tures for every level. The challenge for relational tree-based learners is that the search algorithm is greedy; identifying high-level relations requires traversing several "is a" relationships first, and thus they might not be found in a greedy search. Expanding internal nodes to longer clauses has been implemented with some success [Natarajan et al., 2010; Anderson and Pfahringer, 2009], although this does have the effect of rapidly increasing the number of features to consider during branching. The use of SRL algorithms could also allow the use of relations like patient physicians and providers, which form complex relations less "patient-disease"-oriented but ones that still may be central to patient care. Questions regarding disease heritability could also be addressed through relational family-based analyses.

Given our initial success, we plan to extend our work by including more potential risk factors for learning (i.e., include all the measurements on all the patients). This will be challenging as the number and frequencies of the measurements will differ greatly across patients. In our current model, we used time as the last argument of our predicates. While there is a vast body of work in learning and reasoning with temporal models in propositional domains, the situation is not the same for relational models. We plan to investigate a principled approach to learning and reasoning with relational dynamic models that will allow physicians to monitor the cardiovascular risk levels of patients over time and develop personalized treatment plans. Finally, we plan to build a complete machine learning system for identifying risk factors across many diseases given the longitudinal data available in the EHR.

Acknowledgments

We acknowledge the gracious support by the grant 1UL1RR025011 from the Clinical and Translational Science Award (CTSA) program of the National Center for Research Resources, National Institutes of Health, the Wisconsin Genomics Initiative, NIGMs grant R01GM097618-01 and NLM grant R01LM011028-01, the Translational Science Institute (TSI) of Wake Forest Baptist Health, the Computation and Informatics in Biology and Medicine training grant 5T15LM007359, and the University of Wisconsin-Madison Medical Scientist Training Program.

References

Anderson, G., and Pfahringer, B. 2009. Relational random forests based on random relational rules. In *Interational Joint Conferences in Artificial Intelligence*.

Antonopoulos, S. 2002. Third report of the national cholesterol education program (ncep) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (adult treatment panel iii) final report. *Circulation* 106(3143):3421.

Blockeel, H., and Raedt, L. D. 1998. Top-down induction of first-order logical decision trees. *Artificial Intelligence* 101:285–297.

Bg-Hansen, E.; Larsson, C. A.; Gullberg, B.; Melander, A.; Bostrm, K.; Rstam, L.; and Lindblad, U. 2007. Predictors of acute myocardial infarction mortality in hypertensive patients treated in primary care. *Scandinavian Journal of Primary Health Care* 25(4):237–243.

Craven, M., and Shavlik, J. 1996. Extracting tree-structured representations of trained networks. In *Neural Information Processing Systems*, 24–30.

Dietterich, T.; Ashenfelter, A.; and Bulatov, Y. 2004. Training conditional random fields via gradient tree boosting. In *International Conference on Machine Learning*.

Friedman, J. H. 2001. Greedy function approximation: A gradient boosting machine. *Annals of Statistics* 1189–1232.

Getoor, L., and Taskar, B. 2007. *Introduction to Statistical Relational Learning*. MIT Press.

Greenland, P.; Alpert, J. S.; Beller, G. A.; Benjamin, E. J.; Budoff, M. J.; Fayad, Z. A.; Foster, E.; Hlatky, M.; Hodgson, J. M. B.; and Kushner, F. G. 2010. 2010 accf/aha guideline for assessment of cardiovascular risk in asymptomatic adults. *Journal of the American College of Cardiology* j. jacc. 2010.09. 001v1.

Group, D. P. C. 2002. Prediction of mortality from coronary heart disease among diverse populations: is there a common predictive function? *Heart* 88:222–228.

Gutmann, B., and Kersting, K. 2006. TildeCRF: Conditional Random Fields for Logical sequences. In *European Conference on Machine Learning*.

Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; and Witten, I. H. 2009. The weka data mining software: an update. *Special Interest Group on Knowledge Discovery and Data Mining Explorations Newsletter* 11(1):10–18.

Heritage Provider Network, I. 2011. Heritage health prize.

Kersting, K., and Driessens, K. 2008. Non–parametric policy gradients: A unified treatment of propositional and relational domains. In *International Conference on Machine Learning*.

Manson, J. A. E.; Tosteson, H.; Ridker, P. M.; Satterfield, S.; Hebert, P.; O'Connor, G. T.; Buring, J. E.; and Hennekens, C. H. 1992. The primary prevention of myocardial infarction. *New England Journal of Medicine* 326(21):1406–1416.

McCarty, C. A.; Wilke, R. A.; Giampietro, P. F.; Wesbrook, S. D.; and Caldwell, M. D. 2005. Marshfield clinic personalized medicine research project (pmrp): design, methods and recruitment for a large population-based biobank. *Personalized Medicine* 2(1):49–79.

McCarty, C. A.; Peissig, P.; Caldwell, M. D.; and Wilke, R. A. 2008. The marshfield clinic personalized medicine research project: 2008 scientific update and lessons learned in the first 6 years. *Personalized Medicine* 5(5):529–542.

Natarajan, S.; Khot, T.; Kersting, K.; Guttmann, B.; and Shavlik, J. 2010. Boosting Relational Dependency networks. In *Inductive Logic Programming*.

Natarajan, S.; Joshi, S.; Tadepalli, P.; Kristian, K.; and Shavlik, J. 2011. Imitation learning in relational domains: A functional-gradient boosting approach. In *Interational Joint Conferences on Artificial Intelligence*.

Neville, J.; Jensen, D.; Friedland, L.; and Hay, M. 2003. Learning Relational Probability trees. In *Knowledge Discovery and Data Mining*.

Sutton, R.; McAllester, D.; Singh, S.; and Mansour, Y. 2000. Policy gradient methods for reinforcement learning with function approximation. In *Neural Information Processing Systems*.

Wilson, P. W. F.; D'Agostino, R. B.; Levy, D.; Belanger, A. M.; Silbershatz, H.; and Kannel, W. B. 1998. Prediction of coronary heart disease using risk factor categories. *Circulation* 97(18):1837.