Preface
The 3rd International Workshop on
Knowledge Discovery in Healthcare Data (KDH)

Introduction

The Knowledge Discovery in Health care Data (KDH) workshop series was established in 2016 to bring together AI and clinical researchers, fostering collaborative discussions and presenting AI research efforts to solve pressing problems in health care. This is the workshop's third year; held along with IJCAI/ECAI in Stockholm, Sweden and focusing on learning health care systems. For the first time, this workshop featured a challenge: The Machine Learning Blood Glucose Level Prediction Challenge.

The notion of the learning health care system has been put forward to denote the translation of routinely collected data into knowledge that drives the continual improvement of medical care. This notion has been described in many forms, but each follows a similar cycle of assembling, analyzing and interpreting data from multiple sources (clinical records, guidelines, patient-provided data including wearables, omic data, etc.), followed by feeding the acquired knowledge back into clinical practice. This framework aims to provide personalized recommendations and decision support tools to aid both patients and care providers, to improve outcomes and personalize care.

This framework also extends the range of actions possible in response to patient monitoring data, for example, alerting patients or automatically adjusting insulin doses when blood glucose levels are predicted to go out of range. Blood glucose level prediction is a challenging task for AI researchers with the potential to improve the health and well-being of people with diabetes. In the Machine Learning Blood Glucose Level Prediction (BGLP) Challenge, researchers came together to compare the efficacy of different machine learning prediction approaches on a standard set of real patient data.

The workshop received 22 submissions that were peer-reviewed by at least two reviewers each. After the review phase, 9 technical papers and 7 BGLP Challenge papers were accepted for presentation at the workshop. Among the accepted papers, the current trend of applying deep learning can be seen here as well, three papers use deep learning methods on health care data, while other methods used are: case-based reasoning, natural language processing or time series analysis. Another trend seen in the presentations was the need for open data sets that can drive the field forward and build on each other's work. This topic was addressed by the invited talk as well as by the included BGLP Challenge.

Invited Speaker: Jesse D. Raffa, MIT Critical Data, Cambridge, MA

Bio: Dr. Jesse Raffa, PhD, is a Research Scientist at the Laboratory of Computational Physiology at MIT. His background is in data science, biostatistics and epidemiology, having completed a PhD in biostatistics from the University of Waterloo, Canada, and more recently, a postdoc at the University of Washington, USA. His methodological interests include modeling complex longitudinal data and reproducible research. He has collaborated with colleagues in a diverse set of fields including: virology, addiction, psychiatry, the social sciences, genetics and critical care. He was the recipient of the New Investigator of the Year for Clinical Science by the Canadian Association of HIV/AIDS Research in 2004, and winner of a Distinguished Student Paper award by the Eastern North American section of the International Biometrics Society in 2013. Dr. Raffa has organized several critical care datathons since joining the Laboratory of Computational Physiology, has chaired a session at the Joint Statistical Meetings, was a core-instructor for the MITx massive open online course, Global Health Informatics to Improve Quality of Care, and created the residential MIT course: HST.953, Collaborative Data Science in Medicine.

Title: The Global Open Source Severity of Illness Scale (GOSSIS): Opportunities and Challenges

Abstract: This talk provides an overview of MIT Critical Data's effort to develop a global open severity of illness scale for critical care patients in collaboration with international partners. There is an increasing need for an openly available severity of illness scale which is well documented and easy to deploy. Many current offerings exist, but they are often: proprietary, expensive or developed at a single center or geographic region. Thus far, we have collaborators who have contributed data from North and South America, Asia and Oceania. We will discuss the current technical approach for handling this heterogeneous set of data, where differences in data collection practices and patient case mix can severely affect the ability to predict patient outcomes. The ability of models trained in one setting and applied in other settings will also be explored, with the ultimate aim to foster international collaboration in critical care research.

Accepted Papers

The following technical papers presenting original research works were accepted.

