CSE 515 Literature Survey

Mitchell Koch, Justin Huang

Friday, May 17, 2013

For our project on structured learning of relation association graphs we chose to focus our literature survey on previous relation entailment work [2], on work in relation extraction using matrix factorization [5], as well as on related work in probabilistic modeling of relations between entities [7].

Berant, et al. 2012

This paper describes a method for learning textual entailments for relations. For example, given a question like, *"what affects blood pressure?"*, a possible answer is *"alcohol reduces blood pressure"*. This makes it important to know that *X reduce Y* entails *X affect Y*. Rather than try to learn these rules using only local features, Berant et al. present a system that learns an entire entailment graph subject to global constraints.

In the entailment graph, the nodes are binary predicates with one of the arguments instantiated. For example, *X associated-with nausea*, *X reduce nausea*, and *X help-with nausea* might be nodes in the graph. One of the arguments must be instantiated for scalability reasons. However, in previous work [1], Berant et al. describe an extension to this system that allows the arguments to be instantiated with types, such as *disease causes symptom*. The main global constraint on the graph is transitivity, which requires that if $X \rightarrow Y$, and $Y \rightarrow Z$, then it must be the case that $X \rightarrow Z$.

The graph is created in two steps. The first step is to train a supervised classifier that, given two predicates, estimates how likely it is that one entails the other. The second step is to optimize the sum of the scores for the edges given the transitivity constraint.

The authors describe two ways of representing predicates: binary templates such as *X reduce Y*, and unary templates such as *X reduce* and *reduce Y*. On top of that, they also have several ways of representing the arguments to the predicates in their data. Finally, they have two different metrics for comparing the similarity of two predicates. One of those metrics is symmetric, meaning it can tell if two predicates are similar, but without knowing which entails which. The other metric, however, is asymmetric and gives the information about which predicate entails the other. The authors combine these dimensions together to produce 12 different scores, which are used as the representation

of the predicate pairs to the classifier. The training data for the classifier is automatically generated using entailments from WordNet.

Given the output of the classifier on every pair of nodes, the next step is to optimize the graph subject to the transitivity constraint. To do this, the authors frame the problem as a linear programming problem. They define an indicator function I_{uv} denoting that node u entails node v. They then optimize the sum of the edge scores. If S_{uv} is the score between u and v, then the global score they optimize is $\sum_{u \neq v} S_{uv} I_{uv}$. The transitivity constraint is encoded as $\forall_{u,v,w} I_{uv} + I_{vw} - I_{uw} \leq 1$. They then solve this using standard integer linear programming packages.

This paper provides a framework for thinking about incorporating both local and global data into their system, which we might consider for ours. In particular, the authors give a formal definition of the posterior distribution of the graph given the features $P(G|F)$, where F is the set of all I_{uv}. Their interpretation of the entailment graph and the independence assumptions are both things we could draw inspiration from. One of the strong points of the system is how it creates a consistent graph by enforcing transitivity. On the other hand, that leads to scalability issues, as the global transitivity constraint forces them to solve an NP-hard integer linear programming problem. As a result, they can only generate entailments for relations when one of the arguments is already fixed.

Riedel, et al. 2013

This paper focuses on the problem of aggregate relation extraction: extracting relation triples consisting of two arguments and a relation type. This paper unifies doing relation extraction with a fixed schema with an approach which does not use a fixed schema, like Open IE [3]. They describe this as a "universal schema", the union of both the surface form predicates as in Open IE and relations in fixed knowledge bases like Freebase. They use a matrix factorization approach similar to collaborative filtering, where relations are like products and tuples for those relations are like customers. A big strength of this approach is that it can take advantage of these different kinds of relations and learn jointly from them.

They set up their model to estimate the probability of a tuple being in a relation using log linear models over parameters defined using a latent feature model with latent representations for relations and tuples like in generalized PCA. They use a neighborhood approach to interpolate the confidence based on similar relations by using weights for the strengths between relations. They also use an entity model to allow representation of latent types, allowing for representing the compatibility between argument slots and entities.

They use ranked pairs to avoid the need for negative examples using the concept of implicit feedback from collaborative filtering. They maximize an objective with the goal being pairs that are in the observed data rank higher than those not in the observed data. This is maximized using Stochastic Gradient

Descent.

