Over the past years, Machine Learning (ML) approaches have taken large strides in their predictive accuracy and ease of use, resulting in ML being used in increasing number of domains. At the same time, information has grown exponentially in terms of its size and complexity. Inter-related objects (people, atoms, words, etc.) spread across multiple relations (friends, bonded, dependent, etc.) is now a common occurrence in many domains such as molecular chemistry, medical diagnosis, social networks and information extraction.

To deal with noisy multi-relational data, Statistical Relational Learning (SRL) models have been proposed. Unlike most ML approaches that rely on a fixed number of features for every example, SRL models can handle an arbitrary number of features. For instance, a patient can have all their test results, where the number of tests may vary between patients, as features. But due to the increased complexity, SRL models do not scale to large domains, especially when learning the structure of the probabilistic dependencies (e.g., discovering the dependence of parents' chromosomes on a person's blood type). My research has mainly concentrated on developing more accurate, scalable structure-learning approaches for SRL models to make them more generally and easily applicable. Since these approaches do not rely on an expert designed model, I was able to use them in diverse domains ranging from natural language processing to medical diagnoses to network analysis.

Structure Learning

I started my thesis research on learning the structure for two popular SRL models - Relational Dependency Networks (RDN) [1] and Markov Logic Networks (MLN) [2]. I used functional-gradient boosting to incrementally learn the structure of these models. My approach was able to learn a more accurate model than state-of-the-art structure-learning approaches in a fraction of the time. To show the generality of this approach, I applied it to various tasks such as citation segmentation, link prediction and information extraction (IE). For the IE task, I designed a NLP pipeline to parse articles using a Stanford parser and creating first-order logic facts from the structured output such as parse trees, dependency graphs, etc. We used relational structure learning to build a model for extracting NFL game information (winner, loser, score, etc.) from sports articles.

Partially labeled data

Due to the complexity of the structure learning task, most approaches do not handle missing data in relational models. I developed an Expectation-Maximization based approach for learning the structure of three relational models [3]. We derived the EM steps for functional gradients and experimentally showed that our approach can handle missing data. This work has been invited to a Machine Learning special issue.

Although our approach can handle some missing negative labels, NLP domains may have only the positive examples labeled with all the negative labels missing. One-class classification (OCC) approaches have been proposed in propositional domains to handle this task, but cannot be directly applied to relational models. To handle OCC in relational models, I developed a relational distance metric that can be used to perform density estimation. The distance measure is updated to maximize the likelihood of the marked examples based on the current density estimates. Each update to the distance measure finds novel features that help in differentiating among relational examples. This work has been submitted to AAAI'14 and is under review [4].

Applications and Big Data

I have worked on learning models for multiple tasks such as Alzheimer's prediction [5] and temporal information extraction (IE) [6]. In the former task, we used MRI segments to perform three-way relational classification to differentiate between normal, mild cognitively impaired and Alzheimer's patients. In the latter task, we learned a model along with expert advice to order events temporally.

In addition to these, I have also worked on a large-scale relation-extraction task, namely NIST's TREC Knowledge Base Acceleration (KBA) 2013. In this task, we were provided with 9 TB of documents from which we had to extract relations for entities of interest. Moreover, we had to detect only changes in the relation values to recommend changes, say to a Wikipedia editor. For this problem, I developed a system using a relation extractor followed by a novelty detection system to detect changes in extracted relations. To scale to such a large set of documents, I used a high-recall document filter along with a distributed grid computing framework to work on partitions of the document set.

Future Work

Tractable Models

One of the biggest challenges in relational models is scaling inference to large domains. Unlike propositional models, relational models do not assume independence between examples thereby reducing the scope of performing parallel inference. But by learning models that allow for efficient inference, either due to symmetries or decomposability, I can use these models on web-scale domains. Tractable inference can also be achieved by learning directed acyclic models, which have the added benefit of interpretability, but introduce the additional challenge of detecting cycles in relational models efficiently.

Adaptive Learning

My current work on one-class classification can be applied to multiple applications with partially labeled data such as anomaly detection where only few anomalies might be provided to train the model. But given the subjectivity of the definition of anomalies, just learning a model to fit to a few observed anomalies may lead to a model that does not capture the intended definition of anomaly (e.g. bombing events can be anomalous based on location or number of people).

Apart from the subjectivity of anomalous event, the definition of anomaly or any other class label may change gradually or abruptly over time. Our boosting approach can naturally be extended to handle concept drift to incrementally update the model as the concept definition changes. The challenge lies in ignoring noisy examples while still updating the model for both short term (abrupt) and long term (gradual) changes to concept definition.

Application

With the abundance of data on the internet, there is a large amount of labeled data being generated that can lead to interesting challenges. One such source of natural language text are internet forums viz reddit.com. Many of these forums tag the posts providing us with a large multi-class classification dataset. Apart from the content of the post, it also provides meta-information such as the user, time of posting and comments. Moreover these posts are linked to each other and also to webpages on other domains providing a rich relational structure that can be exploited for this task. Such a dataset can also be used to detect concept drift (change in definition of 'current news'), perform anomaly detection (e.g. posts with high votes, but zero comments), learn tractable relational models and perform domain adaptation (from one sub-forum to another). I plan to work on generating a rich relational dataset from these forums which I can then use to evaluate and direct my proposed work described above. The dataset will also be valuable to the research community as a source for challenging problems and to evaluate their approaches.

My research has been guided by solid theoretical problems and inspired by real world applications. I believe that research should focus on the theoretical problems to better understand the models and thereby ensure broad applicability across multiple domains. At the same time, by applying these model in real world tasks, we can have an impact in people's lives either through a simple web search or identifying medical risks.

References

[1] S. Natarajan, T. Khot, K. Kersting, B. Guttmann, and J. Shavlik, "Gradient-based boosting for statistical relational learning: The relational dependency network case," *Machine Learning*, 2012.

[2] T. Khot, S. Natarajan, K. Kersting, and J. Shavlik, "Learning Markov logic networks via functional gradient boosting," in *ICDM*, 2011.

[3] T. Khot, S. Natarajan, K. Kersting, and J. Shavlik, "Learning relational probabilistic models from partially observed data - opening the closed-world assumption," in *ILP*, 2013.

[4] T. Khot, S. Natarajan, and J. Shavlik, "Relational one-class classification: A Non-parametric approach," in *AAAI*, 2014. Under Review.

[5] S. Natarajan, B. Saha, S. Joshi, A. Edwards, E. Moody, T. Khot, K. Kersting, C. Whitlow, and J. A. Maldjian, "Relational learning helps in three-way classification of alzheimer patients from structural magnetic resonance images of the brain," *International Journal of Machine Learning and Cybernetics*, 2013.

[6] T. Khot, S. Srivastave, S. Natarajan, and J. Shavlik, "Learning relational structure for temporal relation extraction," in *StarAI Workshop in UAI*, 2012.