ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Multi-Label Classification in Sparsely Labeled Network

Prathyusha
Malla Reddy College Of Engineering & Technology, Maisammaguda Dhulapally, Secunderabad, Pin 500100

ABSTARCT

Classification in sparsely labeled networks is challenging to traditional neighborhood-based methods due to the lack of labeled neighbors. Multi-label classification is a critical problem in many areas of data analysis such as image labeling and text categorization. In this paper we propose a probabilistic multi-label classification model based on novel sparse feature learning. By employing an individual sparsity inducing `1-norm and a group sparsity inducing `2,1-norm, the proposed model has the capacity of capturing both label interdependencies and common predictive model structures. We formulate this sparse norm regularized learning problem as a non-smooth convex optimization problem, and develop a fast proximal gradient algorithm to solve it for an optimal solution. Our empirical study demonstrates the efficacy of the proposed method on a set of multi-label tasks given a limited number of labeled training instances In this paper, we propose a novel behavior-based collective classification (BCC) method to improve the classification performance in sparsely labeled networks. In BCC, nodes' behavior features are extracted and used to build latent relationships between labeled nodes and unknown ones. Since mining the latent links does not rely on the direct connection of nodes, decrease of labeled neighbors will have minor effect on classification results. In addition, the BCC method can also be applied to the analysis of networks with heterophily as the homophily assumption is no longer required. Experiments on various public data sets reveal that the proposed method can obtain competing performance in comparison with the other state-of-the-art methods either when the network is labeled sparsely or when homophily is low in the network

INTRODUCTION:

Multi-label classification is a critical problem in many areas of data analysis, where each data instance can be assigned into multiple categories. For example, in image labeling [Zhou and Zhang, 2006] or video annotation [Qi et al., 2007], a given scene usually contains multiple objects of interests. In text categorization [Schapire and Singer, 2000], a given article or webpage can be assigned into multiple topic categories. In gene and protein function prediction [Elisseeff and Weston, 2002], multiple functions are typically associated with each gene and protein. Due to its complex nature, the labeling process of a multi-label data set is typically more expensive or time-consuming comparing to single-label cases, since the annotator needs to evaluate each class label even when the positive labels appear in a very sparse pattern.

Cite this Article as: Prathyusha "Multi-Label Classification in Sparsely Labeled Network", International Journal & Magazine of Engineering, Technology, Management and Research (IJMETMR), ISSN 2348-4845, Volume 7 Issue 6, June 2020, Page 24-33.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

To mitigate the needs hence the cost of collecting labeled data, learning effective multi-label classifiers from a small number of training instances thus is important to be investigated. One straightforward approach for multi-label classification is to cast the multi-label learning problem as a set of independent single label classification problems [Lewis et al., 2004; Chen et al., 2007]. This simple method has the obvious drawback of ignoring useful correlation information between the predictions of multiple labels. Developing methods to exploit the label dependency information and capture shared prediction structures among the multiple labels is critical in multi-label learning. In the literature, many approaches have been proposed to address multi-label learning by either exploiting label dependencies [Elisseeff and Weston, 2002; Godbole and Sarawagi, 2004; Guo and Gu, 2011; Schapire and Singer, 2000; Petterson and Caetano, 2011], or capturing the common prediction structures of the multiple binary prediction tasks associated with the individual classes [Yan et al., 2007; Zhang and Zhou, 2008; Yu et al., 2005; Ji et al., 2010]. But very few have taken both aspects into account for multi-label learning. In this paper we propose a novel probabilistic multi-label classification model to simultaneously exploit both label dependency knowledge and shared prediction structures across labels based on sparse feature learning. Sparse feature learning has been effectively exploited in simultaneous multi-task learning problems by enforcing an `2,1 norm [Argyriou et al., 2006; Liu et al., 2009; Obozinski et al., 2006]. Different from these works which consider only common input features, our model contains two types of features: structural label dependency features associated with each individual single-label prediction task, and common input features which are shared across the multiple single-label prediction tasks. We first propose to learn sparse label dependency structures by associating an `1-norm regularization with the label dependency features, aiming to overcome possible overfitting issues. It induces a sparse conditional dependency network under probabilistic multi-label predictors. Then by adding another `2,1-norm regularization over the input features, we formulate the overall probabilistic multi-label learning problem as a joint convex optimization problem with combined sparse norm regularizations, where an `1-norm is used for the sparse structural feature selection, and an `2,1-norm is used for selecting the discriminative input features shared across multiple binary predictors. We develop a fast proximal gradient algorithm to solve the proposed optimization problem for an optimal solution. Our empirical results on a number of multi-label data sets demonstrate the efficacy of the proposed approach when the number of training instances is small, comparing to a few related probabilistic methods

