Analysis of Classification Algorithm on Hypergraph

Linli Zhu[1*] and Wei Gao[2]

[1]School of Computer Engineering, Jiangsu University of Technology, Changzhou, China, 213001; [2]School of Information, Yunnan Normal University, Kunming, China,650500

Abstract: Classification learning problem on hypergraph is an extension of multi-label classification problem on normal graph, which divides vertices on hypergraph into several classes. In this paper, we focus on the semi-supervised learning framework, and give theoretic analysis for spectral based hypergraph vertex classification semi-supervised learning algorithm. The generalization bound for such algorithm is determined by using the notations of zero-cut, non-zero-cut and pure component. Furthermore, we derive a generalization performance bound for near-zero-cut partition with optimal parameter λ .

Keywords: Classification algorithm, hypergraph, pure component, pure sub-hypergraph, scaling factor, tuning parameter.

1. INTRODUCTION

Spectral clustering of weighted graph, an important component of spectral graph theory and statistic learning theory, has been widely used in many fields, such as computer network, data mining, image segmentation and ontology similarity computation (see [1-7]). Hypergraph is a subset system for limited set, which is the most general discrete structure, and it is the generalization of the common graph. For many practical problems, adopting the concept of hypergraph is more usefully than adopting the concept of graph. At present, the model of hypergraph has been applied in many fields, such as: VLSI layout, electricity network topology analysis. The goal of classification algorithm on hypergraph is to divide vertices into several classes, and spectral method may play a key role in it (see [8-11]). Some applications on such classification can be referred in [12-18].

Let $V=\{v_1,v_2,\ldots,v_m\}$ be a limited set, E is family of subset of V, *i.e.*, $E\subseteq 2^V$. Then $H=(V,E)$ is a hypergraph on V. the element of V is called a vertex, the elements of E is called a hyperedge. Let $|V|$ be the order of H, $|E|$ be the scale of H. $|e|$ is basic number of hyperedge e. $r(H)=\max_j |e_j|$ is rank of hyperedge e, and $s(H)= \min_j |e_j|$ lower rank of hyperedge e. If $|e|=k$ for each hyperedge e of E (that is $r(H)=s(H)=k$), then H is a k-uniform hypergraph. If k=2, then H is just a normal graph.

A hypergraph H is called a simple hypergraph or a sperner hypergraph, if any two hyperedges are not contained with each other. Let $H^{'}=(V,E^{'})$ is a hypergraph on V, if $E^{'}\subset E$, then $H^{'}$ is a part-hypergraph of H. For $S\subseteq V$, $H[S]=\{e\in E:e\subseteq S\}$ is called a sub-hypergraph of H induced by S.

Hypergraph H can be represented by graph by using the set of vertices to represent the elements of V. If $|e_j|=2$, using a continuous curve which attach to the elements of e_j to representing e_j; If $|e_j|=1$, using a loop which contains e_j to represent e_j; If $|e_j|\geq 3$, using a simple close curve which contains all the elements of e_j to represent e_j.

In this paper, we assume H is a weighted hypergraph, each edge given a weight $w(e)$. The degree of vertex v_j in hypergraph H is denoted as

$$\deg_j(H) = \sum_{e\in E} w(e)h(v,e),$$

where,

$$h(v,e) = \begin{cases} 1, & if \quad v\in e \\ 0, & if \quad v\notin e \end{cases}.$$

Let $\delta(e)=\sum_{v\in V} h(v,e)$. Then, the normalized laplacian $L(H)\in \mathbb{R}^{m\times m}$ on hypergraph H is defined by:

$$L_{ij}(H) = \begin{cases} -\sum_{\{i,j\}\subseteq e} w(e)\dfrac{1}{\delta(e)} & i\neq j \\ \deg_j(H) & \text{otherwise} \end{cases}.$$

