Learning from medical data streams: an introduction

Pedro Pereira Rodrigues[1], Mykola Pechenizkiy[2], Mohamed Medhat Gaber[3], and João Gama[1]

[1] LIAAD - INESC Porto, L.A. & University of Porto, Portugal
[2] Eindhoven University of Technology, The Netherlands
[3] University of Portsmouth, United Kingdom
pprodrigues@med.up.pt,m.pechenizkiy@tue.nl
mohamed.gaber@port.ac.uk,jgama@fep.up.pt

Abstract. Clinical practice and research are facing a new challenge created by the rapid growth of health information science and technology, and the complexity and volume of biomedical data. Machine learning from medical data streams is a recent area of research that aims to provide better knowledge extraction and evidence-based clinical decision support in scenarios where data are produced as a continuous flow. This year's edition of AIME, the Conference on Artificial Intelligence in Medicine, enabled the sound discussion of this area of research, mainly by the inclusion of a dedicated workshop. This paper is an introduction to LEMEDS, the Learning from Medical Data Streams workshop, which highlights the contributed papers, the invited talk and expert panel discussion, as well as related papers accepted to the main conference.

1 Introduction

Artificial Intelligence in Medicine is facing a new challenge, created by the rapid growth in information science and technology in general and the complexity and volume of data in particular. Medical settings are using sensors and networks of health information systems to integrate data from patients from which it is necessary to extract some sort of knowledge. The main issue is that this data production often takes the form of continuous flows of data.

Medical domains include several settings where data is produced in a streaming fashion, such as anatomical and physiological sensors, or incidence records and health information systems. New services like Google Health[1] appear allowing users to store and track information about their medical history, to connect to and stream data from medical devices. Medical data streams become widespread and call for development of intelligent tool for making use of these data. Decision support, alerting services, ambient intelligence, assisted leaving and personalization services are just few examples of expected uses of actionable knowledge extracted from medical data streams. All of them are characterized by the high-speed at which huge amounts of data are produced, and often require fast and

[1] http://www.google.com/health/

accurate information retrieval and analysis, that can effectively support clinical decisions.

Dealing with continuous, and possibly infinite, flows of data require different approaches for machine learning and knowledge discovery. Particular issues to address include summarization of infinite data, incremental and decremental learning, resource-awareness, real-time monitoring of changes and recurrences, etc. This is an incremental task that requires incremental learning algorithms that integrate artificial intelligence in medical domains. Streaming artificial intelligence is increasingly important in the research community, as new algorithms are needed to process medical data in reasonable time.

Furthermore, medical domains introduce extra peculiarities to the learning problem. For example, health information systems now deal with heterogeneous data sources, possibly distributed across healthcare institutions. Moreover, this data integration requirement yields possibly privacy-preserving issues, the same time it forces the system to take time, resources, and costs into consideration.

Currently, generic techniques for intelligent analysis and learning from streaming data are widely spread in the machine learning research community. Also, in the medical domain technological issues of data collection and storage, access, integration, information fusion, etc are also widely studied in the health informatics research community. However, adoption and development of tailored techniques for medical stream mining and clinical decision support is still to come.

2 Learning from Medical Data Streams

The artificial intelligence community has long identified machine learning as a prospective branch suited to address medical data [5]. However, its application to medical streams presents several issues that need to be solved. In this section we present an introduction to the *Learning from Medical Data Streams* workshop (LEMEDS 2011), organized in conjunction with *13th Conference on Artificial Intelligence in Medicine* (AIME 2011), highlighting the most recent works proposed in the field of learning from medical data streams.

2.1 LEMEDS 2011 Contributed Papers

The first edition of the *Learning from Medical Data Streams* workshop has included contributions from diverse fields of research that address medical data streams [3, 7, 9–11].

The fact that more and more medical data is being produced by sensors that measure physiological parameters [6] includes new challenges to artificial intelligence in general, and machine learning in particular. Jones et al. [3] proposed to interpret biosignals to improve mobile health monitoring for clinical decision support, using body sensor networks. The paper presents two possible applications and discusses the possibilities of applying machine learning in this ubiquitous streaming scenario, yielding a sound discussion in the learning from data streams forum. Biomedical signals have also been addressed by the two following works.

Rodrigues et al. [10] propose to improve cardiotocography monitoring using streaming statistics of both the fetal heart rate and the uterine contractions signals. The statistics will then be used to early detect changes in the monitored signals, and help in the prediction of birth outcome. It is an interesting position paper that has not experimentation yet, but should foster a sound discussion on the subject.

