Multinomial Relation Prediction in Social Data

A Dimension Reduction Approach

Review By: Pulkit Jain (10543)

Introduction/Problem Addressed

The explosion of internet in the recent days has led to emergence of social data. Web services that produce/operate with this data involve social actions such as tagging, annotating and other actions which involve heterogeneous and multiple objects. User actions are modeled as relations. Prediction of these actions is important for many fundamental services such as search engines and recommendation systems. The work done addresses the action prediction problem using dimensionality reduction approach.

Proposed Solution

The solution proposed is based upon reduction of a relation instance into multiple binomial instances and using attribute information of involved objects.

- Nonlinear dimensionality reduction is applied to binomial relations which embeds objects and relations into a common subspace where each object is close to the relation it participates in.
- Minimizing distances (Euclidian distance metric used) between the embedding of objects and relations results in solving an optimization problem which is formulated as a generalized eigenvalue problem. This **gives us exact solutions that are globally optimal** and are **robust to the sparse data**.
- Incorporating the attribute information in the solution only requires replacing the low dimensional embedding used previously by linear projections of the attribute information to obtain embedding in the required subspace. This addresses the "**cold start"** problem where the sparsity of data is high.

Thus the input to the process/algorithm followed includes sets of objects that can participate in relations, and a set of observed relation instances. The output obtained is a list of other possible relations amongst the objects sorted in order of likelihood of the relation.

Input:	$S_1, S_2 \ldots S_K$	- Set of Objects
	$O (\subset S_1 \times S_2 \times \ldots \times S_K)$	- Set of Observed Relation Instances
	$\Phi_1, \Phi_2 \ldots \Phi_K$	- K matrices representing object attributes
Output:	$S_1 \times S_2 \times \ldots \times S_K \setminus O$	- Possible relations sorted in order of likelihood

Results/Contributions

- Datasets from Twitter for prediction of "favorite" and "retweet" actions and del.icio.us (a social tagging service) were used for measuring the performance.
- The data was also tested against the existing PARAFAC and TUCKER tensor decomposition methods along with the proposed solutions to build comparisons.
- The proposed solution showed highest robustness when the data was sparse in terms of the relation instances.
- The solution did not perform well when tested against data that was sparse in relations that involved particular objects, but the performance was high when attribute information of objects was used along with the relation information.

- Overall, **the proposed solution is highly robust to data sparsity in comparison to the existing methods** of prediction using tensor decompositions.

REFERENCES

[1]Nozomi Nori;Danushka Bollegala; Hisashi Kashima 2012. Multinomial Relation Prediction in Social Data. A Dimension Reduction Approach. In Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence.

[2] Lin, Y.-R.; Sun, J.; Castro, P.; Konuru, R.; Sundaram, H.;and Kelliher, A. 2009. Metafac: community discovery via relational hypergraph factorization. In Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

[3] CP Decomposition, Wikipedia. http://en.wikipedia.org/wiki/CP_decomposition

[4]Tucker Decomposition, Wikipedia. http://en.wikipedia.org/wiki/Tucker_decomposition