Fixed m scaling factor $\boldsymbol{S}_j$ (j=1,…,m) (Normally, we can choose $\boldsymbol{S}=\boldsymbol{I}$, or $\boldsymbol{S}_j=\deg_j(H)$), Let $\boldsymbol{S}=\mathrm{diag}(\{\boldsymbol{S}_j\})$. Then $\boldsymbol{S}$-normalized Laplacian on hypergraph is given as:

$$L_S(H) = \boldsymbol{S}^{-1/2}L(H)\boldsymbol{S}^{-1/2}.$$

The corresponding regularization is relying on:

$$\boldsymbol{f}_{\cdot,k}^T L_{\boldsymbol{S}}(H)\boldsymbol{f}_{\cdot,k} = \frac{1}{2}\sum_{e\in E}\frac{1}{\delta(e)}\sum_{\{j,j'\}\subseteq e} w(e)\left(\frac{f_{j,k}}{\sqrt{S_u}} - \frac{f_{j',k}}{\sqrt{S_v}}\right)^2 .$$

The definition of function $\boldsymbol{f}$ will be given later. It implies that, if vertices j, j' contained in the same hyperedge, then they have higher similarity and it is possible to be classified in the same class.

Now, we give the following definition for component on hypergraph: A sub-hypergraph $H_0=(V_0, E_0)$ of H is called pure component, if H_0 is connected, E_0 is induced by restricting E on V_0, and the labels y have identical values on V_0. A pure sub-hypergraph $H'=\bigcup_{l=1}^{q} H_l$ of H divides V into q disjoint sets $V=\bigcup_{l=1}^{q} V_l$ such that each sub-hypergraph $H_l=(V_l, E_l)$ is a pure component. Denote by $\lambda_i(H_l)=\lambda_i(L(H_l))$ the i-th smallest eigenvalue of $L(H_l)$.

For example, if all hyperedges of H which connect vertices with different labels are removed, then the resulting sub-hypergraph is a pure sub-hypergraph. For each pure component H_l, its first eigenvalue $\lambda_1(H_l)$ is always zero. The second eigenvalue $\lambda_2(H_l)>0$ since H_l is connected.

Let Y be output space, which contain K possible values, and each value represents a class. Each vertex v_j on hypergraph corresponding to a output value y_j. $Z_n=\{j_i:1\le i\le n\}$ is n indices random draw from $\{1,\ldots,m\}$ uniformly and without replacement. Manually label the n vertices v_{j_i} with labels $y_{j_i}\in Y$, and then automatically label the remaining m-n vertices. The aim of transductive classification learning on hypergraph is to estimate the labels on the remaining m-n vertices.

With fixed $\boldsymbol{y}=\{y_1,\ldots,y_m\}$. The goal is to reconstruct it from subset of labels. In statistic learning theory setting, label y_j is regarded as vector in $\mathbb{R}^K$, *i.e.*, $y_j=k$ means y_j is corresponding to vector $\boldsymbol{f_j}=\{f_{j,1},\ldots, f_{j,K}\}$, where k-th entry is 1, all others are 0. We can decode the corresponding label estimation $\hat{y}_j$ as:

$$\hat{y}_j=\hat{y}(\boldsymbol{f}_j)=\arg\max_k\{f_{j,k}: k=1,\cdots,K\}.$$

Let y_j be a true lable,

$$I(\cdot)=\begin{cases}1, & if\cdot is\quad true\\ 0, & if\cdot is\quad false\end{cases},$$