Sebastião et al. [11] developed a learning-based advisory system for detecting changes in depth of anesthesia signals. The paper addresses an important problem in the operational settings that can help to adapt administered doses for patients. The problem is formulated and addressed as the problem of handling concept drift in online settings, with the obtained experimental results, based on real data collected at one of the hospitals, being very promising.

Considering a higher-level approach to stream processing, McGregor et al. [7] presented a process mining framework to improve clinical guidelines in clinical care. The proposal is based on an extension of the CRISP-DM model, which considers temporal abstractions and multiple dimensions (CRISP-TDMn) and PaJMa to model the temporal abstractions as patient journeys. The paper presents a very interesting approach to knowledge discovery in a challenging scenario where data is produced as several heterogeneous streams.

Intensive care units are, undoubtedly in current healthcare services, the main clinical setting where data streams are being produced. But other medical data streams exist which differ from biomedical signals. An example is presented by Rodrigues et al. [9], where the authors describe a setting of integrated electronic health records, trying to improve the visualization mechanism of the increasing amount of clinical documents available in a central hospital. This is performed through the proposition of new bayesian approaches. As a position paper, this paper clearly presents the problem spaces and the research presents a valid and realistic problem that exists within healthcare today.

2.2 LEMEDS 2011 Invited Talk

The workshop chairs are honoured to include an invited talk by *Peter Lucas* (*Radboud University Nijmegen, The Netherlands*), one of the most knowledgeable researchers in the field of Artificial Intelligence in Medicine, with an emphasis on Bayesian techniques for clinical decision support. The talk, entitled *"Disease Monitoring and Clinical Decision Support"*, focus on the properties of biomedical data streams (e.g. from sensors) that impose constraints on how collected data can be exploited, and the new opportunities to monitor the progress of diseases in patients, reviewing some of these requirements and illustrating them by various real-world applications [6].

2.3 LEMEDS 2011 Panel Discussion

Given that it is still a young research area, the LEMEDS workshop aims at convening researchers from related fields in order to find and consolidate a network of interests. This way, the workshop will promote a panel discussion on the *"Challenges and roadmap for machine learning from medical data streams"*, with the participation of three scholar experts:

– *Carlo Combi* (*University of Verona, Italy*), an expert on temporal information systems, with an emphasis on the management of clinical information;
– *Carolyn McGregor* (*University of Ontario Institute of Technology, Canada*), an expert on health informatics, with an emphasis on data streams processing in critical care settings; and
– *João Gama* (*University of Porto, Portugal*), an expert on machine learning, with an emphasis on learning from ubiquitous data streams.

Topics that are suggested to be discussed include: main domains where medical data is produced as a stream; applications for LEMEDS; related fields of research; issues that differentiate this research area from other related fields; and best forums/venues for researchers to publish and discuss LEMEDS.

2.4 AIME 2011 Contributed Papers

This year's edition of AIME included six papers which address stream-related medical data [1, 2, 4, 8, 12, 13]. Although they might not directly include streaming machine learning techniques, they present scenarios and approaches which are relevant for discussion here.

Clinical time series are often produced in a stream. Enright et al. [1] proposed to analyze clinical time series using mathematical models and dynamic Bayesian networks. One type of medical data that is usually produced in a stream are physiological readings (e.g. heart rate). García-García et al. [2] used statistical machine learning to assess physical activity based on accelometry and heart rate readings. Wieringa et al. [12] also addressed physical activity, by defining an ontology-based dynamic feedback to the users. On a related topic, Jovic and Bogunovic [4] presented a Java-based framework to extrat features from cardiac rhythm. Rees et al. [8] presented the intelligent ventilator project, where physiological models are used in decision support, while Williams and Stanculescu [13] proposed to automate the calibration of a neonatal condition monitoring system. These are clearly related with intensive care units, a usual setting where data are produced as streams. The adaptation and application of such methods to streaming settings is a relevant path of research that should be considered.

3 Future Paths

LEMEDS is a recent trend of research that is yet to be consolidated. Thus, the corpus of contributions that has already been produced for AIME and LEMEDS, in 2011, supports the idea that, not only the involved research questions are both relevant and timely, but also knowledge in this domain is expanding and a small community is emerging, coming from related areas such as health informatics, machine learning, and clinical decision support. Given this, we believe that further activity is definitely going to exist and the field will produce valuable applications to improve healthcare.

