Relational Logistic Regression: the Directed Analog of Markov Logic Networks

Seyed Mehran Kazemi, David Buchman, Kristian Kersting, Sriraam Natarajan, and David Poole
cs.ubc.ca/~smkazemi/ cs.ubc.ca/~davidbuc/
www-ai.cs.uni-dortmund.de/PERSONAL/kersting.html
homes.soic.indiana.edu/natarasr/ cs.ubc.ca/~poole/

Abstract

Relational logistic regression (RLR) was presented at the 14th International Conference on Principles of Knowledge Representation and Reasoning (KR-2014). RLR is the directed analogue of Markov logic networks. Whereas Markov logic networks define distributions in terms of weighted formulae, RLR defines conditional probabilities in terms of weighted formulae. They agree for the supervised learning case when all variables except a query leaf variable are observed. However, they are quite different in representing distributions. The KR-2014 paper defined the RLR formalism, defined canonical forms for RLR in terms of positive conjunctive (or disjunctive) formulae, indicated the class of conditional probability distributions that can and cannot be represented by RLR, and defined many other aggregators in terms of RLR. In this paper, we summarize these results and compare RLR to Markov logic networks.

Introduction

Relational probabilistic models (Getoor and Taskar 2007; de Raedt et al. 2008) or template-based models (Koller and Friedman 2009) represent the probabilistic dependencies among relations of individuals. In these models, individuals about which we have the same information are exchangeable (i.e. the individuals are treated identically when we have no evidence to distinguish them), and the probabilities are about relations among individuals and can be specified independently of actual individuals. Relational models can be directed or undirected.

Markov logic networks (MLNs) (Richardson and Domingos 2006; Domingos et al. 2008) are undirected models that define distributions in terms of weighted formulae. MLNs can also be used to define conditional probabilities when all the neighbours of a variable are observed. However, when weighted formulae are used to define conditional probabilities in a directed model, which we call relational logistic regression (RLR) (Kazemi et al. 2014), this results in quite different models.

Markov Logic Networks and Relational Logistic Regression

Markov logic networks (MLNs) (Richardson and Domingos 2006; Domingos et al. 2008) and relational logistic regression (RLR) (Kazemi et al. 2014) are defined in terms of weighted formulae. In MLNs the formulae are used to define joint probability distributions, whereas in RLR the formulae are used to define conditional probabilities. MLNs and RLR are identical when all but a query leaf variable are observed, but otherwise differ.

A **weighted formula** is a triple $\langle L, F, w \rangle$ where L is a set of logical variables, F is a formula where all of the free logical variables in F are in L, and w is a real-valued weight.

An MLN is a set of weighted formulae[1] where the probability of any world (complete assignment to all ground random variables) is proportional to the exponential of the sum of the weights of the instances of the formulae that are true in the world.

RLR is a form of aggregation, defining a conditional probability in terms of a set of weighted formula for a given relation (defining the parents of the relation). The conditional probability of the relation given its parents is proportional[2] to the exponential of the sum of the weights of the instances of the formulae that are true for that parent assignment and for the assignment to the relation.

Example 1 Suppose we have weighted formulae:

$$\begin{array}{l} \langle \{\}, q, \alpha_0 \rangle \\ \langle \{x\}, q \wedge \neg r(x), \alpha_1 \rangle \\ \langle \{x\}, q \wedge r(x), \alpha_2 \rangle \\ \langle \{x\}, r(x), \alpha_3 \rangle \end{array}$$

Treating this as an MLN, if the truth value of $r(x)$ for every individual x is observed:

$$P(q \mid obs) = sigmoid(\alpha_0 + n_F \alpha_1 + n_T \alpha_2) \qquad (1)$$

[1]MLNs typically do not explicitly include the set of logical variables as part of the weighted formulae, but use the free variables in F. If one wanted to add an extra logical variable, x, one could conjoin $true(x)$ to F where $true$ is a property that is true of all individuals.