RELATEDWORK Multi-label classification has received increasing attention from machine learning community in recent years, due to its wide applications in practice. There is a rich body of work on multi-label learning in the literature. We provide a review to the most related methods in this section. One simple approach for multi-label classification is to cast the multi-label learning problem as a set of independent single label classification problems [Lewis et al., 2004; Chen et al., 2007]. Such an approach however is

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

unsatisfactory, since the different labels occurring in a multi-label classification problem are not independent. On the contrary, they often exhibit strong correlations or dependencies. Capturing these correlations in different manners have led to many advanced developments in multi-label classification. A significant number of multi-label learning approaches have been proposed to exploit label dependencies in classification model formulation, including ranking based methods [Elisseeff and Weston, 2002; Schapire and Singer, 2000; Shalev-Shwartz and Singer, 2006; Fuernkranz et al., 2008], pairwise label dependency methods [Zhu et al., 2005; Petterson and Caetano, 2011], probabilistic classifier chains [Dembczynski et al., 2010], large-margin methods [Guo and Schuurmans, 2011; Godbole and Sarawagi, 2004; Hariharan et al., 2010], and probabilistic graphical models [Ghamrawi and McCallum, 2005; de Waal and van der Gaag, 2007; Bielza et al., 2011; Zaragoza et al., 2011; Guo and Gu, 2011]. Most of these methods however involve resource-consuming optimization procedures or extensive model-structure learning processes. On the other hand, another set of methods attempt to exploit the relationships between multiple binary classification models in multi-label learning by capturing their common prediction structures [Yan et al., 2007; Zhang and Zhou, 2008; Yu et al., 2005; Ji et al., 2010]. These two types of multi-label prediction approaches have in general all achieved good empirical performance. However, very few methods have taken both aspects of label dependencies and shared model structures into account for multi-label learning. The Bayesian network models for multi-label learning [de Waal and van der Gaag, 2007; Bielza et al., 2011] although take steps in this direction by learning separate feature subnetwork, class subnetwork, and feature-class bridge subnetwork, they nevertheless are limited to problems with a small number of discrete feature variables. The proposed approach in this paper aims to integrate the strengths of label dependency based methods and common prediction structure based methods within a novel convex sparse feature learning framework. From the perspective of capturing label dependency, our work is closely related to the simple multi-label learning methods in [Godbole and Sarawagi, 2004; Guo and Gu, 2011]. The work in [Godbole and Sarawagi, 2004] uses a very intuitive and simple procedure to exploit multi-label dependency information. It first trains a set of binary SVM classifiers, one for each of the K classes. Then it uses the K binary classifiers to produce K label features to augment the original features of each instance. Finally another set of K binary SVM classifiers are trained on the augmented instances. Its testing process follows a corresponding procedure. This work is simple and straightforward, but lacks of principled explanation. The work of [Guo and Gu, 2011] generalizes this intuitive idea into a principled probabilistic framework based on directed conditional dependency networks (CDNs), where each label variable takes all the other label variables as its parents. In a CDN model, the training of multiple binary prediction models can be interpreted as maximizing the approximated joint conditional distributions of the label variables. The training process is even simpler than the SVM method in [Godbole and Sarawagi, 2004]. It only requires training one set of K

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

independent binary probabilistic classifiers, and a simple Gibbs sampling procedure is used for conducting inference in the testing phase. Nevertheless, the CDN model in [Guo and Gu, 2011] uses a fully connected directed graph as the label dependency structure, which can easily fall into the trap of overfitting, especially when there are a limited number of training instances. With sparse feature learning, the proposed approach in this paper aims to integrate the strength of the CDN model but overcome its drawbacks.