Then, the error for classification is:

$$err(\boldsymbol{f}_j, y_j)=I(\hat{y}(\boldsymbol{f}_j)\neq y_j).$$

For estimating the concatenated vector $\boldsymbol{f}=[\boldsymbol{f_j}]=[f_{j,k}]\in\mathbb{R}^{mK}$ *via* a subset of labeled vertices, we need to impose restrictions using a quadratic regularizer:

$$\boldsymbol{f}^T\boldsymbol{Q_K}\boldsymbol{f}=\sum_{k=1}^{K}\boldsymbol{f}_{\cdot,k}^T\boldsymbol{K}^{-1}\boldsymbol{f}_{\cdot,k},$$

where, $\boldsymbol{K}\in\mathbb{R}^{m\times m}$ is full rank positive definite kernel matrix, $\boldsymbol{f}_{\cdot,k}=\{f_{1,k},\ldots, f_{m,k}\}\in\mathbb{R}^m$. *i.e.*, the predictive vector for each class k is regularized separately. Note that we use $\boldsymbol{K}$ to denote the kernel matrix and K to denote the number of classes.

For a fixed vector $\boldsymbol{f}\in\mathbb{R}^{mK}$, we use loss function $\phi(\boldsymbol{f_j}, y_j)$ to measure the quality of its component $\boldsymbol{f_j}=\{f_{j,1},\ldots, f_{j,K}\}\in\mathbb{R}^K$. Thus, the empirical risk on Z_n, subject to $\boldsymbol{f}^T\boldsymbol{Q}_K\boldsymbol{f}$ is given by:

$$\hat{f}(Z_n)=\arg\min_{f\in\mathbb{R}^{mK}}\left[\frac{1}{n}\sum_{j\in Z_n}\phi(\boldsymbol{f}_j, y_j)+\lambda\boldsymbol{f}^T\boldsymbol{Q}_K\boldsymbol{f}\right], \tag{1}$$

where, $\lambda>0$ is a regularization parameter. We are interested in the following special class of loss function in this paper:

$$\phi(\boldsymbol{f_j}, y_j)=\sum_{k=1}^{K}\phi_0(f_{j,k},\delta_{k,y_j}),$$

where,

$$\delta_{a,b}=\begin{cases}1, & a=b\\ 0, & a\neq b\end{cases}.$$

Let $\nabla_1\phi_0(x,y)$ be a sub-gradient of $\phi_0(x,y)$ with respect to x. We need the following assumption:

Assumption 1： There exist positive constants a, b, and c such that

(1) $\phi_0(x,y)$ is non-negative and convex in x.

(2) When y=0,1, and $\phi_0(x,y)\le a, \left|\nabla_1\phi_0(x,y)\right|\le b$.

(3) $c=\inf\{x: \phi_0(x,1)\le a\}-\sup\{x: \phi_0(x,0)\le a\}$.

We first give classification error on hypergraph as follows:

Theorem 1. Consider (1) with loss function satisfying Assumption 1. Then for $\forall p>0$, the expected generalization error of the learning algorithm (1) on training samples Z_n (uniformly drawn without replacement form hypergraph vertices $\{1,\ldots,m\}$) can be bounded by:

$$E_{Z_n}\frac{1}{m-n}\sum_{j\in\bar{Z}_n} err(\hat{\boldsymbol{f}}_j(Z_n), y_j)$$

$$\le\frac{1}{a}\inf_{f\in\mathbb{R}^{mK}}\left[\frac{1}{m}\sum_{j=1}^{m}\phi(\boldsymbol{f}_j, y_j)+\lambda\boldsymbol{f}^T\boldsymbol{Q}_K\boldsymbol{f}\right]+\left(\frac{b\,\mathbf{tr}_p(\boldsymbol{K})}{\lambda nc}\right)^p,$$

where $\bar{Z}_n=\{1,\ldots,m\}-Z_m$,

$$\mathbf{tr}_p(\boldsymbol{K})=\left(\frac{1}{m}\sum_{j=1}^{m}\boldsymbol{K}_{j,j}^p\right)^{1/p},$$

and $\boldsymbol{K}_{j,j}$ denoted as j-th diagonal entry of matrix $\boldsymbol{K}$.

