Conceptual framework for big medical data analysis

Aleksandar Stojmenski and Ivan Chorbev

Faculty of Computer Science and Engineering, University Ss Cyril and Methodius
Skopje, Republic of Macedonia
{aleksandar.stojmenski, ivan.chorbev}finki.ukim.mk

Abstract. The human health and the diseases that appear in the health systems are characterized with complex interactions between multiple factors. The rapidly increasing number of newly obtained diverse medical data is crucial for solving current and future medical challenges. Transforming knowledge from these kinds of data into efficient medical solutions combines transdisciplinary work with participants from medical, computer, data science background. In this paper we propose a conceptual framework for transdisciplinary research on medical data using computer science algorithms. The data in the system is obtained from sensor networks, diagnostic techniques and medical case studies which are measuring the performance of various organs of the human body. Part of the data for analysis and processing is already gathered from our previous researches. After the data analysis phase, we plan to improve the current or propose new algorithms, tools, methods and mathematical models for ICT systems aimed for working with medical data. This kind of systems would facilitate the current medical approaches and the treatment for certain diseases. Reaching this goal is not trivial, since we have a huge and increasing amount of available data and different new techniques for processing this data. That's why we need synchronization and homogenization of the whole available medical knowledge, storing it into a familiar format for researchers of different fields. Currently, two possible applications of the system are identified, extraction of max ejection fraction using echocardiography [1][2] and diseases prediction using ECG signal and heart rate processing [3][4][5]. The main goal of the framework is to provide more effective, efficient and more economically affordable solutions for emerging and future approaches to medical systems. Achieving this goal is not a trivial task and thus includes many challenges. One of the challenges is interpretability. Deep learning systems generally should perform well in medical data processing, but they are "black boxes," and they don't provide an explanation of the reasoning behind the conclusions. In our work, we plan to incorporate various mechanisms such as differentiable attention [6] to help doctors understand the internal workings of the neural network and its potential failure modes. Another challenge is choosing the best technique for big data analysis and extraction of medical knowledge that was not discovered before in the same context. There are vast amounts of data generated on daily basis, and there is also a vast number of available techniques which can be applied on the data. Choosing the correct method in combination with the chosen data will be analyzed. We plan to use the proposed conceptual framework on different datasets and compare the results with other proposed solutions in the literature.

Keywords: Medical data, analysis, image processing, disease prediction

References

1. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A., van Ginneken, B. and Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. Medical image analysis, 42, pp.60-88 (2017).
2. Xu, Y., Mo, T., Feng, Q., Zhong, P., Lai, M., Eric, I. and Chang, C., 2014, May. Deep learning of feature representation with multiple instance learning for medical image analysis. In 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1626-1630. IEEE (2014).
3. Ballinger, B., Hsieh, J., Singh, A., Sohoni, N., Wang, J., Tison, G.H., Marcus, G.M., Sanchez, J.M., Maguire, C., Olgin, J.E. and Pletcher, M.J., 2018. DeepHeart: Semi-Supervised Sequence Learning for Cardiovascular Risk Prediction. arXiv preprint arXiv:1802.02511 (2018).
4. Karpathy, A., 2015. The unreasonable effectiveness of recurrent neural networks. Andrej Karpathy blog. [Available from: http://karpathy.github.io/2015/05/21/rnn-effectiveness/]
5. Madani, A., Arnaout, R., Mofrad, M. and Arnaout, R., 2017. Fast and accurate classification of echocardiograms using deep learning. arXiv preprint arXiv:1706.08658 (2017).
6. Xu, K.; Ba, J.; Kiros, R.; Cho, K.; Courville, A.; Salakhudinov, R.; Zemel, R.; and Bengio, Y. 2015. Show, attend and tell: Neural image caption generation with visual attention. In International Conference on Machine Learning, pp. 2048–2057 (2015).