Summary 22: Graph Regularized Transductive Classification on Heterogenous Information Networks

Susan Cherry

This paper deals with classification on heterogenous information networks. Information networks are large numbers of data objects that are linked to each other, such as citation networks and co-author networks. Classification can allow us to discover hidden structure of information networks and the different roles played by the objects in the network. Previous work typically focused on homogeneous information networks but in real life multiple types of objects often form heterogeneous information networks. One example includes bibliographic information networks, which contain papers, authors, and venues. This paper considers the problem of classifying heterogeneous networked data into classes. Another example is a movie network, which includes movies, directors, actors, and keywords that specify the genre.

The following characteristics make transductive classification on heterogenous networks very challenging: the complexity of the network structure, lack of features, and lack of labels. This paper introduces a graph-based regularization framework that addresses these problems and classifies the data with a network topology and object/link types. In section 3, they introduce concepts, notation, and formal definition of the problem. In section 4, the describe the graph-based regularization framework. They give information using a bibliographic information network, which can be formalized as follows: The estimated confidence measure of two objects belongs to class k and f_{ip}^k and f_{jp}^k should be similar if the objects are linked together. The confidence estimation f_i^k should be similar to the ground truth.

Next they introduce the algorithm formally. They show that the homogenous version of the algorithm is equivalent to the graph based learning method and then describe the connection between the algorithm and 21 on heterogenous information networks. Then they analyze the computational complexity of the iterative solution and the closed form solution.

Finally, they move to experimental results. They use bibliographic data and try to classify it into research communities. They compare GNetMine with Learning with Local and Global Consistency, Weighted-vote Relational Neighbor Classifier, and Network Only Link Based classification. Overall, they find that GNetMine performs best on all types of objects and outperforms LLGC by preserving consistency on the subgraphs that correspond to each type of links separately. This allows it to minimize aggregated error. Then the authors move to model selection. They find that GNetMine achieves better performance than all of the other algorithms again. Overall, the authors find that parameter selection doesn't significantly affect the performance of GNetMine.

In conclusion, this paper prevents a new graph based regularization framework for transductive classification on heterogeneous information networks. Experiments show that this algorithm performs better than existing algorithms.