reviews, " Multimedia Tools and Applications, Nov 2017. [Online]. Available: https://doi.org/10.1007/s11042-017-5336-z

[34] -, "Demographic prediction based on user reviews about medications, " Computación y Sistemas, vol. 21, no. 2, 2017.

[35] S. Karimi, A. Metke-Jimenez, M. Kemp, and C. Wang, "Cadec: A corpus of adverse drug event annotations, " Journal of biomedical informatics, vol. 55, pp. 73-81, 2015.

[36] R. Ginn, P. Pimpalkhute, A. Nikfarjam, A. Patki, K. O'Connor, A. Sarker, K. Smith, and G. Gonzalez, "Mining twitter for adverse drug reaction mentions: a corpus and classification benchmark, " in Proceedings of the fourth workshop on building and evaluating resources for health and biomedical text processing, 2014.

[37] A. Yates and N. Goharian, "Adrtrace: detecting expected and unexpected adverse drug reactions from user reviews on social media sites, " in European Conference on Information Retrieval. Springer, 2013, pp. 816-819.

[38] A. Shelmanov, I. Smirnov, and E. Vishneva, "Information extraction from clinical texts in russian, " in Computational Linguistics and Intellectual Technologies: Annual International Conference "Dialog", vol. 17, 2015.

[39] V. Solovyev and V. Ivanov, "Knowledge-driven event extraction in Russian: corpus-based linguistic resources, " Computational intelligence and neuroscience, vol. 2016, p. 16, 2016.

[40] M. V. Lapayev, A. B. Vodyakho, Aleksandr Ivanovichand Smirnov, and N. A. Zhukova, "Sistema obrabotki tekstovykh meditsinskikh dannykh."

[41] Y. D. Izotova and D. S. Tarasov, "Izvlecheniye farmatsevticheski znachimykh aspektnykh terminov model'yu rekurrentnoy neyronnoy seti iz tekstov na yestestvennom yazyke pri malykh vyborkakh, " in NEYROINFORMATIKA-2016, 2016, pp. 65-73.

[42] D. Y. Turdakov, N. A. Astrakhantsev, Y. R. Nedumov, A. A. Sysoev, I. A. Andrianov, V. D. Mayorov, D. G. Fedorenko, A. V. Korshunov, and S. D. Kuznetsov, "Texterra: A framework for text analysis, " Programming and Computer Software, vol. 40, no. 5, pp. 288-295, 2014.

[43] A. Kutuzov and E. Kuzmenko, WebVectors: A Toolkit for Building Web Interfaces for Vector Semantic Models. Cham: Springer International Publishing, 2017, pp. 155-161. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-52920-2-15

[44] N. V. Loukachevitch and A. Levchik, "Creating a general russian sentiment lexicon." in LREC, 2016.

[45] S. Kiritchenko, X. Zhu, and S. M. Mohammad, "Sentiment analysis of short informal texts, " Journal of Artificial Intelligence Research, vol. 50, pp. 723-762, 2014.