Bidirectional Semi-supervised Learning with Graphs

Tomoharu Iwata[1] and Kevin Duh[2]

[1] NTT Communication Science Laboratories
iwata.tomoharu@lab.ntt.co.jp
[2] Nara Institute of Science and Technology
kevinduh@is.naist.jp

Abstract. We present a machine learning task, which we call bidirectional semi-supervised learning, where label-only samples are given as well as labeled and unlabeled samples. A label-only sample contains the label information of the sample but not the feature information. Then, we propose a simple and effective graph-based method for bidirectional semi-supervised learning in multi-label classification. The proposed method assumes that correlated classes are likely to have the same labels among the similar samples. First, we construct a graph that represents similarities between samples using labeled and unlabeled samples in the same way with graph-based semi-supervised methods. Second, we construct another graph using labeled and label-only samples by connecting classes that are likely to co-occur, which represents correlations between classes. Then, we estimate labels of unlabeled samples by propagating labels over these two graphs. We can find a closed-form global solution for the label propagation by using matrix algebra. We demonstrate the effectiveness of the proposed method over supervised and semi-supervised learning methods with experiments using synthetic and multi-label text data sets.

Keywords: semi-supervised learning, label propagation, multi-label classification.

1 Introduction

The performance of a classifier can be improved as the number of labeled samples is increased. However, we might not have enough labeled samples to achieve a reasonable performance because their generation incurs cost and requires time. To overcome the shortage of labeled samples, there has been great interest in methods that effectively increase training samples. For example, semi-supervised learning [1] augments training samples by using unlabeled samples. The other examples include domain adaptation [2] and class adaptation [3], where the former utilizes samples from different domains and the latter utilizes samples from different taxonomies.

In this paper, we consider a new way to improve multi-label classification performance, where we have label-only samples as well as labeled and unlabeled

P. Flach et al. (Eds.): ECML PKDD 2012, Part II, LNCS 7524, pp. 293–306, 2012.

Table 1. Notation

Symbol	Description
$\{(\boldsymbol{x}, \boldsymbol{y})\}$	labeled samples
$\{\boldsymbol{x}\}$	unlabeled samples
$\{\boldsymbol{y}\}$	label-only samples
N	number of labeled samples
U	number of unlabeled samples
O	number of label-only samples
K	number of classes
M	number of features

Table 2. Given data in supervised, semi-supervised and bidirectional semi-supervised learning

Problem	Given data
supervised	$\{(\boldsymbol{x}, \boldsymbol{y})\}$
(one-directional) semi-supervised	$\{(\boldsymbol{x}, \boldsymbol{y})\}, \{\boldsymbol{x}\}$
bidirectional semi-supervised	$\{(\boldsymbol{x}, \boldsymbol{y})\}, \{\boldsymbol{x}\}, \{\boldsymbol{y}\}$

samples. The labeled samples are a set of pairs of feature and label vectors $\{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^{N}$, and the unlabeled samples are a set of feature vectors $\{\boldsymbol{x}_i\}_{i=N+1}^{N+U}$ without label information. The label-only samples are a set of label vectors $\{\boldsymbol{y}_i\}_{i=N+U+1}^{N+U+O}$, where corresponding feature information is unavailable. We call this setting *bidirectional semi-supervised learning*. Semi-supervised learning is defined as a task with abundant inputs (unlabeled samples) but few input-output pairs (labeled samples); so we define bidirectional semi-supervised learning as a task with abundant inputs and abundant outputs (label-only samples), but few input-output pairs. Table 1 summaries our notation, and Table 2 summaries the given data in supervised, semi-supervised and bidirectional semi-supervised learning.

This setting can be found in many real applications. For example, let us consider recommendation problems in two different domains, where the feature vector is a user's preference for items in a given domain, and the label vector is the user's binary preference for items in another domain [4, 5]. Here, we suppose that we want to estimate the preferences in the second domain. Users who have preferences in the both domains can be used for labeled samples, those who have preferences only in the first domain can be used for the unlabeled samples, and those who have preferences only in the second domain can be used for the label-only samples. Cross-lingual information retrieval [6–8] is another application, where the feature vector is a query and the label vector consists of relevant documents in a different language. Here, we might have a lot of documents in a different language for label-only samples. Other applications include automatic image annotation problems [9–11], where the feature vector is image features and the label vector is the annotations. The label-only samples can be obtained

using text corpus under the assumption that correlation between annotations is related to correlation between word in the text corpus. Similarly, in multi-label text classification problems, label correlation might be available from outside label-only sources. In general, bidirectional semi-supervised problems may arise when we have disjoint datasets of features and labels.