Title: A brief survey of machine learning methods for classification in networked data and an application to suspicion scoring

This paper surveys work from the field of machine learning on the problem of within-network learning and inference. To give motivation and context to the rest of the survey, we start by presenting some (published) applications of within network inference. After a brief formulation of this problem and a discussion of probabilistic inference in arbitrary networks, we survey machine learning work applied to networked data, along with some important predecessors— mostly from the statistics and pattern recognition literature. We then describe an application of within-network inference in the domain of suspicion scoring in social networks. We close the paper with pointers to toolkits and benchmark data sets used in machine learning research on classification in network data. We hope that such a survey will be a useful resource to workshop participants, and perhaps will be complemented by others. We describe a guilt-by-association system that can be used to rank entities by their suspiciousness. We demonstrate the algorithm on a suite of data sets generated by a terroristworld simulator developed under a DoD program. The data sets consist of thousands of people and some known links between them. We show that the system ranks truly malicious individuals highly, even if only relatively few are known to be malicious ex ante. When used as a tool for identifying promising data-gathering opportunities, the system focuses on gathering more information about the most suspicious people and thereby increases the density of linkage in appropriate parts of the network. We assess performance under conditions of noisy prior knowledge (score quality varies by data set under moderate noise), and whether augmenting the network with prior scores based on profiling information improves the scoring (it doesn't). Although the level of performance reported here would not support direct action on all data sets, it does recommend the consideration of network-scoring techniques as a new source of evidence in decision making. For example, the system can operate on networks far larger and more complex than could be processed by a human analyst.

EXISTING SYSTEM:

While the rapid development of information technology has greatly improved our ability to collect data in recent years, traditional methods of network classification are facing new challenges: in the era of big data, substantial proportion of nodes are typically unlabeled in many settings. For such sparsely labeled networks, the neighbors of an unknown node are mostly unlabeled as well; consequently, many neighborhood-based

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

methods cannot achieve satisfied performance for such kind of networks. For this reason, a lot of efforts have been made recently in order to develop new techniques for sparse labeling problem, such as semi-supervised learning, active learning and latent link mining.

DISADVANTAGES:

- All the above methods can handle the sparse labeling problem to some extent, however, the interacting behavior of nodes, which is important to the formation of network structure, is not considered.
- When the number of nodes in one class is much larger than the other class, unknown nodes are more likely to be classified as the same category as the majority.

PROPOSED SYSTEM:

Propose a novel behavior based collective classification (BCC) method for network data in this study. In the new method, firstly, we extract the behavior feature of nodes in the network; then, instead of including all labeled nodes in the classification process, we screen valuable nodes which are most relevant for the classification; Finally, since latent links can be estimated between unknown nodes and valuable nodes by analyzing their behavior feature, collective classification is performed based on the latent links to infer the class of unknown nodes. Experiment reveals that the method performs competitively on several public real-world datasets and can overcome the challenge of classification in sparsely labeled networks and networks with lower homophily.

In sparsely labeled networks, the labels of nodes are much fewer, making it difficult to leverage label dependencies to make accurate prediction. Without considering the label information, it can be found that the network structure can still provide useful information. Therefore, most researches focus on utilizing the network structure to predict unknown nodes. For example, CN method estimates the similarity of nodes by local structure (the number of common neighbors). However, it becomes ineffective when handling the sparsely labeled network classification task in some situations.

Instead of including all labeled nodes, BCC only allows the most relevant nodes for classification to improve the performance on sparsely labeled networks. So in the next, we screen valuable nodes by using correlation analysis and similarity analysis respectively. Given an unknown node u, we first compare the correlation between u and each labeled node, then, nodes with correlation coefficients exceeding a threshold will be added into the valuable node set Vu. After that, we compare the similarity between u and each node in Vu, and add the top-K similar nodes into set V0 u, which is then used to classify the unknown node u by voting. It should be noted that, our method is flexible to integrate other techniques in each step, e.g., classification by voting can be replaced by other classifiers, such as SVM, linear regression and so on. Finally, in order to deal with challenges of classification in extremely sparsely labeled network, we perform collective inference, in which the newly labeled nodes will be added to the labeled node set and used for inferring the rest unknown nodes. BCC method consists of four steps for classification, and in this section, we introduce the implement of each step in detail. Firstly, we will describe

