IEEE 30th Jubilee Neumann Colloquium, NC 2017, 2018, vol.2018-January, pages 47-52

End-to-end deep framework for disease named entity recognition using social media data

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

Abstract

© 2017 IEEE. A growing interest in the natural language processing methods applied to healthcare applications has been observed in the recent years. In particular, new drug pharmacological properties can be derived patient observations shared in social media forums. Developing approaches designed to automatically retrieve this information is of no low interest for personalized medicine and wide-scale drug tests. The full potential of the effective exploitation of both textual data and published biological data for drug research often goes untapped mostly because of the lack of tools and focused methodologies to curate and integrate the data and transform it into new, experimentally testable hypotheses. Deep learning architectures have shown promising results for a wide range of tasks. In this work, we propose to address a challenging problem by applying modern deep neural networks for disease named entity recognition. An essential step for this task is recognition of disease mentions and medical concept nor-malization, which is highly difficult with simple string matching approaches. We cast the task as an end-to-end problem, solved using two architectures based on recurrent neural networks and pre-trained word embeddings. We show that it is possible to assess the practicability of using social media data to extract representative medical concepts for pharmacovigilance or drug repurposing.

http://dx.doi.org/10.1109/NC.2017.8263281

Keywords

deep learning, disease named entity extraction, disease named entity normalization, healthcare, Medical systems, recurrent neural networks

References

[1] A. Kotov, "Social media analytics for healthcare," pp. 309-340, 2015. [Online]. Available: http://www.crcnetbase.com/doi/abs/10.1201/ b18588-11

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

[3] 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.

[4] Z. 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.

[5] S.-P. Choi, S. Lee, H. Jung, and S.-k. Song, "An intensive case study on kernel-based relation extraction," Multimedia Tools and Applications, vol. 71, no. 2, Jul 2014. [Online]. Available: https://doi.org/10.1007/s11042-013-1380-5

[6] 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.

[7] W.-Y. S. Chou, Y. M. Hunt, E. B. Beckjord, R. P. Moser, and B. W. Hesse, "Social media use in the united states: implications for health communication," Journal of medical Internet research, vol. 11, no. 4, 2009.

[8] 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.

[9] H. Sharif, F. Zaffar, A. Abbasi, and D. Zimbra, "Detecting adverse drug reactions using a sentiment classification framework," 2014.

[10] D. Yalamanchi, "Sideffective-system to mine patient reviews: sentiment analysis," Ph.D. dissertation, Rutgers University-Graduate School-New Brunswick, 2011.

[11] E. Cambria, T. Benson, C. Eckl, and A. Hussain, "Sentic proms: Application of sentic computing to the development of a novel unified framework for measuring health-care quality," Expert Systems with Applications, vol. 39, no. 12, pp. 10 533-10 543, 2012.

[12] Y. Deng, M. Stoehr, and K. Denecke, "Retrieving attitudes: Sentiment analysis from clinical narratives." in MedIR@ SIGIR, 2014, pp. 12-15.

[13] G. C. M. D. C. Harman, "Quantifying mental health signals in twitter," ACL 2014, vol. 51, 2014.

[14] D. Preotiuc-Pietro, J. Eichstaedt, G. Park, M. Sap, L. Smith, V. Tobolsky, H. A. Schwartz, and L. Ungar, "The role of personality, age and gender in tweeting about mental illnesses," in NAACL HLT, vol. 2015, 2015, p. 21.

[15] T.-T. Dang and T.-B. Ho, "Mixture of language models utilization in score-based sentiment classification on clinical narratives," in International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems. Springer, 2016, pp. 255-268.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] A. I. and T. E., "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.

[24] T. Huynh, Y. He, A. Willis, and S. Ruger, "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.

[25] R. Leaman, C.-H. Wei, and Z. Lu, "tmchem: a high performance approach for chemical named entity recognition and normalization," Journal of cheminformatics, vol. 7, no. 1, p. S3, 2015.

[26] Y. Tsuruoka, J. McNaught, J. c. Tsujii, and S. Ananiadou, "Learning string similarity measures for gene/protein name dictionary look-up using logistic regression," Bioinformatics, vol. 23, no. 20, pp. 2768-2774, 2007.