A number of papers address problems with activity monitoring using sensors and wearables. Holmes et al. present an approach for analysing patient domestic activity as part of recovery monitoring after surgery. The work uses a combination of ambient and wearable sensors to measure variables such as location, intensity of movement and types of physical activity. The authors present results from a real scenario in an out-of-lab environment to back their work.

Diaz et al. present a methodology for analysing changes in physical activity in school children following specific intervention programs. The work describes how different activity levels are recognised from raw accelerometer data and how length and frequency are computed for different activity levels. Clustering is then used to determine physical activity changes in the trial group, in comparison with a control group.

Vonstad et al. present a machine learning model to classify the quality of movements measured using a camera tracking system. The authors specifically focus on classifying weight-shifting movements, common in stroke rehabilitation patients, and evaluate different classification algorithms (Random Forests, K-nearest Neighbour and Support Vector Machines) .

Massie et al. present preliminary work to exploit sensor data to develop a fall prediction system for the residents of FitHomes, a Scottish Smart Home initiate lead by Albyn Housing Society Ltd, UK.

Finally, Agrawal et al. describe data gathering and analysis for a virtual trainer to assist nursing and care professionals move patients safely, usinga Kinect camera and pressure sensors on shoes to collect data for annotating as correct or incorrect (with error label).

The proceedings also include a number of papers implementing Machine Learning approaches for a number of prediction and classification tasks, including Natural Language Processing. To begin with, Biagi et al. propose the use of Compositional Data (CoDa) analysis to classify daily blood glucose patterns in people with type 1 diabetes (T1DM) based on features including the proportions of the day spent in different glycemic regions as well as glucose concentration patterns on different days for the patient. This work can provide valuable insight to care providers if patterns found can be linked with daily activities, with interesting preliminary results presented in the paper. Gupta et al. present an application using pre-trained TimeNets to learn features from time series patient health data to predict mortality in the ICU. Finally, Adduru et al. present a novel method to create a paraphrase corpus from webpages discussing medical topics.

The BGLP Challenge papers describe blood glucose (BG) level prediction approaches and the corresponding results on the OhioT1DM dataset. The participants used a wide variety of machine learning approaches, ranging from simple autoregressive models and ridge regression, to more sophisticated deep learning models such as recursive neural networks (RNN) with long short-term memory (LSTM) units, dilated RNNs, or temporal convolution networks (TCN).

Martinsson et al. use an LSTM-based approach in two instantiations: one that predicts just the BG level, trained with a mean squared error (MSE) loss, and one that predicts both the mean and variance of BG levels under a Gaussian distribution, trained with a negative log-likelihood loss. Both models use only a 30 minute history of blood glucose behavior and achieve comparable RMSE results. Analysis of patient-level results shows that a predicted higher variance correlates with a higher RMSE.

Chen et al. use a 3-layered Dilated RNN (DRNN), which enables a vanilla RNN to learn temporal dependencies at different resolutions, thus capturing the difference in temporal effects on blood glucose due to previous BG levels, carbohydrate intake, and insulin. When training data is small due to large regions of missing data, the subject-specific DRNN models are pre-trained on 10% of data from the other subjects. The trained DRNN models process a history of 30 minutes and are instantiated with vanilla RNN cells, which are shown to obtain better results compared to the more complex LSTM cells and gated recurrent units (GRU).

Zhu et al. use a WaveNet approach containing blocks of 5-layered Dilated CNN (DCNN). Like Chen et al. (this volume), they predict blood glucose using a history of BG levels, carbohydrate intake, insulin, and normalized time. To account for large blocks of missing data, the training set for each subject is augmented with 10% of data from the other subjects. Experimental results show that using carbohydrate intake, insulin, and normalized time improves the RMSE, compared with using BG levels alone as input.