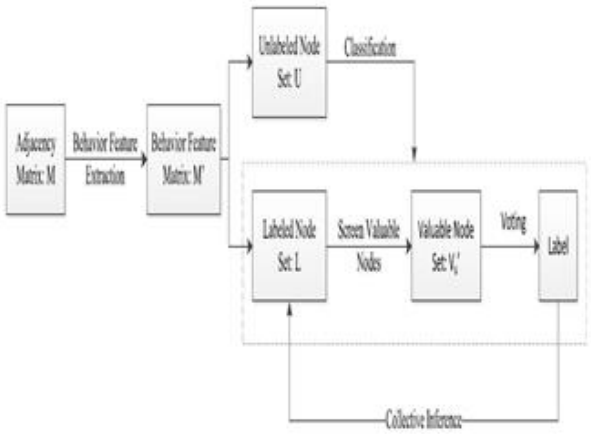
ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

how to extract behavior feature, which has shown more discriminative ability in sparsely labeled networks. In order to handle the imbalanced dataset, we only allow the most relevant nodes in the classification process by using correlation and similarity analysis. Then we introduce the strategy of voting for classification. Collective inference procedure is used to handle the extremely sparse labeling problem, which is described afterwards. Finally, the algorithm is given to show the details of our method.

ADVANTAGES:

- In BCC, the behavior feature of nodes is extracted for classification, which has shown more discriminative ability to traditional methods.
- Then, instead of using all the labeled nodes, we screen the most-relevant nodes according to the calculation of correlation and similarity, which can overcome the effects of noise and imbalanced dataset.itle: Suspicion scoring based on guilt-by-association, collective inference, and focused data access.

ARCHITECTURE:



MODULES

- **SEMI-SUPERVISED LEARNING**
- **BEHAVIOR BASED COLLECTIVE CLASSIFICATION**
- **SCREEN VALUABLE NODES FOR CLASSIFICATION**

SEMI-SUPERVISED LEARNING:

Making use of both labeled and unlabeled data, semi supervised learning is an effective method for classification in sparsely labeled networks. One type of this method is to design a classification function which is sufficiently smooth with respect to the intrinsic structure collectively revealed by labeled and unlabeled points. Zhou et al. Propose a simple iteration algorithm, which considered global and local consistency by introducing a regularization parameter. By modeling the network with constraint on label consistency, Zhu et al. propose a Gaussian Random Field (GRF) method by introducing a harmonic function, of which the value is the average of neighboring points. Another type of semi-supervised learning methods is the graph-cut method , which assumes that more closely connected nodes tend to belong to the same category. The core idea is to find a cut set with the minimum weight by using different criteria. However, the high cost of computing often lead to poor performance of the algorithm when applied in large networks. Some other algorithms use random walk on the network to obtain a simple and effective solution by propagating labels from labeled nodes to unknown nodes. Based on passaging time during random walks with bounded lengths, Callut et al. and Newman [30] introduce a novel technique, called D-walks, to handle semi-supervised classification problems in

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

large graphs. Zhou and Schlkopf dene calculus on graphs by using spectral graph theory, and propose a regularization framework for classification

Problems on graphs. However, many semi-supervised learning methods rely heavily on the assumption that the network exhibits homophily, i.e., nodes belonging to the same class tend to be linked with each other. Meanwhile, the implementation of semi-supervised learning algorithm often requires a large amount of matrix computation, and thus is infeasible for processing large datasets. Many methods have been developed to overcome these limitations. For example, Tong et al. propose a fast random walk with restart algorithm to improve the performance on largescale dataset. Lin et al. propose a highly scalable method, called Multi-Rank-Walk (MRW), which requires only linear computation time in accordance to the number of edges in the network . Mantrach et al. design two iterative algorithms which can be applied in networks with millions of nodes to avoid the computation of the pair wise similarities between nodes. Gallagher et al.design an even-step random walk with restart (Even-step RWR) algorithm, which mitigates the dependence on network homophily effectively.

BEHAVIOR BASED COLLECTIVE CLASSIFICATION:

Since behavior feature can provide a different kind of information that may be useful in sparsely labeled networks, we propose a novel Behavior-based Collective Classification method (BCC) in this paper to handle the sparse labeling problem. The process of BCC in network data consists of four steps: behavior feature extraction, screening valuable nodes, classification by voting and collective inference. We assume that nodes may belong to the same class if their behavior features are similar. Therefore, given the adjacency matrix M of a network, we will extract nodes' behavior feature at first to obtain the feature matrixM0 , of which the i-th row vector is the behavior feature of node i. Instead of including all labeled nodes, BCC only allows the most relevant nodes for classification to improve the performance on sparsely labeled networks. So in the next, we screen valuable nodes by using correlation analysis and similarity analysis respectively. Given an unknown node u, we first compare the correlation between u and each labeled node, then, nodes with correlation coefficients exceeding a threshold will be added into the valuable node set Vu. After that, we compare the similarity between u and each node in Vu, and add the top-K similar nodes into set V0 u, which is then used to classify the unknown node u by voting. It should be noted that, our method is flexible to integrate other techniques in each step, e.g., classification by voting can be replaced by other classifiers, such as SVM, linear regression and so on. Finally, in order to deal with challenges of classification in extremely sparsely labeled network, we perform collective inference, in which the newly labeled nodes will be added to the labeled node set and used for inferring the rest unknown nodes.