Unlabeled samples contain information about the distribution of samples in the feature space. Semi-supervised learning methods uses this information for improving performance. Similarly, label-only samples contain information about the distribution in the label space, or information about correlations between classes. Therefore, we can expect that the multi-label classifier performance can be improved by using label-only samples.

We propose a simple and effective graph-based method for bidirectional semi-supervised learning. A number of graph-based semi-supervised learning methods, or label propagation, have been proposed [12–15] because of its simplicity and easy implementation. They have used for a wide variety of applications, such as text classification [16], image recognition [17] and protein function prediction [18]. With the graph-based semi-supervised method, a graph is constructed using labeled and unlabeled samples by connecting samples that have similar feature vectors, where each node corresponds to a sample. Then, labels are estimated by propagating labels over the constructed graph with the assumption that connected samples tend to have the same label. An advantage of graph-base methods is that we can obtain a global closed-form solution.

The proposed method is an extension of the graph-based semi-supervised methods. First, we construct a graph using labeled and unlabeled samples in the similar way with graph-based semi-supervised learning. Second, we construct another graph using labeled and label-only samples by connecting classes that are likely to co-occur, where each node corresponds to a class. Then, we estimate labels by using these two graphs with the assumption that labels of correlated classes in similar samples tend to be the same. We can obtain a global closed-form solution for the proposed method. We can use similar techniques that have been extensively studied for graph-based semi-supervised methods for the proposed method, such as techniques for constructing effective graphs and algorithms for efficient estimation.

The remainder of this paper is organized as follows. In Section 2, we formulate the proposed method, and describe a closed-form solution and an iterative estimation method. In Section 3, we briefly review related work. In Section 4, we demonstrate the effectiveness of the proposed method with experiments using synthetic and multi-label text data sets. Finally, we present concluding remarks and a discussion of future work in Section 5.

2 Proposed Method

We suppose that there are N labeled samples $\{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^{N}$, U unlabeled samples $\{\boldsymbol{x}_i\}_{i=N+1}^{N+U}$, and O label-only samples $\{\boldsymbol{y}_i\}_{i=N+U+1}^{N+U+O}$. A feature vector is represented by $\boldsymbol{x}_i = (x_{im})_{m=1}^{M}$, where x_{im} is the mth element of the ith sample's

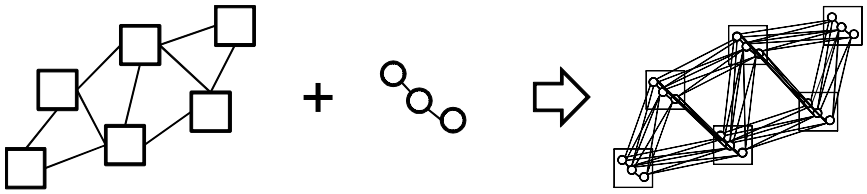


Fig. 1. Sample, class and sample-class graphs. Square and circle nodes represent sample and class, respectively.

feature vector, and M is the number of features. A label vector is represented by $\boldsymbol{y}_i = (y_{ik})_{k=1}^K$, where $y_{ik} = 1$ if the ith sample is categorized into class k, $y_{ik} = -1$ otherwise, $y_{ik} \in \{-1, 1\}$, and K is the number of classes. Each sample can be assigned to multiple classes. Classes that do not appear in labeled samples can appear in label-only samples. Our task is to assign labels to unlabeled samples.

First, we construct a sample graph, where nodes are the labeled and unlabeled samples. An edge between two nodes represents the similarity of feature vectors of the two samples. The edge weight can be calculated by using Gaussian kernel as follows,

$$w_{ij} = \exp\left(-\frac{\alpha}{2} \parallel \boldsymbol{x}_i - \boldsymbol{x}_j \parallel^2\right), \tag{1}$$

where α is the precision parameter. We can also build the sample graph with k nearest neighbors, where nodes are connected if they are k nearest neighbors in Euclidean distance, and $w_{ij} = 0$ otherwise. Figure 1 (a) shows an example of a sample graph, where a square node represents a sample.