2. GENERALIZATION BOUND WITH HYPERGRAPH -CUT

For fixed label $y=\{y_j\}_{j=1,\ldots,m}$ on V, the hypergraph cut for the $\boldsymbol{S}$-normalized Laplacian $L_{\boldsymbol{S}}(H)$ as:

$$cut(L_S, y) = \frac{1}{2}\sum_{e\in E}\frac{1}{\delta(e)}\sum_{\{j,j'\}\subseteq e} w(e)(\frac{1}{S_j}+\frac{1}{S_{j'}})$$
$$+\ \frac{1}{2}\sum_{e\in E}\frac{1}{\delta(e)}\sum_{\{j,j'\}\subseteq e} w(e)(\frac{1}{\sqrt{S_j}}-\frac{1}{\sqrt{S_{j'}}})^2 .$$

In this section, we apply Theorem 1 to analyze hypergraph learning by using hypergraph-cut.

The learning theoretical definition of hypergraph-cut penalizes a normalized version of between-class hyperedge weights on one hand, and penalizes within-class hyperedge weights when such a hyperedge connects two vertices with different scaling factors on the other hand. For un-normalized Laplacian, we delete the second term on the right hand side of hypergraph-cut definition, *i.e.*, it only penalizes weights corresponding to edges connecting the vertices with different labels. Under such situation, the learning theoretical definition corresponds to the hypergraph-theoretical definition:

$$cut(L_S, y) = \sum_{e\in E}\frac{1}{\delta(e)}\sum_{\{j,j'\}\subseteq e} w(e) .$$

According to the learning theoretical hypergraph-cut definition, the generalization result for the estimator in (1) with ***K*** defined is given:

$$K^{-1} = \alpha S^{-1} + L_S(H) = \boldsymbol{S}^{-1/2}(\alpha\boldsymbol{I} + L(H))\boldsymbol{S}^{-1/2} \qquad (2)$$

where, $\alpha > 0$ is called a tuning parameter in order to ensure that ***K*** is strictly positive definite. The corresponding regularization condition can be regarded as:

$$\boldsymbol{f}^T\boldsymbol{Q}_K\boldsymbol{f} = \sum_{k=1}^{K}(\alpha\sum_{j=1}^{m}\frac{f_{k,j}^2}{\boldsymbol{S}_j} + \frac{1}{2}\sum_{e\in E}\frac{1}{\delta(e)}\sum_{\{j,j'\}\subseteq e} w(e)(\frac{f_{j,k}}{\sqrt{S_u}}-\frac{f_{j',k}}{\sqrt{S_v}})^2).$$

Usually, other trick is by setting $\boldsymbol{K}^{-1} = \alpha\boldsymbol{I} + L_S(G)$. The corresponding conclusions are similar to that of (2).

Note that the bound of Theorem 1 depends on λ and ***K***, and this inspires us to consider the more detailed bound with optimal λ. The assumption is stated as follows which is used for our analysis.

Assumption 2 Consider (1) with regularization condition (2). ϕ is loss function and satisfies Assumption 1 and $\phi_0(0,0) = \phi_0(1,1) = 0$.

The Assumption 2 on the loss function here holds for the least squares method and other standard loss functions such as SVM.

Theorem 2 Consider (1) with Assumption 2 is satisfied. Then for any $p > 0$, there exists a sample independent regularization parameter λ such that the expected generalization error can be bounded by:

$$E_{Z_n}\frac{1}{m-n}\sum_{j\in\bar{Z}_n} err(\hat{\boldsymbol{f}}_j(Z_n), y_j) \le \frac{C_p(a,b,c)}{n^{p/(p+1)}}(\alpha s + cut(L_S, y))^{p/(p+1)} tr_p(\boldsymbol{K})^{p/(p+1)},\ C_p(a,b,c) = (\frac{b}{ac})^{\frac{p}{p+1}}(p^{\frac{1}{p+1}} + p^{-\frac{p}{p+1}}) \qquad (3)$$

where, $s = \sum_{j=1}^{m} S_j^{-1}$.