SCREEN VALUABLE NODES FOR CLASSIFICATION:

The labeled nodes are much fewer in sparsely labeled network, so traditional methods tend to

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

utilize all the labeled nodes in the classification process. However, involving unrelated

nodes in the classification process will only bring noise data and lead to poor performance. Moreover, when classes of labeled nodes are imbalanced, unknown nodes will be more likely to be labeled the same as the majority. To solve this issue, we show how to find the most relevant nodes, from the perspective of correlation and similarity of behavior feature, to reduce the impact of noise data.

1) CORRELATION OF BEHAVIOR FEATURE

Correlation analysis is an important method to measure the relationship between two observed variables. We assume that nodes of the same class should have higher correlation of their behavior feature. Therefore, given an unknown node u, the labeled node set L, and Pearson correlation threshold P, we can screen out the valuable node set Vu by:

$$V_u = \{v | v \in L \wedge corr(v, u) > P\},$$

Correlation analysis is an important method to measure the relationship between two observed variables. We assume that nodes of the same class should have higher correlation of their behavior feature. Therefore, given an unknown node *u*, the labeled node set *L*, and Pearson correlation threshold *P*, we can screen out the valuable node set *Vu* by:

$$V_u = \{v | v \in L \wedge corr(v, u) > P\},$$

where *corr*(*v*,*u*) represents pearson correlation value between node *v* and *u*. *corr*(*v*;,*u*) can be calculate by

$$corr\ (v, u) = \frac{1}{N-1} \sum_{i=1}^{N} \left(\frac{v_i - \bar{v}}{s_v}\right)\left(\frac{u_i - \bar{u}}{s_u}\right),$$

where *N* is the number of nodes in the network, *v* is the mean value of node *v*'s behavior feature vector, s_v is the standard deviation of node *v*'s behavior feature vector, and analogously for N*u* and s_u. As we will see in the experiments, labeled nodes of higher correlation with *u* will have bigger influence in the classification process.

2) SIMILARITY OF BEHAVIOR FEATURE

Correlation analysis is able to discover the latent relationship of behavior features, but not enough for finding the most relevant nodes in weighted networks. For example, in Table 1, it can be found that the connection behavior of node A and node B are almost same, except subtle changes when connecting node F. As we know, experimental datasets are crawled from real-world networks. In the crawling process, information may be lost inevitably, which means node A and node B may have the same connection behaviors with node F in real-world network. In this situation, it is obvious that the connection behavior of node *B* is more similar with *A* compared to *C*. However, by using the correlation analysis, *C* will have a higher correlation value with *A* (*corr*(*A*,*C*) = 1, *corr*(*A*, *B*)= 0.99).

In order to improve the ability to handle this problem, we implement a similarity analysis procedure after the correlation analysis. We assume that nodes of the same class should have more similar

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

behavior features. Since nodes' behavior features are expressed as probability distributions, symmetric Kullback-Leibler (KL) divergence [44] can be used to measure the similarity:

$$D_{sKL}(i,j) = \frac{1}{2}[\sum_{n=1}^{N} p_{(i,n)} \ln \frac{p_{(i,n)}}{p_{(j,n)}} + \sum_{n=1}^{N} p_{(j,n)} \ln$$

Where p (i,j) is the probability of connection from node i to node j.

A node with smaller KL divergence will indicate that it has similar behavior feature to the unknown node and thus is more valuable for the classification. Therefore, given the unknown

node u, we calculate the similarity of node u with each node in V_u, and add the top-K similar nodes to set V'_u.