Second, we construct a class graph, where nodes are the classes that appear in the labeled and label-only samples. An edge between two nodes represents the similarity of the two classes, or how likely the two classes co-occur. The edge weight can be calculated by using Gaussian kernel in the similar way to the sample graph,

$$v_{kl} = \exp\left(-\frac{\beta}{2} \parallel \boldsymbol{y}^{(k)} - \boldsymbol{y}^{(l)} \parallel^2\right), \tag{2}$$

where β is the precision parameter, and $\boldsymbol{y}^{(k)} = (y_{1k}, \cdots, y_{Nk}, y_{N+U+1,k}, \cdots, y_{N+U+O,k})$ is an $N + O$ dimensional vector that consists of the kth elements of label vectors in the labeled and label-only samples. $\boldsymbol{y}^{(k)}$ can be used with $L2$ normalization so that the weights correlate to their cosine similarities. Figure 1 (b) shows an example of a class graph, where a circle node represents a class. Note that there are $N + U$ nodes in the sample graph, and K nodes in the class graph.

Then, we estimate labels for unlabeled samples using the sample and class graphs. We suppose that f_{ik} is a real valued relaxation of y_{ik}, which is to be estimated. We assume that labels of correlated classes (which are connected in

the class graph) in similar samples (which are connected in the sample graph) are likely to be similar. This can be achieved by minimizing the following function,

$$E = \frac{1}{2} \sum_{i,j=1}^{N+U} \sum_{k,l=1}^{K} w_{ij} v_{kl} (f_{ik} - f_{jl})^2, \tag{3}$$

with the constraint of $f_{ik} = y_{ik}$ on the labeled data $i = 1, \cdots, N$. When w_{ij} is high (feature vectors of i and j are similar) and v_{kl} is high (classes k and l are correlated), the estimated value for class k of the ith sample needs to be similar to that for class l of the jth sample so as to minimize the objective function E. Therefore, by minimizing the objective function, we can find estimated labels, where correlated classes in similar samples have similar labels.

The proposed method can be seen as label propagation on a sample-class graph that is build by combining sample and class graphs as shown in Figure 1 (c), where the correlated classes in the similar samples are connected. In the sample-class graph, each node corresponds to a class of each sample, and the number of nodes is $(N + U)K$.

With graph-based semi-supervised learning methods, the following objective function is minimized,

$$E = \frac{1}{2} \sum_{i,j=1}^{N+U} w_{ij} \sum_{k=1}^{K} (f_{ik} - f_{jk})^2, \tag{4}$$

where they assume that similar samples have similar labels. However, they do not consider the correlation between classes. The graph-based semi-supervised learning methods can be seen as label propagation on a sample graph without class graphs. The proposed method with $v_{kl} = 1$ if $k = l$ and $v_{kl} = 0$ otherwise coincides with the graph-based semi-supervised method.

2.1 Closed-Form Solution

We can find a closed-form global solution for the minimization of the objective function E by using matrix algebra. The proposed method propagates labels on the sample-class graph as shown in Figure 1. Therefore, we can use the same algorithm for finding the solution with label propagation for semi-supervised learning [15]. Let $\boldsymbol{A}$ be an $(N+U)K \times (N+U)K$ matrix, whose $(iK+k, jK+l)$th element is $A_{iK+k,jK+l} = w_{ij} v_{kl}$ as follows,

$$\boldsymbol{A} = \begin{pmatrix} w_{11}v_{11} & w_{11}v_{12} & \cdots & w_{11}v_{1K} \\ w_{11}v_{21} & w_{11}v_{22} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ w_{N+U,1}v_{K1} & \cdots & \cdots & w_{N+U,N+U}v_{KK} \end{pmatrix}, \tag{5}$$

which represents the sample-class graph. Let $\boldsymbol{D}$ be a diagonal matrix, whose ith diagonal element is

$$D_{ii} = \sum_{j=1}^{(N+U)K} A_{ij}. \tag{6}$$

The graph Laplacian matrix $\boldsymbol{L}$ is defined as,

$$\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}. \tag{7}$$

Let $\boldsymbol{f} = (f_{11}, f_{12}, \cdots, f_{1K}, f_{21}, \cdots, f_{N+U,K})^\top$ be the vector of f values on all of the labeled and unlabeled samples. The objective function E can be written as,

$$E = \boldsymbol{f}^\top \boldsymbol{L} \boldsymbol{f}. \tag{8}$$

Then, we can obtain the following closed-form solution by solving the constrained optimization problem using Lagrange multipliers,

$$\boldsymbol{f}_l = \boldsymbol{y}_l, \tag{9}$$

$$\boldsymbol{f}_u = -\boldsymbol{L}_{uu}^{-1} \boldsymbol{L}_{ul} \boldsymbol{y}_l. \tag{10}$$