[27] N. Limsopatham and N. Collier, "Normalising medical concepts in social media texts by learning semantic representation." in ACL (1), 2016.

[28] A. Sarker, D. Molla-Aliod, C. Paris et al., "Outcome polarity identification of medical papers," 2011.

[29] A. Coulter and J. Ellins, "The quality enhancing interventions project: patient-focused interventions," London: The Health Foundation, 2006.

[30] L. Xia, A. L. Gentile, J. Munro, and J. Iria, "Improving patient opinion mining through multi-step classification." in TSD, vol. 5729. Springer, 2009, pp. 70-76.

[31] D. Z. Adams, R. Gruss, and A. S. Abrahams, "Automated discovery of safety and efficacy concerns for joint & muscle pain relief treatments from online reviews," International Journal of Medical Informatics, vol. 100, pp. 108-120, 2017.

[32] M. d. P. Salas-Zarate, J. Medina-Moreira, K. Lagos-Ortiz, H. Luna-Aveiga, M. A. Rodriguez-Garcia, and R. Valencia-Garcia, "Sentiment analysis on tweets about diabetes: An aspect-level approach," Computational and mathematical methods in medicine, vol. 2017, 2017.

[33] N. Ofek, C. Caragea, L. Rokach, P. Biyani, P. Mitra, J. Yen, K. Portier, and G. Greer, "Improving sentiment analysis in an online cancer survivor community using dynamic sentiment lexicon," in Social Intelligence and Technology (SOCIETY), 2013 International Conference on. IEEE, 2013, pp. 109-113.

[34] P. Biyani, C. Caragea, P. Mitra, C. Zhou, J. Yen, G. E. Greer, and K. Portier, "Co-training over domain-independent and domain-dependent features for sentiment analysis of an online cancer support community," in Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. ACM, 2013, pp. 413-417.

[35] R. G. Rodrigues, R. M. das Dores, C. G. Camilo-Junior, and T. C. Rosa, "Sentihealth-cancer: a sentiment analysis tool to help detecting mood of patients in online social networks," International journal of medical informatics, vol. 85, no. 1, pp. 80-95, 2016.

[36] 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.

[37] S. Nikolenko and E. Tutubalina, "Demographic prediction based on user reviews about medications," Computacion y Sistemas, vol. 21, no. 2, pp. 227-241, 2017.

[38] M. Conway and D. OConnor, "Social media, big data, and mental health: current advances and ethical implications," Current opinion in psychology, vol. 9, pp. 77-82, 2016.

[39] A. Benton, M. Mitchell, and D. Hovy, "Multitask learning for mental health conditions with limited social media data," in Proceedings of the 15th Conference of the EACL, vol. 1, 2017, pp. 152-162.

[40] G. Koscielny, P. An, D. Carvalho-Silva, J. A. Cham, L. Fumis, R. Gasparyan, S. Hasan, N. Karamanis, M. Maguire, E. Papa et al., "Open targets: a platform for therapeutic target identification and validation," Nucleic acids research, vol. 45, no. D1, pp. D985-D994, 2016.

[41] J. Lafferty, A. McCallum, F. Pereira et al., "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in Proceedings of the eighteenth international conference on machine learning, ICML, vol. 1, 2001, pp. 282-289.

[42] J. L. Elman, "Finding structure in time," Cognitive science, vol. 14, no. 2, pp. 179-211, 1990.

[43] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM networks," in Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on, vol. 4. IEEE, 2005, pp. 2047-2052.

[44] A. Graves, S. Fernandez, and J. Schmidhuber, "Bidirectional LSTM networks for improved phoneme classification and recognition," Artificial Neural Networks: Formal Models and Their Applications-ICANN 2005, pp. 753-753, 2005.

[45] 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.

[46] Y. Barak and D. Aizenberg, "Switching to aripiprazole as a strategy for weight reduction: a meta-analysis in patients suffering from schizophrenia," Journal of obesity, vol. 2011, 2010.