In the following context, we will give some applications of examples for Theorem 2.

3. ALGORITHM ALAYSIS USING ZERO-CUT IN HYPERGRAPH

In this section, we consider an application of Theorem 2 for the normalized Laplacian with the zero-cut assumption that each connected component of the hypergraph has a single label. Under this assumption, our goal is to estimate the label for each connected component.

Theorem 3 Consider (1) such with Assumption 2 and the regularization condition is $\boldsymbol{K}^{-1} = \alpha I + L$. Assume that cut($L$, y) = 0, and the hypergraph has q connected components with sizes $m_1 \le \ldots \le m_q$ ($\sum_l m_l = m$). For any $p>0$, let $\alpha \to 0$, and with optimal λ, we obtain the generalization bound

$$E_{Z_n}\frac{1}{m-n}\sum_{j\in\bar{Z}_n} err(\hat{\boldsymbol{f}}_j, y_j)$$
$$\le \frac{C_p(a,b,c)}{n^{p/(p+1)}}(\sum_{l=1}^{q}(\frac{m}{m_l})^{p-1})^{1/(p+1)} + \mathrm{O}(\alpha),$$

where C_p is defined by (3). More specific, we get

$$E_{Z_n}\frac{1}{m-n}\sum_{j\in\bar{Z}_n} err(\hat{\boldsymbol{f}}_j, y_j) \le \min[2\sqrt{\frac{b}{ac}\cdot\frac{q}{n}}, \frac{b}{ac}\cdot\frac{m}{nm_1}] + \mathrm{O}(\alpha).$$

Under the zero-cut assumption, when $\alpha \to 0$, the generalization error can be bounded as $O(\sqrt{\frac{q}{n}})$. However, a faster convergence rate of $O(\frac{1}{n})$ can also be achieved, although the bound for generalization error depends on the inverse of the smallest component size through $\frac{m}{m_1} \ge q$. *i.e.*, we can reach a better convergence at the $O(\frac{1}{n})$ level under the condition that the sizes of the components are balanced. If the component sizes are significantly different, the convergence may behave like $O(\sqrt{\frac{q}{n}})$.

4. NON-ZERO CUT AND PURE COMPONENTS

We find the assumption that each connected component has only one label (*i.e.*, the cut is zero) and it is too restrictive, and in many applications, this assumption is not reasonable. In this section, we relax the assumption and obtain similar bounds.

Theorem 4. Consider (1) with Assumption 2 is satisfied. Let $H' = \bigcup_{l=1}^{q} H_l$ be a pure sub-hypergraph of H. For any

$p \geq 1$, there exists a fixed tuning parameter α and a sample-independent regularization parameter λ, such that

$$E_{Z_n} \frac{1}{m-n} \sum_{j \in \bar{Z}_n} err(\hat{\boldsymbol{f}}_j, y_j)$$

$$\leq \frac{C_p(a,b,c)}{n^{p/(p+1)}} (s^{1/2} (\sum_{l=1}^{q} \frac{s_l(p)/m}{m_l^p})^{1/2p}$$

$$+ \, cut(L_S, y)^{1/2} (\sum_{l=1}^{q} \frac{s_l(p)/m}{\lambda_2(G_l)^p})^{1/2p})^{2p/(p+1)},$$

where, C_p is defined in (3), $m_l=|V_l|$, $s = \sum_{j=1}^{m} \boldsymbol{S}_j^{-1}$, and $s_l(p) = \sum_{j \in V_l} S_j^p$.