Here, $\boldsymbol{f} = (\boldsymbol{f}_l, \boldsymbol{f}_u)$ and

$$\boldsymbol{L} = \begin{pmatrix} \boldsymbol{L}_{ll} & \boldsymbol{L}_{lu} \\ \boldsymbol{L}_{ul} & \boldsymbol{L}_{uu} \end{pmatrix}, \tag{11}$$

which are partitioned with respect to values of labeled and unlabeled samples.

2.2 Iterative Algorithm

We can also obtain a solution by using an iterative algorithm [19, 20]. When the graphs are sparse, the iterative algorithm is efficient and reduces required memory. Sparse graphs can be obtained by using k nearest neighbor graph construction for sample and class graphs. With the closed-form solution, we need the inverse of a $UK \times UK$ matrix $\boldsymbol{L}_{uu}$, whose inverse is not sparse even if $\boldsymbol{L}_{uu}$ is sparse. The numbers of unlabeled samples U and classes K might be large, and it might require a huge memory space for storing the $UK \times UK$ dense matrix. On the other hand, the iterative algorithm does not need to calculate any dense matrices. First, we initialize estimates as follows,

$$\boldsymbol{f}_l^{(0)} \leftarrow \boldsymbol{y}_l, \tag{12}$$

$$\boldsymbol{f}_u^{(0)} \leftarrow (0, 0, \cdots, 0). \tag{13}$$

Then, we iterate the following updates until convergence,

$$\boldsymbol{f}^{(t+1)} \leftarrow \boldsymbol{D}^{-1} \boldsymbol{A} \boldsymbol{f}^{(t)}, \tag{14}$$

$$\boldsymbol{f}_l^{(t+1)} \leftarrow \boldsymbol{y}_l, \tag{15}$$

where $\boldsymbol{f}^{(t)}$ is the estimates of the tth iteration. We can find the unique fixed point by the iterative algorithm.

2.3 Induction

The method described before is transductive, which means that the method estimates labels of the given unlabeled samples. When we newly obtain unlabeled samples to be estimated, we can estimate their labels by combining previously given samples and the newly obtained samples. However, it is inefficient. We can efficiently estimate labels for out of samples using the estimation result for the previously given samples as follows,

$$f_{ik} = \frac{\sum_{j=1}^{N+U} \sum_{l=1}^{K} w_{ij} v_{kl} f_{jl}}{\sum_{j=1}^{N+U} \sum_{l=1}^{K} w_{ij} v_{kl}}, \tag{16}$$

where f_{jl} is the estimated labels of the previously given samples.

3 Related Work

Bidirectional semi-supervised learning is a new type of machine learning task, and the proposed method is a simple method based on graphs. A lot of graph-based semi-supervised learning methods have been proposed. However, most of them are for single-label classification. Those methods can be applied to multi-label classification by estimating each label independently. However, by considering the class interdependence for multi-label classification, the performance can be improved and some those types of graph-based semi-supervised learning methods have been proposed [21–23, 4]. For example, [22] proposed a method that estimates labels so that multiple labels for each sample satisfy the given correlations between classes. Other multi-label classifiers also uses class correlation [24]. However, they utilize only classes that appeared in the labeled samples. On the other hand, the proposed method can utilize classes that do not appear in the labeled samples by representing class correlation by a graph, and propagate labels over the graph. With the proposed method, even if two classes do not directly co-occur, label information can propagate through edges. In [21], correlation between labels for each sample is considered. In contrast, the proposed method considers correlation between labels not only in one sample but also in multiple similar samples.

4 Experiments

4.1 Data

We demonstrate the effectiveness of the proposed method using the following two data sets: Swissroll and Patent.