Acknowledgments

The workshop chairs would like to thank all the participants that made this event possible. First, the authors of contributed papers are acknowledged for their participation. Then, we kindly thank the participation of our invited speaker, *Peter Lucas*, and our experts panel: *Carlo Combi*, *Carolyn McGregor* and *João Gama*. Also, we would like to thank the other Program Committee members for their help in peer-reviewing the contributed papers; thanks *Miguel Coimbra*, *Antoine Cornuéjols*, *Matjaz Kukar*, *Mark Last*, *Florent Masseglia*, *Ernestina Menasalvas*, *Josep Roure Alcobé*, *Cristina Santos*, *Alexey Tsymbal* and *Indre Zliobaite*. Ultimately, the chairs thank the attendants of the workshop, to whom the event is intended after all.

References

1. Enright, C.G., Madden, M.G., Madden, N.: Clinical time series data analysis using mathematical models and DBNs. In: Proceedings of the 13th Conference on Artificial Intelligence in Medicine. Lecture Notes in Artificial Intelligence, Springer Verlag, Bled, Slovenia (July 2011)
2. García-García, F., García-Sáez, G., Chausa, P., Martínez-Sarriegui, I., Benito, P.J., Gómez, E.J., Hernando, M.E.: Statistical machine learning for automatic assessment of physical activity intensity using multi-axial accelerometry and heart rate. In: Proceedings of the 13th Conference on Artificial Intelligence in Medicine. Lecture Notes in Artificial Intelligence, Springer Verlag, Bled, Slovenia (July 2011)
3. Jones, V., Batista, R., Bults, R., Akker, H.O.D., Widya, I., Hermens, H., Veld, R.H.I., Tonis, T., Vollenbroek-Hutten, M.: Interpreting streaming biosignals: in search of best approaches to augmenting mobile health monitoring with machine learning for adaptive clinical decision support. In: Proceedings of Learning from Medical Data Streams Workshop. Bled, Slovenia (July 2011)
4. Jovic, A., Bogunovic, N.: HRVFrame: Java-based framework for feature extraction from cardiac rhythm. In: Proceedings of the 13th Conference on Artificial Intelligence in Medicine. Lecture Notes in Artificial Intelligence, Springer Verlag, Bled, Slovenia (July 2011)
5. Lucas, P.: Bayesian analysis, pattern analysis, and data mining in health care. Current Opinion in Critical Care 10, 399–403 (2004)
6. Lucas, P.: Disease monitoring and clinical decision support. In: Proceedings of Learning from Medical Data Streams Workshop. Bled, Slovenia (July 2011)
7. McGregor, C., Catley, C., James, A.: A process mining driven framework for clinical guideline improvement in critical care. In: Proceedings of the Learning from Medical Data Streams Workshop. Bled, Slovenia (July 2011)

8. Rees, S.E., Karbing, D.S., Allerod, C., Toftegaard, M., Thorgaard, P., Toft, E., Kjærgaard, S., Andreassen, S.: The intelligent ventilator project: Application of physiological models in decision support. In: Proceedings of the 13th Conference on Artificial Intelligence in Medicine. Lecture Notes in Artificial Intelligence, Springer Verlag, Bled, Slovenia (July 2011)
9. Rodrigues, P.P., Dias, C., Cruz-Correia, R.: Improving clinical record visualization recommendations with bayesian stream learning. In: Proceedings of the Learning from Medical Data Streams Workshop. Bled, Slovenia (July 2011)
10. Rodrigues, P.P., Sebastião, R., Santos, C.C.: Improving cardiotocography monitoring: a memory-less stream learning approach. In: Proceedings of the Learning from Medical Data Streams Workshop. Bled, Slovenia (July 2011)
11. Sebastião, R., Silva, M., Gama, J., Mendonça, T.: Contributions to an advisory system for changes detection in depth of anesthesia signals. In: Proceedings of the Learning from Medical Data Streams Workshop. Bled, Slovenia (July 2011)
12. Wieringa, W., Akker, H.O.D., Jones, V.M., Akker, R.O.D., Hermens, H.J.: Ontology-based generation of dynamic feedback on physical activity. In: Proceedings of the 13th Conference on Artificial Intelligence in Medicine. Lecture Notes in Artificial Intelligence, Springer Verlag, Bled, Slovenia (July 2011)
13. Williams, C.K., Stanculescu, I.: Automating the calibration of a neonatal condition monitoring system. In: Proceedings of the 13th Conference on Artificial Intelligence in Medicine. Lecture Notes in Artificial Intelligence, Springer Verlag, Bled, Slovenia (July 2011)