[2]In MLNs there is a single normalizing constant, guaranteeing the probabilities of the worlds sum to 1. In RLR, normalization is done separately for each possible assignment to the parents.

where *obs* has $r(x)$ is true for n_T individuals, and false for n_F individuals out of a population of n individuals (so $n = n_T + n_F$), and $sigmoid(x) = 1/(1+e^{-x})$.

Note that in the MLN, α_3 is not required for representing the conditional probability (because it cancels out), but can be used to affect $P(r(A_i))$, where A_i is an individual of x.

In Kazemi et al. (2014), the sigmoid, as in Equation (1), is used as the definition of RLR. (Kazemi et al. (2014) assumed all formulae were conjoined with $q\wedge$, and omitted $q\wedge$ from the formulae.) RLR only defines the conditional probability of q given each combination of assignments to the $r(x)$ (using Equation (1)); when not all $r(x)$ are observed, separate models of the probability of $r(x)$ are needed.

In summary: RLR uses the weighted formulae to define the conditional probabilities, and MLNs use the weighted formulae to define the joint probability distribution.

Example 2 Suppose people want to go to a party and the party is fun for them if they know at least one social person in the party. In this case, we have a parametrized random variable $funFor(x)$ which is a child of two parametrized random variables, $knows(x,y)$ and $social(y)$. The following weighted formulae can be used to model the dependence of the relation $funFor(x)$ on its parents:

$$\langle\{x\}, funFor(x), -5\rangle$$
$$\langle\{x,y\}, funFor(x)\wedge knows(x,y)\wedge social(y), 10\rangle$$

RLR sums over the above weighted formulae and takes the *sigmoid*, resulting in:

$$P(funFor(x) = True|\Pi) = sigmoid(sum), \quad sum = -5 + 10n_T$$

where Π is the assignment of values to the parents, and n_T represents the number of individuals y for which $knows(x,y)\wedge social(y)$ is $True$ for the given x. When $n_T = 0$, $sum < 0$ and the probability is close to 0; when $n_T > 0$, $sum > 0$ and the probability is close to 1.

Example 3 This example is similar to Example 1, but uses only positive conjunctions[3], and also involves multiple logical variables of the same population.

$$\begin{array}{l}
\langle\{\}, q, \alpha_0\rangle \\
\langle\{x\}, q\wedge true(x), \alpha_1\rangle \\
\langle\{x\}, q\wedge r(x), \alpha_2\rangle \\
\langle\{x\}, true(x), \alpha_3\rangle \\
\langle\{x\}, r(x), \alpha_4\rangle \\
\langle\{x,y\}, q\wedge true(x)\wedge true(y), \alpha_5\rangle \\
\langle\{x,y\}, q\wedge r(x)\wedge true(y), \alpha_6\rangle \\
\langle\{x,y\}, q\wedge r(x)\wedge r(y), \alpha_7\rangle
\end{array}$$

In RLR and in MLNs, if the truth value of $r(x)$ for every individual x is observed:

$$P(q|obs) = sigmoid(\alpha_0 + n\alpha_1 + n_T\alpha_2 + n^2\alpha_5 + n_T n\alpha_6 + n_T^2\alpha_7)$$

[3]Here $true(x)$ is true of every x. This notation is redundant. For the traditional MLN notation, one can remove the explicit set of logical variables and keep the $true(\cdot)$ relations. If you are happy with the explicit logical variables, you can remove the $true(\cdot)$ predicates. Removing both is incorrect. Keeping both is harmless. You could add formulae that involve negations or disjunctions, but they are redundant (Kazemi et al. 2014).

where *obs* has $r(x)$ true for n_T individuals out of a population of n. The use of two different logical variables of the same population in a formula (e.g. x and y in the last formula of this example) allows a sigmoid-of-a-quadratic dependency on the population sizes (n_T, n_F, n).

This example uses only positive conjunctions (but needs to have the "extra" logical variables). This is one of the canonical representations for RLR. Using the canonical forms for RLR, Kazemi et al. (2014) proved that in RLR, each conditional probability is the sigmoid of a polynomial of counts. Moreover, any conditional probability that is a sigmoid of polynomials of counts can be represented by RLR.