The Swissroll data [25, 26] are synthetic, where samples of three dimensional feature vectors are lying on a two dimensional nonlinear manifold as shown in Figure 2. We augmented the swissroll data set with multiple labels. We generated label vectors so that they became similar if their feature vectors were located

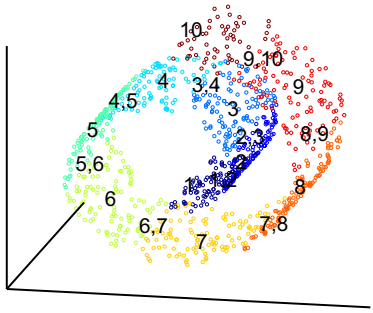


Fig. 2. Swissroll data. Each point represents a feature vector and its color represents the labels. The numbers show that nearby samples are assigned to those classes.

closely in the two dimensional nonlinear manifold, where we set the number of classes at $K = 10$. We generated samples, where the number of labeled, unlabeled and label-only samples were $N = 10$, $U = 1,000$ and $O = 1,000$ respectively.

The Patent data consist of patents published in Japan from January to March in 2004, to which International Patent Classification (IPC) codes were attached by experts according to their content. We used bag-of-words of a patent for the feature vector, where the number of words was $M = 104,621$, and we normalized each feature vector by $L2$ norm. We used the most frequently occurred 500 IPC codes in the corpus for the classes, $K = 500$. We sampled 10 labeled samples, $1,000$ unlabeled samples, and $5,000$ label-only samples from the corpus.

Figure 3 shows a class graph of the Patent data set. Here, each node represents a class, and it is visualized by [27] so that connected nodes are located closely. Some classes form clusters, and we can see the structure of classes from the visualization result. In the Patent data set, there are correlated classes, such as 'transmitter' and 'receiver', 'distinct material semiconductor' and 'distinct alignment semiconductor', and 'system to generate signals for adjusting focus' and 'automatic focus adjusting system'.

4.2 Measurements

For the evaluation measurements, we used mean reciprocal rank (MRR) and normalized discounted cumulative gain (NDCG) [28], which were widely used in evaluating ranking problems. We used ranking measurements because they give higher scores when true classes are ranked higher than false classes even if the estimated classes did not exactly match with the true classes. They were also used for multi-label classification [29–31].

The MRR is the average of the reciprocal ranks, and the MRR of the ith sample is given as follows,

$$MRR_i = \frac{1}{|\boldsymbol{y}_i|} \sum_{k:y_{ik}=1} \frac{1}{\mathrm{rank}_{ik}}, \tag{17}$$

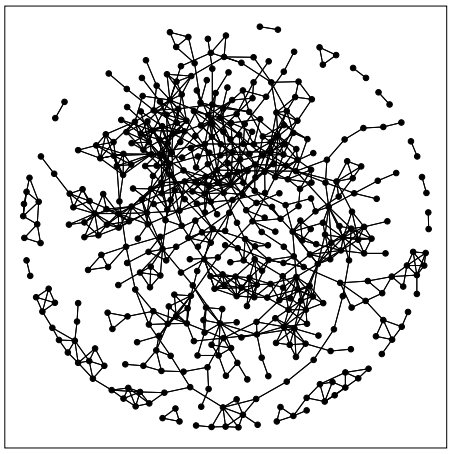


Fig. 3. A class graph of the Patent data set, where each node represents a class or IPC code

where rank_{ik} is the rank of class k of the ith sample in the estimated result, and $|\boldsymbol{y}_i|$ represents the number of classes that satisfy $y_{ik} = 1$.

The DCG is calculated as follows,

$$DCG_i = g_{i1} + \sum_{k=2}^{K} \frac{g_{ik}}{\log_2 k}, \tag{18}$$

where $g_{ik} = 1$ if the kth ranked estimated class is the true class for the ith sample, and $g_{ik} = 0$ otherwise. NDCG can be obtained by normalizing DCG by the maximum possible DCG, and NDCG lies on the interval 0.0 to 1.0. With the proposed method, classes were ranked according to values f_{ik} for each unlabeled sample. Higher MRR and NDCG represent better classification performance.

4.3 Compared Methods

We compared the proposed method, which uses all of the labeled, unlabeled and label-only samples, with a supervised method, which uses only labeled samples, and a semi-supervised method, which uses labeled and unlabeled samples.