Midroni et al. explored gradient-boosted decision trees (XGBoost), Random Forests (RF), and RNNs, using a diverse set of features and their combinations. The best test results are obtained using XGBoost. Subsequent feature ablation experiments with XGBoost show that optimal test RMSE is obtained using only blood glucose and self-reported information such as meals, finger-stick glucose, stress, illness, exercise and work.

Bertachi et al. use feed-forward neural networks (NNs) on two tasks: BG prediction and hypoglycemia prediction. For BG prediction, physiological model equations estimate the insulin on board, the glucose absorption rate, and the activity on board. Together with two BG derived features, these are used as input to a set of independently trained NNs, one for each of 6 predefined BG ranges.

Contreras et al. use a grammatical evolution approach to generate models of BG dynamics, based on the output of the same physiological equations used by Bertachi et al. (this volume). They also introduce a sinusoidal term to account for the circadian variations of the patients' physiology. Besides the standard RMSE, two additional metrics are used to define the fitness function: the glucose specific RMSE and the Clarke error grid. Corresponding results are then reported for 30, 60, and 90 minute predictions.

Xie and Wang compare the classic autoregression with exogenous inputs (ARX) with a set of popular ML algorithms, including XGBoost, support vector regression (SVR), as well as deep learning models such as LSTMs and temporal convolution networks (TCNs). All models use the total insulin delivery rate, meal sizes, and the heart rate. Two multi-step

prediction strategies are defined and evaluated: a recursive method that predicts one step ahead multiple times, and a direct method that predicts multiple steps ahead. Of all the models tested, the simple ARX model is shown to obtain the best RMSE on test data.

The final session of the workshop was a community discussion with a panel of the workshop organizers chaired by Razvan Bunescu. During this session the panelists discussed the challenges faced when creating open data sets in the health sciences as well as encouraging the community to open their source code. There was a clear agreement that sharing all types of sources is beneficial and should be encouraged.

We very much appreciate the support of the workshop chair, Kevin Leyton-Brown, as well as this year's conference chair Jeffrey S. Rosenschein and program chair Jérôme Lang. Further we would like to thank Fredrik Heintz for the local arrangements and Vesna Sabljakovic-Fritz for her administrative support.

We sincerely hope that the participants enjoyed this year's workshop program and that this collection of papers will inspire and encourage more AI-related research for and within healthcare in the future.

Kerstin Bach, Razvan Bunescu,
Oladimeji Farri, Aili Guo,
Sadid Hasan, Zina Ibrahim,
Cindy Marling, Jesse Raffa,
Jonathan Rubin, Honghan Wu

Stockholm, July 2018

Program Committee

- Imon Banerjee, Stanford University
- Isabelle Bichindaritz, State University of New York at Oswego
- Ali Cinar, Illinois Institute of Technology
- Kevin Bretonnel Cohen, University of Colorado School of Medicine
- José Manuel Colmenar, Universidad Rey Juan Carlos
- Bryan Conroy, Philips Research North America
- Sergio Consoli, Philips Research, Data Science Department
- Alexandra Constantin, Bigfoot Biomedical
- Vivek V Datla, Philips Research North America
- Spiros Denaxas, University College London
- Franck Dernoncourt, Massachusetts Institute of Technology
- Andrea Facchinetti, University of Padova
- Michele Filannino, MIT
- Pau Herrero, Imperial College London
- Ignacio Hidalgo, Universidad Complutense de Madrid
- Yuan Ling, Philips Research North America
- Bo Liu, Auburn University
- Stewart Massie, Robert Gordon University
- Claudia Moro, PUCPR
- Tristan Naumann, Massachusetts Institute of Technology
- Anna Rumshisky, University of Massachusetts, Lowell
- Sadiq Sani, Robert Gordon University Aberdeen
- Alexander Schliep, Gothenburg University
- Rushdi Shams, OneClass - Notesolution Inc.
- Giovanni Sparacino, University of Padova
- Shawn Stapleton, Philips Research North America
- Ozlem Uzuner, George Mason University
- Josep Vehi, University of Girona