3) MULTI-LABEL CLASSIFICATION IN MAJORITY-VOTING:

After the above screening process, the valuable node set V'_u, is then used to classify unknown nodes. We use the majority voting strategy, which means that the label of an unknown

node is determined by the class of nodes which belongs to the majority in V'_u:

$$C(u|V'_u) = \arg\max_{C_j} \sum_{x \in V'_u} I(C(x) = C_j), \quad j = 1, \cdots$$

in which $C(u)$ represents the class of node u, J is the total number of classes in the network, and C_j is the j-th class. I (.) is a discriminate function such that when $C(x) = C_j$, I (.) = 1 and otherwise I (.) =0.

4) COLLECTIVE INFERENCE:

In order to improve the classification performance in sparsely labeled network, collective inference procedure is introduced in our method, in which newly labeled nodes will be used

for inferring the rest unknown nodes. Consequently, as the classification process goes on, the labeled node set expands constantly and existing knowledge continues to accumulate to guide subsequent classification process.

However, introducing collective inference process will come with a new problem: unknown nodes that have been labeled will affect subsequent prediction process, so labeling

is relevant to the order of how unknown nodes are classified. To mitigate such effect, we propose an iteration strategy. In the i-th iteration, the labeled node set Li will use the labels at the end of the previous iteration. Then, each initial unknown node will be classified by using behavior based classification method and get a new label. If the node has never been labeled in the previous iteration, it will be added to L_i, otherwise we will update L_i with the new label. The iteration continues until labels of all initial unknown nodes stay unchanged in Li or the maximum number of iterations is reached.

This process inherits the idea of iterative classification (IC) method. However, instead of using local neighbors, our method relies on latent links created by behavior feature. Since we extract a few valuable nodes to participate in the classification, it does not need to update numerous nodes in each iteration and the process typically converges efficiently in a limited number of iterations.

When the labeled data is very sparse, the performance of traditional collective classification might be largely degraded due to

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

the lack of sufficient neighbors. However, in our method, latent links can be mined between labeled nodes and unknown nodes by using behavior feature, even nodes do not connect directly. It means that in our method, the label of node *u* is only affected by valuable nodes in V^{l}_{u}, rather than its local neighbors. Therefore, decrease of labeled neighbors will have minor effect on classification performance, making BCC more suitable for handling sparse labeling problem. Moreover, we can see that the proposed method does not rely on the homophily assumption, so it can be applied to network with lower homophily as well.

CONCLUSION

In order to improve classification accuracy in sparsely labeled networks, we propose a novel behavior based collective classification method, BCC, in this study. In BCC, the behavior feature of nodes is extracted for classification, which has shown more discriminative ability to traditional methods. Then, instead of using all the labeled nodes, we screen the most-relevant nodes according to the calculation of correlation and similarity, which can overcome the effects of noise and imbalanced dataset. Finally, collective inference is introduced to utilize both labeled nodes and unlabeled nodes, which can relieve the sparse labeling problem effectively.

REFERENCES:

1. Aha, D.W., Kibler, D., & Albert, M.K. (1991), 'Instance-based learning algorithms', Machine Learning, vol. 6, no. 1, pp. 37-66.
2. Boutell, M.R., Luo, J., Shen, X. & Brown, C.M. (2004), 'Learning multi-label scene classification', Pattern Recognition, vol. 37, no. 9, pp. 1757-71
3. Chang, C.-C., & Lin, C.-J. (2004), 'LIBSVM : a library for support vector machines', Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm
4. Clare, A. & King, R.D. (2001), 'Knowledge Discovery in Multi-Label Phenotype Data', paper presented to Proceedings of the 5th European Conference on Principles of Data Mining and Knowledge Discovery (PKDD 2001), Freiburg, Germany.
5. Diplaris, S., Tsoumakas, G., Mitkas, P. & Vlahavas, I. (2005), 'Protein Classification with Multiple Algorithms', paper presented to Proceedings of the 10th Panhellenic Conference on Informatics (PCI 2005), Volos, Greece, November.
6. Elisseeff, A. & Weston, J. (2002), 'A kernel method for multi-labelled classification', paper presented to Advances in Neural Information Processing Systems 14.
7. Freund, Y. & Schapire, R.E. (1997), 'A decision-theoretic generalization of on-line learning and an application to boosting', Journal of Computer and System Sciences, vol. 55, no. 1, pp. 119-39.
8. Godbole, S. & Sarawagi, S. (2004), 'Discriminative Methods for Multi-labeled Classification', paper presented to Proceedings of the 8th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2004).