For the supervised method (SL), we used a maximum entropy model, which is a discriminative classifier, and it has achieved high performance for text classification [32]. The maximum entropy model estimates the probability distribution that maximizes entropy under the constraints imposed by the given data. The probability that the ith sample is classified into class k is calculated as follows,

$$P(k|i) = \frac{\exp(\boldsymbol{\theta}_k^\top \boldsymbol{x}_i)}{\sum_{l=1}^{K} \exp(\boldsymbol{\theta}_l^\top \boldsymbol{x}_i)}, \tag{19}$$

Table 3. Average MRR and NDCG in the Swissroll data set and their standard deviation

	MRR	NDCG
SL	0.520 ± 0.049	0.738 ± 0.047
SSL	0.602 ± 0.048	0.814 ± 0.053
Proposed	$\mathbf{0.675 \pm 0.047}$	$\mathbf{0.900 \pm 0.039}$

Table 4. Average MRR and NDCG in the Patent data set and their standard deviation

	MRR	NDCG
SL	0.029 ± 0.021	0.157 ± 0.021
SSL	0.025 ± 0.018	0.153 ± 0.019
Proposed	$\mathbf{0.034 \pm 0.023}$	$\mathbf{0.166 \pm 0.026}$

where $\boldsymbol{\theta}_k$ is a parameter vector for class k. The labels were ranked according this estimated probability. The parameters can be obtained using maximum a posteriori (MAP) estimation with Gaussian priors. We chose the hyper-parameters for the Gaussian priors from $\{10^{-2}, 10^{-1}, 1\}$ that achieved the best performance.

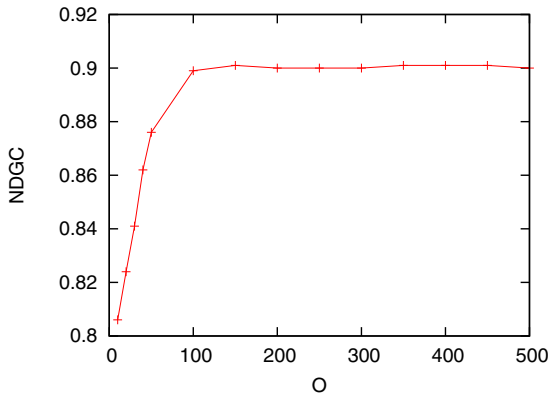
For the semi-supervised method (SSL), we used a graph-based semi-supervised method, or label propagation [12]. The graph-based semi-supervised method coincides with the proposed method when the class graph is constructed with $v_{kl} = 1$ if $k = l$, and $v_{kl} = 0$ otherwise as described before. We set the precision parameter for Gaussian kernel at $\alpha = 1$, and the number of neighbors at 10 when we construct the graph.

With the proposed method, we set the precision parameters for Gaussian kernel at $\alpha = 1$ and $\beta = 1$, and the number of neighbors at 10 when we construct both of the sample and class graphs. With the semi-supervised and proposed method, we estimated labels using iterative algorithms.

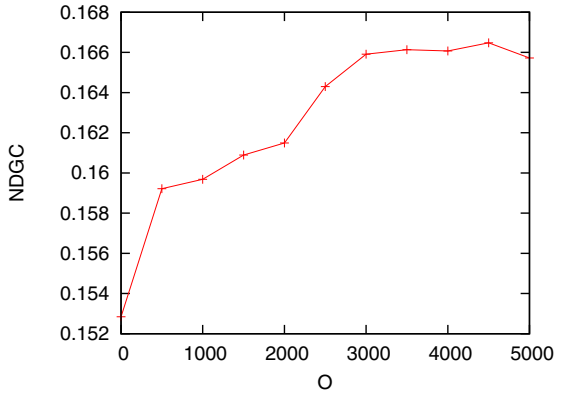
4.4 Results

Tables 3 and 4 show the averages of MRR and NDCG and their standard deviations over 100 experiments with Swissroll and Patent data sets, respectively. The proposed method achieved the best performance in both data sets. This result indicates that the proposed method can appropriately assign labels through its use of label-only samples as well as unlabeled samples.

Figure 4 shows NDCGs achieved by the proposed method with different numbers of label-only samples in Swissroll and Patent data sets. As the number of label-only samples increases, the NDCG increases. The MRR showed the same tendency of the NDCG. This result implies that the proposed method can obtain relationships between classes more precisely by using more label-only samples.