Example 4 Suppose we have a set of movies and the ratings people gave to these movies in a star-rating system. We define a movie to be "popular" if the average of its ratings is more than 3.5. In this case, we have a parametrized random variable $popular(m)$ which is a child of a parametrized random variable $rate(p,m)$. The following weighted formulae can approximate the conditional dependence of $popular(m)$ on its parent (the second weighted formula is repeated 5 times for different values of $i \in \{1,2,3,4,5\}$):

$$\langle\{m\}, popular(m), -\varepsilon\rangle$$
$$\langle\{m,p\}, popular(m)\wedge rate(p,m) = i, i-3.5\rangle$$

RLR sums over the above weighted formulae and takes the sigmoid resulting in:

$$P(popular(m) = True|\Pi) = sigmoid(-\varepsilon + s - 3.5n)$$

where s denotes the sum of the ratings and n represents the number of ratings for this movie. The value inside the sigmoid is positive if $mean = \frac{s}{n} > 3.5$ and is negative otherwise (ε is used to ensure the the value inside the sigmoid is negative when $mean = 3.5$). This example demonstrates how the aggregator *mean* can be represented by RLR. It has been demonstrated in (Kazemi et al. 2014) that many other popular aggregators such as OR, AND, noisy-OR, noisy-AND, max, more-than-t *true*s, more-than-$t\%$ *true*s, and mode can be represented in terms of RLR.

Related Work

Logistic regression has been previously used in relational domains for learning purposes. Popescul et al. (2002) use inductive logic programming (ILP) to generate first-order rules for a target relation, create features by propositionalizing the rules, and then use logistic regression to learn a classifier based on these features.

Similar ideas to RLR have been used for discriminative learning in MLNs. Huynh and Mooney (2008) use an ILP technique to generate discriminative clauses for a target relation and then use logistic regression with L1-regularizer to learn the weights with automatic feature selection (automatic structure learning). They empirically demonstrate how their discriminative model outperforms MLNs when there is a single target relation. Their work can be considered as one approach for learning an RLR conditional probability for a relation when all its parents are observed.

Conclusion

The paper introducing relational logistic regression (RLR) (Kazemi et al. 2014):

- Defines RLR in terms of weighted formulae
- Investigates how conditional probabilities are defined based on the number of (tuples of) individuals for which each weighted formula holds
- Presents non-redundant canonical forms for the formulae, e.g., positive conjunctions
- Proves that a conditional dependency of child relation on population sizes of its parents can be represented by RLR if and only if it can be represented as the sigmoid of a polynomial of these population sizes
- Shows how other aggregators can be represented by RLR, including OR, AND, noisy-OR, noisy-AND, mean$> t$, max, more-than-t *true*s, more-than-t% *true*s, and mode.

References

de Raedt, L.; Frasconi, P.; Kersting, K.; and Muggleton, S. 2008. *Probabilistic Inductive Logic Programming: Theory and Applications*. Springer-Verlag.

Domingos, P.; Kok, S.; Lowd, D.; Poon, H.; and Richardson, M. 2008. *Markov Logic. In Raedt, L.D., Frasconi, P., Kersting, K. and Muggleton, S., eds., Probabilistic Inductive Logic Programming*. Springer.

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

Huynh, T. N., and Mooney, R. J. 2008. Discriminative structure and parameter learning for markov logic networks. In *Proc. of the international conference on machine learning*.

Kazemi, S. M.; Buchman, D.; Kersting, K.; Natarajan, S.; and Poole, D. 2014. Relational logistic regression. In *Proc. 14th International Conference on Principles of Knowledge Representation and Reasoning (KR-2014).*

Koller, D., and Friedman, N. 2009. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, Cambridge, MA.

Popescul, A.; Ungar, L. H.; Lawrence, S.; and Pennock, D. M. 2002. Towards structural logistic regression: Combining relational and statistical learning. In *KDD Workshop on Multi-Relational Data Mining*.

Richardson, M., and Domingos, P. 2006. Markov logic networks. *Machine Learning* 42:107–136.