Theorem 4 can be regarded as a natural generalization of Theorem 3 for $p \geq 1$. As common graph, it quantitatively illustrates the importance of analyzing hypergraph learning by using a partition of the original hypergraph into well-connected pure components. The second eigenvalue λ_2 (H_i) measures how well-connected H_i is. A more intuitive quantity that measures the connectedness of hypergraph $H=(V,E)$ is the isoperimetric number h_H defined as

$$h_H = \inf_{S \subset V} \sum_{j \in S, j' \in V-S, \{i,j\} \subseteq e} \frac{w(e)}{\delta(e) \min(|S|, |V-S|)}.$$

Similar as standard spectral graph theory [19], we can cheek that λ_2 (H_i) $\geq$ $\frac{h_{H_i}^2}{2 \max_j \deg_j(H_j)}$.

If the vertices are well-connected everywhere, then the isoperimetric number of hypergraph H is large. Specifically, if $\deg_j(H)$ is of the order $|V|$, and $w(e)= 1$ when $e \in E$, then $\sum_{j \in S, j' \in V-S, \{i,j\} \subseteq e} \frac{w(e)}{\delta(e)}$ is of the order $|S||V-S|$, and $h_H = O(|V|)$ for a well-connected hypergraph. Let H' be a pure-sub-hypergraph of H with well-behaved, *i.e.*, each pure component H_l of H' is well-connected in the above sense. We infer that $\frac{\lambda_2(H_l)}{m_l} \geq u(H')$,for some constant $u(H')$ which is independent on the vertex number of the pure components (but only how well-connected each pure component is). By such condition, we replace $\sum_{l=1}^{q} m_l \lambda_2(G_l)^{-p}$ by $u(G')^{-p} \sum_{l=1}^{q} m_l^{1-p}$ in Theorem 4 and get a simplified bound:

$$E_{Z_n} \frac{1}{m-n} \sum_{j \in \bar{Z}_n} err(\hat{\boldsymbol{f}}_j, y_j)$$

$$\leq \frac{C_p(a,b,c)}{n^{p/(p+1)}} (\sum_{l=1}^{q} \frac{s_l(p)/m}{(m_l/m)^p})^{1/(p+1)} (\sqrt{\frac{s}{m}}$$

$$+ \sqrt{\frac{cut(L_S, y)}{u(H')m}})^{2p/(p+1)}$$

where

$$u(H') = \min_l \frac{\lambda_2(G_l)}{m_l}.$$

We consider the following two special situations: $p = 1$ and $p \longrightarrow \infty$:

$$E_{Z_n} \frac{1}{m-n} \sum_{j \in \bar{Z}_n} err(\hat{\boldsymbol{f}}_j, y_j) \leq 2 \sqrt{\frac{b}{ac} \cdot \frac{\sum_{l=1}^{q} \frac{s_l(1)}{m_l}}{n}} (\sqrt{\frac{s}{m}} + \sqrt{\frac{cut(L_S, y)}{u(H')m}}), \tag{4}$$

$$E_{Z_n} \frac{1}{m-n} \sum_{j \in \bar{Z}_n} err(\hat{\boldsymbol{f}}_j, y_j) \quad \leq \frac{b}{ac} \cdot \frac{\max_l \max_{j \in V_l} (S_j / m_l)}{n} (\sqrt{s}, + \sqrt{\frac{cut(L_S, y)}{u(H')m}})^2. \tag{5}$$

These bounds can be regarded as the generalizations of those in Theorem 4. If we take $\boldsymbol{S}=\boldsymbol{I}$, then the number of pure components q affects the $O(\frac{1}{\sqrt{n}})$ convergence rate in (4) as $\sum_{l=1}^{q} \frac{s_l(1)}{m_l} = q$. If the vertices number of the components are balanced, better convergence at the $O(\frac{1}{n})$ level as in (5) can be achieved; otherwise, the convergence may just be as $O(\sqrt{\frac{q}{n}})$ level. This fact inspires a scaling matrix $\boldsymbol{S}$ that compensates for the unbalanced vertices number of pure component, which we will discuss in the following context.