(a) Swissroll



(b) Patent

Fig. 4. NDCG with different numbers of label-only samples

5 Conclusion

We presented bidirectional semi-supervised learning, which is a novel machine learning task to improve performance by using label-only samples as well as labeled and unlabeled samples. We then proposed a simple and effective graph-based method for bidirectional semi-supervised learning. The proposed method assumes that correlated classes are likely to have the same labels among the similar samples. The correlated classes can be found by using labeled and label-only samples, and the similar samples can be found by using labeled and unlabeled samples. In experiments with synthetic and text data sets, we confirmed that the proposed method can improve the performance of multi-label classification.

Although our results have been encouraging as a first step towards bidirectional semi-supervised learning, we must extend our approach in a number of directions. First, we can extend the proposed method by using more advanced

techniques for graph-based semi-supervised learning because the proposed method uses the same framework with the graph-based semi-supervised learning methods. Examples include methods for learning weight matrices [12], and efficient algorithms for label estimation [33, 34].

Second, we need to investigate methods for bidirectional semi-supervised learning other than the proposed graph-based method. In the semi-supervised learning, a wide variety of methods have been proposed such as transductive SVMs [35, 36], methods using generative models [1, 37] as well as graph-based methods. These methods might be helpful for considering new bidirectional semi-supervised learning methods.

Finally, we would like to evaluate the proposed method in other real applications. Application examples include collaborative filtering [4, 5], cross-lingual information retrieval [6–8], and image annotation [10, 11].

References

1. Nigam, K., McCallum, A.K., Thrun, S., Mitchell, T.M.: Text classification from labeled and unlabeled documents using EM. Machine Learning 39(2/3), 103–134 (2000)
2. Daume III, H., Marcu, D.: Domain adaptation for statistical classifiers. Journal of Artificial Intelligence Research 26, 101–126 (2006)
3. Iwata, T., Tanaka, T., Yamada, T., Ueda, N.: Improving classifier performance using data with different taxonomies. IEEE Transactions on Knowledge and Data Engineering 23(11), 1668–1677 (2011)
4. Li, T., Yan, S., Kweon, T.M.I.S.: Local-driven semi-supervised learning with multi-label. In: IEEE International Conference on Multimedia and Expo., ICME 2009, pp. 1508–1511 (2009)
5. Nakatsuji, M., Fujiwara, Y., Tanaka, A., Uchiyama, T., Ishida, T.: Recommendations over domain specific user graphs. In: Proceeding of the 2010 Conference on ECAI 2010: 19th European Conference on Artificial Intelligence, pp. 607–612 (2010)
6. Dumais, S.T., Landauer, T.K., Littman, M.L.: Automatic cross-linguistic information retrieval using latent semantic indexing. In: Proceedings of Workshop on Cross-Linguistic Information Retrieval in SIGIR 1996, pp. 16–23 (1996)
7. Xu, J., Weischedel, R., Nguyen, C.: Evaluating a probabilistic model for cross-lingual information retrieval. In: Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2001, pp. 105–110 (2001)
8. Platt, J.C., Toutanova, K., Yih, W.: Translingual document representations from discriminative projections. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, EMNLP 2010, pp. 251–261 (2010)
9. Blei, D.M., Jordan, M.I.: Modeling annotated data. In: SIGIR 2003: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 127–134 (2003)
10. Socher, R., Fei-Fei, L.: Connecting modalities: Semi-supervised segmentation and annotation of images using unaligned text corpora. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pp. 966–973 (2010)

