Individuals and Organizations Completing Research in the IMEDS Lab
Page Group at UW-Madison

Research by Page Group at UW-Madison

The Page Group consists of David Page, Vitor Santos Costa, Aubrey Barnard, Jeremy Weiss, Peggy Peissig, and Sriraam Natarajan. Access is requested for Barnard (afbarnard@wisc.edu), Peissig (Peissig.Peggy@mcrf.mfldclin.edu), Santos Costa (vitor@biostat.wisc.edu), and Weiss (jcweiss2@wisc.edu). Barnard and Santos Costa already have accounts.

Research Objectives and Proposed Approaches
We wish to transfer three novel approaches for ADE discovery to OMOP (make them available as methods); the aim 1 approach has already been tested in the OMOP Lab but the approaches in Aims 2 and 3 have only been tested on other data so far. Because all three approaches are for ADE discovery, and we wish to evaluate them on OMOP tasks and make them available as methods, the proposed work meets the permitted uses.

Aim 1: *Make available as an OMOP method the approach in our AAAI-12 paper (Page et al., 2012).* We anticipate doing so with one month of additional testing for robustness on the OMOP server.

Aim 2: *Test whether a Markov network approach can improve ROC area in recovery of adverse drug events beyond the performance achieved in our AAAI-12 paper (Page et al., 2012).* The approach is discussed in Aubrey Barnard's prelim document attached, but here is a brief summary. Markov networks are now widely used in machine learning and artificial intelligence. In our approach the nodes of the Markov network correspond to logical atomic formulae of the form either (1) *event(P,E,T)*, where *P* is a patient, *E* is a drug or condition era, and *T* is a start time for that era, or (2) *before(P,E1,E2)*, where *P* is a patient, and *E1* and *E2* are events. Any three nodes *before(P,Ei,Ej)*, *event(P,Ei,Ti)*, and *event(P,Ej,Tj)* form a clique in the Markov network, with a corresponding potential that has an entry (weight) of zero in the case that $Ti > Tj$ and a weight of one otherwise. Each individual node also has a potential that is learned from the data via the standard gradient ascent algorithm for Markov network parameters, such that nodes that differ only on the patient (first argument) have their potentials tied together. A possible adverse drug event, in which drug *D* causes condition *C*, will be scored by the weight attached to the truth of the node *before(P,D,C)*, that is, how consistently *D* precedes *C* across patients. The expected advantage of the Markov network approach over simply scoring by how frequently *D* precedes *C*, as in our AAAI paper, is that the gradient ascent parameter learning algorithm should automatically adjust for confounding by correlated events (observed events, though not unobserved), for the following reason. Such confounding should correspond to another path in the Markov network between *D* and *C*, and increasing weights along that path should decrease the weight on the node *before(P,D,C).*

Aim 3: *Test whether continuous-time Bayesian networks (CTBNs) can outperform the approaches of Aims 1 and 2.* Bayesian networks, which are based on directed graphs, are an alternative to Markov networks, which are based on undirected graphs. Bayesian networks have been tailored specifically to longitudinal data via dynamic Bayesian networks (DBNs). Nevertheless, DBNs assume all data arrive in regular, well-defined time-steps, such as minutely, daily, or weekly, which typically is not true of observational clinical data. To address irregular longitudinal data, DBNs in recent years have been largely supplanted by CTBNs for such

applications. CTBNs are directed (possibly cyclic) graphs, where the rate at which a variable changes values is governed by intensity parameters that are conditional on the parents of that variable. We have recently developed a novel structure learning and parameter learning algorithm for CTBNs (Weiss et al., 2012) and propose to test it on the adverse drug event task. In this setting the nodes will be drugs and conditions of interest. A potential adverse drug event, with drug D causing condition C, will be scored by how much the intensity increases for condition C changing from false to true when drug D changes from false to true for a patient. The expected advantage of this CTBN approach is that, unlike the approaches of Aims 1 and 2, it takes into account how quickly a condition occurs after a drug in a soft fashion (that is, without setting a window after drug exposure in which the condition must occur). OMOP's well-defined drug exposure eras will be a particular advantage for CTBNs.

Impact

The three aims cover three highly active research areas in machine learning and data mining currently: relational learning, learning undirected graphical models, and learning factored continuous-time models. All three are appropriate to ADE discovery, and this work will further evaluate them and make them available. The first aim will make available to the research community an artificial intelligence and machine learning approach to adverse drug events that already has shown promise. Graphical models, especially Markov networks and CTBNs, are widely used in the intersection of statistics and machine learning. The second and third aims will test these leading graphical graphical model-based approaches in a natural way on the task of discovering adverse drug events, to see if they have potential to improve our ability to uncover such events.

Experience

The researchers for whom access is requested are all exceptionally well qualified for working with this type of data. Aubrey Barnard is a PhD student in UW-Madison's Dept. of Computer Sciences who is a co-author on one paper on ADE discovery already and who has several years' experience with a leading EHR software company. Peggy Peissig has a PhD in Clinical and Translational Science and has 20 years' experience working with Marshfield Clinic's EHR system and data warehouse, both in IT and in research, is active in the EMERGE consortium, and has numerous publications on analysis of clinical data. Vitor Santos Costa has a PhD in Computer Science and holds a research grant in machine learning for adverse drug events. Jeremy Weiss is an MD/PhD student at UW-Madison and has publications in *AI Magazine*, *NIPS* (one of two leading international machine learning research venues), and *Innovative Applications of Artificial Intelligence*.

Data Requirements

We request access to the five data sets that were used in the first two years' of methods evaluations. We expect we will work most with MSLR but would like to test on all five.

Funding Sources

NIGMS grant R01GM097618-01 and NIH 5T15LM007359

Duration

Upon agreement and signature until end of 2014.

Individuals and Organizations Completing Research in the IMEDS Lab
Page Group at UW-Madison

References

David Page, Vítor Santos Costa, Sriraam Natarajan, Aubrey Barnard, Peggy Peissig, and Michael Caldwell (2012). Identifying Adverse Drug Events by Relational Learning. *Proceedings of AAAI-12.*

Jeremy Weiss, Sriraam Natarajan, and David Page (2012). Multiplicative Forests for Continuous-Time Processes. *Proceedings of NIPS-12*.