5. NEAR-ZERO-CUT PARTITION WITH OPTIMAL NORMALIZATION

We discuss a pure sub-hypergraph $H'=\bigcup_{l=1}^{q} H_l$ of H. If the scaling factors $\boldsymbol{S}_j$ are nearly constant within each pure component, then using the Laplacian definition above, we infer a small regularization penalty for the hyperedges within a pure component and between the vertices with similarly output values (*i.e.*, $f_{j,k} \approx f_{j',k}$). Thus, in the next context, we focus on finding the optimal scaling matrix $\boldsymbol{S}$ such that $\boldsymbol{S}_j$ is constant within each pure component V_l. Assume that we use q numbers $[\overline{s_l}]_{l=1,\ldots,q}$, to quantify $\boldsymbol{S}$ which satisfys $\boldsymbol{S}_j = \overline{s_l}$ for $j \in V_{l'}$.

Consider the following quantity:

$$cut(H',y)=\sum_{\{j,j'\}\subseteq e, y_j\neq y_{j'}} w(e)\frac{1}{\delta(e)}$$

$$+\sum_{l\neq l'}\sum_{j\in V_l, j'\in V_{l'}}\sum_{\{j,j'\}\subseteq e}\frac{w(e)}{2\delta(e)}$$

Then, we have

$$cut(L_S,y)\leq\frac{cut(H',y)}{\min_l \overline{s}_l}.$$

We assume that hyperedge weights are small if it contains vertices between pure components. Then, we can ensure that $cut(H',y)$ is small as well.

With the $O(\frac{1}{n})$ convergence rate, we obtain from (5) that

$$E_{Z_n}\frac{1}{m-n}\sum_{j\in\bar{Z}_n} err(\hat{\boldsymbol{f}}_j,y_j)$$

$$\leq\frac{b}{ac}\cdot\frac{\max_l(s_j/m_l)}{n}\left(\sqrt{\sum_{l=1}^{q}\frac{m_l}{\overline{s}_l}}+\sqrt{\frac{cut(H',y)}{u(H')\min_l\overline{s}_l}}\right)^2.$$

If $cut(H',y)$ is small, then the right hand side of dominating term becomes $\frac{\max_l(s_j/m_l)}{n}\sum_{l=1}^{q}\frac{m_l}{\overline{s}_l}$, which can be optimized by choosing $\overline{s}_l=m_l$, and the resulting bound will be as follows:

$$E_{Z_n}\frac{1}{m-n}\sum_{j\in\bar{Z}_n} err(\hat{\boldsymbol{f}}_j,y_j)\ \leq\frac{b}{ac}\cdot\frac{1}{n}\left(\sqrt{q}+\sqrt{\frac{cut(H',y)}{u(H')\min_l\overline{s}_l}}\right)^2.$$

i.e., under the condition $cut(H',y)$ is small, we can choose scaling factor $\overline{s}_l\propto m_l$ for each pure component such that the generalization bound is of the order $O(\frac{1}{n})$, which approximate to $\frac{bq}{acn}$.

6. CONCLUSION

In our paper, we consider the semi-supervised learning framework, and obtain theoretic conclusions for spectral based hypergraph vertex classification of semi-supervised learning algorithm. The contribution of this article is two-fold: first, we deduce the generalization bound for such algorithm in terms of zero-cut, non-zero-cut and pure component; second, the generalization performance bound for near-zero-cut partition with optimal parameter λ is yielded. The result achieved in our paper illustrates the promising application prospects for algorithms using hypergraph model.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

First, we thank the reviewers for their constructive comments in improving the quality of this paper. We also would like to thank the anonymous referees for providing us with constructive comments and suggestions. We wish to acknowledge the National Natural Science Foundation of China (61142007) and the College Natural Science Foundation of Jiangsu Province in China(10KJD520002).

REFERENCES

[1] R. Johnson, T. Zhang, "On the effectiveness of laplacian normalization for graph semi-supervised learning", *Mach. Learn. Res.*, vol. 8, pp. 1489-1517, 2007.