11. Kimura, A., Kameoka, H., Sugiyama, M., Nakano, T., Maeda, E., Sakano, H., Ishiguro, K.: SemiCCA: Efficient semi-supervised learning of canonical correlations. In: Proceedings of IAPR International Conference on Pattern Recognition, ICPR 2010, pp. 2933–2936 (2010)
12. Zhu, X., Ghahramani, Z., Lafferty, J.: Semi-supervised learning using Gaussian fields and harmonic functions. In: Proceedings of the 20th International Conference on Machine Learning, ICML 2003, pp. 912–919 (2003)
13. Zhou, D., Bousquet, O., Lal, T.N., Weston, J., Scholkopf, B.: Learning with local and global consistency. In: Advances in Neural Information Processing Systems 16, pp. 321–328. MIT Press (2004)
14. Belkin, M., Matveeva, I., Niyogi, P.: Regularization and Semi-supervised Learning on Large Graphs. In: Shawe-Taylor, J., Singer, Y. (eds.) COLT 2004. LNCS (LNAI), vol. 3120, pp. 624–638. Springer, Heidelberg (2004)
15. Zhu, X., Goldberg, A.B.: Introduction to Semi-Supervised Learning. In: Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool (2009)
16. Subramanya, A., Bilmes, J.: Soft-supervised learning for text classification. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP 2008, pp. 1090–1099. Association for Computational Linguistics, Stroudsburg (2008)
17. Cao, L., Luo, J., Huang, T.S.: Annotating photo collections by label propagation according to multiple similarity cues. In: Proceedings of the 16th ACM International Conference on Multimedia, MM 2008, pp. 121–130. ACM, New York (2008)
18. Kato, T., Kashima, H., Sugiyama, M.: Robust label propagation on multiple networks. IEEE Transactions on Neural Networks 20(1), 35–44 (2009)
19. Zhu, X., Ghahramani, Z.: Learning from labeled and unlabeled data with label propagation. Technical report, Carnegie Mellon University (2002)
20. Bengio, Y., Delalleau, O., Roux, N.L.: Label propagation and quadratic criterion. In: Chapelle, O., et al. (eds.) Semi-Supervised Learning, MIT Press, Cambridge (2006)
21. Wang, J., Zhao, Y., Wu, X., Hua, X.S.: Transductive multi-label learning for video concept detection. In: Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval, MIR 2008, pp. 298–304. ACM, New York (2008)
22. Zha, Z.J., Mei, T., Wang, J., Wang, Z., Hua, X.S.: Graph-based semi-supervised learning with multiple labels. J. Vis. Comun. Image Represent. 20, 97–103 (2009)
23. Liu, W., Chang, S.F.: Robust multi-class transductive learning with graphs. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 381–388 (2009)
24. Liu, Y., Jin, R., Yang, L.: Semi-supervised multi-label learning by constrained nonnegative matrix factorization. In: Proceedings of the 21st National Conference on Artificial Intelligence, AAAI 2006, pp. 421–426. AAAI Press (2006)
25. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science 290(5500), 2319–2323 (2000)
26. Roweis, S., Saul, L.: Nonlinear dimensionality reduction by locally linear embedding. Science 290(5500), 2323–2326 (2000)
27. Matsubayashi, T., Yamada, T.: The hierarchical individual timestep method for large-scale graph drawing. In: Proc. of the 15th International Symposium on Graph Drawing, GD 2007 (2007)
28. Järvelin, K., Kekäläinen, J.: Cumulated gain-based evaluation of IR techniques. ACM Trans. Inf. Syst. 20, 422–446 (2002)

29. Lin, X., Chen, X.W.: Mr.KNN: soft relevance for multi-label classification. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, CIKM 2010, pp. 349–358. ACM, New York (2010)
30. Angelova, R., Kasneci, G., Weikum, G.: Graffiti: graph-based classification in heterogeneous networks. In: World Wide Web (2011)
31. Yang, Y., Gopal, S.: Multilabel classification with meta-level features in a learning-to-rank framework. Machine Learning (2011)
32. Nigam, K., Lafferty, J., McCallum, A.: Using maximum entropy for text classification. In: Proceedings of IJCAI 1999 Workshop on Machine Learning for Informat ion Filtering, pp. 61–67 (1999)
33. Garcke, J., Griebel, M.: Semi-supervised learning with sparse grids. In: Proc. of the 22nd ICML Workshop on Learning with Partially Classified Training Data (2005)
34. Mahdaviani, M., Freitas, O.D., Fraser, B., Hamze, F.: Fast computational methods for visually guided robots. In: Proc. of the International Conference on Robotics and Automation, ICRA 2005 (2005)
35. Bennett, K.P., Demiriz, A.: Semi-supervised support vector machines. In: Advances in Neural Information Processing Systems, pp. 368–374. MIT Press (1999)
36. Fung, G., Mangasarian, O.L.: Semi-supervised support vector machines for unlabeled data classification. Optimization Methods and Software 15, 29–44 (2001)
37. Fujino, A., Ueda, N., Saito, K.: Semisupervised learning for a hybrid generative/discriminative classifier based on the maximum entropy principle. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(3), 424–437 (2008)