In their experiments, they use articles from different time ranges in the New York Times for training and testing and also split Freebase facts to align with different facts in the train and test sets. For evaluation they compare their system against several other systems and use an information retrieval style evaluation where for each relation, each system gets to choose the tuples it is most confident in and a common pool is chosen from those to evaluate all the systems. A weakness of this evaluation is that it casts the problem of relation extraction as a problem of information retrieval instead of measuring how many facts in a document are actually extracted correctly. This still should allow relative comparisons, however, and they do outperform the other relation extraction methods. Their combined model including the latent entity representation performs the best for predicting Freebase relations, although it appears better to not use the latent entity representation for surface patterns.

One source of background for this work comes from relational clustering, with the goal being to cluster surface patterns to find latent relations; however, in this paper the goal is to predict patterns where they are not observed instead of directly identifying new patterns. Instead of using an assumption of symmetry in clustered patterns, they assume relations entail others.

We intend to try a similar approach for relation entailment by using a similar matrix of entity pairs and relation types from both Open IE, Freebase, as well as those extracted from Freebase using distant supervision as in [4]. Some weaknesses of this approach hinge on the fact that it only provides aggregate relation extraction over a whole corpus as opposed to relation extraction for individual sentences; however, for relation entailment this should be less of an issue as the idea is to find in general what relations associate with each other.

Taskar, et al. 2003

This paper approaches the problem of predicting links and their types for entities involved in relations. It takes the relational Markov network (RMN) framework of [6] and extends it to predict not only labels but also links. This is applied to a graph of university webpages as well as a social network. The background is that of Probabilistic Relational Models, a directed version of RMNs.

For these problems, relational schemas are defined with object types and attributes. In this case, both pages and links are objects, and attributes can be information such as from a bag of words. A RMN is defined to provide cliques and potentials between attributes of related entities using templates, such as between class labels of linked pages. The parameters are learned from data over a single relational instantiation using belief propagation.

For the university webpages dataset, evidence for relations includes hyperlinks from an entity page or an owned page to another entity's page, as well as virtual links based on mentioning a page title or an email address. They tried observing page categories and not observing them. Using both modeling of website sections along with triad transitivity templates outperformed a flat model

evaluating individual links in the observed case. For the unobserved entity label case, they used a similarity model as well as combining with a flat model, which together did better than just the flat model but not in all cases. The social network dataset includes information about students and lists of friends. They tried predicting where some links are observed in the test data and use cliques modeling shared characteristics between students. Here this similarity model outperforms the flat model as well.

They found that both similarity templates, including links sharing a common endpoint, as well as transitivity templates helped. The strength of this method is clear over flat classification of links, because it allows taking into account more information, and it also performs significantly better than using generative PRMs, because the RMNs are discriminative models. This also allows using subgraph patterns not possible to represent in the directed PRMs. A weakness of this approach is that because of the complexity of the problem, belief propagation is used and does not always converge.

We could model the entities and relations that we are dealing with in a similar way. Instead of pages we would have Freebase entities, instead of links we would have relations or surface patterns, and potentially we could have entity types, either from Freebase or latent for labels. The cliques could group together entities appearing in the same sentence or document. Then the relation extraction problem would be to learn more links given some observed links, and the relation entailment problem could be approached by extracting the probabilities of one kind of link given another and forming a corresponding relation entailment graph that agrees.

References

[1] J. Berant, I. Dagan, and J. Goldberger. Global learning of typed entailment rules. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, HLT '11, pages 610–619, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.

[2] J. Berant, I. Dagan, and J. Goldberger. Learning entailment relations by global graph structure optimization. *Comput. Linguist.*, 38(1):73–111, Mar. 2012.

[3] O. Etzioni, M. Banko, S. Soderland, and D. S. Weld. Open information extraction from the web. *Commun. ACM*, 51(12):68–74, Dec. 2008.

[4] R. Hoffmann, C. Zhang, X. Ling, L. S. Zettlemoyer, and D. S. Weld. Knowledge-based weak supervision for information extraction of overlapping relations. In *ACL*, pages 541–550, 2011.

[5] S. Riedel, L. Yao, B. M. Marlin, and A. McCallum. Relation extraction with matrix factorization and universal schemas. In *Joint Human Language Tech-*

nology Conference/Annual Meeting of the North American Chapter of the *Association for Computational Linguistics (HLT-NAACL '13)*, June 2013.

[6] B. Taskar, P. Abbeel, and D. Koller. Discriminative probabilistic models for relational data. In *Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence*, UAI'02, pages 485–492, San Francisco, CA, USA, 2002. Morgan Kaufmann Publishers Inc.

[7] B. Taskar, M. F. Wong, P. Abbeel, and D. Koller. Link prediction in relational data. In *NIPS*, 2003.