[2] S. Agarwal, "*Ranking on graph data*", in Proceedings of the 23rd international conference on machine learning, 2006, pp. 25-32.

[3] S. Agarwal, "Learning to rank on graphs", *Mach. Learn.,* vol. 81, no. 3, pp. 333-357, 2010.

[4] S. Smale and D. X. Zhou, "Geometry on probability spaces", *Construct. Approx..* vol. 30, pp. 311-323, 2009.

[5] S. Agarwal, K. Branson, and S. Belongie, "*Higher order learning with graphs*", In: Proceedings of the 23rd International conference on Machine Learning, Pittsburgh, PA: ICML, 2006, pp. 17- 24.

[6] D. Zhou, J. Huang, and B. Scholkopf, "*Learning from labeled and unlabeled data on a directed graph*", In: Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany:" ICML, 2005, pp. 1041-1048.

[7] I. Pesenson, "Variational splines and paley-wiener spaces on combinatorial graphs", *Constr. Approx.*, vol. 29, pp. 1- 21, 2009.

[8] D. Zhou, J. Huang, and B. Scholkopf, "*Beyond pair-wise classification and clustering using hypergraphs*", Canada: University of Waterloo, 2005.

[9] D. Zhou, J. Huang, and B. Scholkopf, "*Learning with hypergraphs, clustering, Classification, and Embedding*", In: Proceedings of 20th Annual Conference on Neural Information Processing Systems, 2006, Vancouver / Whistler, Canada: IEEE, 2006, pp. 1601-1608.

[10] S. Liang, S. Ji, and J. Ye, "*Hypergraph spectral learning for multi-label classification*", In: Proceeding of the 14th ACM SIGKDD international conference on Know ledge discovery and data mining. Las Vegas, Nevada, USA: ACM, 2008, pp. 668- 676.

[11] G. Chen, J. Zhang, and F. Wang, C. Zhang, Y. Gao, "*Efficient multi- label classification with hypergraph regularization*", In: IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL, USA: IEEE, 2009, vol. 9, pp. 1658-1665.

[12] W. Gao, L. Liang, and Y. Zhang, "Ontology concept similarity computation Based on Graph Learning", *J. Southwest China Normal Univ. (Natural Science Edition)*, vol. 36, no. 4, pp. 64-67, 2011.

[13] W. Gao, L. Zhu, and L. Liang, "Ontology mapping algorithm based on the regularization framework of graph", *J. Southwest Univ. (Natural Science Edition)*, vol. 34, no. 3, pp. 118-121, 2012.

[14] W. Gao, Y. Gao, and L. Liang, "Ontology mapping algorithm based on δ-neighborhood", *J. Yunnan Normal Univ. (Natural Science Edition)*, vol. 31, no. 3, pp. 37-40, 2011.

[15] T. Xu, J. Gan, L. Gao, and W. Gao, "Ontology mapping algorithm based on dimensionality reduction theory", *J. Northwest Normal Univ. (Natural Science Edition)*, vol. 47, no. 4, pp. 52-55, 2011.

[16] W. Gao and L. Liang, "Ontology concept similarity computation based on regularization framework of hypergraph", *Microelect. Comp.*, vol. 28, no. 5, pp. 15-17, 2011.

[17] W. Gao, L. Zhu, and L. Liang, "Ontology mapping algorithm based on the regularization framework of graph", *J. Southwest Univ. (Natural Science Edition)*, vol. 31, no. 3, pp. 118-121, 2012.

[18] W. Gao, L. Zhu, and Y. Zhang, "Ontology concept similarity Computation based on graph learning", *J. Southwest China Normal Univ. (Natural Science Edition)*, vol. 34, no. 4, pp. 64-67, 2011.

[19] F. Chung, "*Spectral graph theory*", Regional Conference Series in Mathematics. American Mathematical Society, Rhode Island, 1998.

Received: September 22, 2014 Revised: November 30, 2014 Accepted: December 02, 2014