修士論文 要旨

グラフ埋め込みにおける正則化の研究

数理情報学専攻 48186206　丸尾 恭四郎

指導教員　山西 健司 教授

1　概要

グラフ構造における解析は古くから行われているが，近年ではソーシャルネットワークやウェブサイトなどが現れたことにより，大規模なグラフに対する解析はますます盛んとなっており，グラフにおける機械学習の手法の需要は高まっている．特に近年ではグラフのノードを，低次元ベクトルへと表現するグラフ埋め込みという手法が注目されている．

一方で，機械学習一般において，パラメーターの次元は小さすぎては表現能力が低く，次元が大きすぎると過学習という問題が起こることがよく知られているが，グラフ埋め込みの文脈においてはこれを最適化する方法については議論されていない．

本研究では，グラフ埋め込みに Group Lasso と呼ばれる正則化を導入することで適切に低次元なベクトルを学習し，また最適化における計算量を削減する手法を提案する．さらに記述長最小原理に基づく MDL-RS と呼ばれる手法で，正則化パラメーターを各次元に対して非一様にし，正則化パラメーターに関しても各次元に対してそれぞれ最適化することで低次元でかつ高性能な埋め込みベクトルを獲得する手法を提案する．本研究の新規性は，グラフ埋め込みに Group Lasso を用いた正則化を行い，さらにその正則化パラメーターも最適化する手法を与えた点にある．さらに数値実験によって，他の正則化や，正則化なしの場合のグラフ埋め込みと比較をし，再現率の意味での性能を高く保ちつつ，より次元が低い埋め込みベクトルが得られたことを示した．

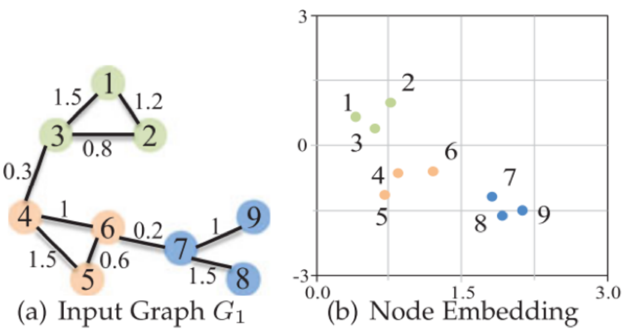

図 1. グラフ埋め込みの例．(a) 入力のグラフ．(b) 埋め込みの結果．[1] より引用．

2　グラフ埋め込みについて

グラフ埋め込みはいくつかの手法が提案されているが，本研究では Skip-Gram[2] に基づく手法に注目する．これは，2 つのノードが近ければ，埋め込み先における，2 つのベクトルの内積が大きくなるように学習する手法である．次のような最適化問題を解くことで学習する：

$$\min_{U\in\mathbb{R}^{d\times n}, V\in\mathbb{R}^{d\times n}} \sum_{i,j} f_{ij}(u_i, v_j),$$

$$f_{ij}(u_i, v_j) = \begin{cases} \sigma(u_i \cdot v_j) & (\text{ノード } i, j \text{ 間にエッジがある}) \\ \sigma(-u_i \cdot v_j) & \text{otherwise}, \end{cases}$$

$$U = (u_1, \ldots, u_n), V = (v_1, \ldots, v_n),$$

$$\sigma(x) = \frac{1}{1+e^{-x}}.$$

(n はグラフのノード数)

U, V は埋め込みベクトルを並べたもので，U, V がグラフ埋め込みにおいて求めるべきベクトルを並べたものである．

3　Group Lasso

正則化学習では，目的関数に正則化関数を加えて学習することで，ノルムを制限したり，スパースな解を得たりすることができる．次のような最適化問題を考える：

$$\min_{U\in\mathbb{R}^{d\times n}, V\in\mathbb{R}^{d\times n}} \left\{ \sum_{i,j} f_{ij}(u_i, v_j) + g(U, V, \lambda) \right\}.$$

ここで現れる g が正則化関数と呼ばれる．

Group Lasso [3] とは L1 正則化を拡張した正則化であり，L1 正則化はパラメーターに 0 になる成分が現れるという特徴がある一方で，Group Lasso はパラメーターをいくつかのグループに分割し，そのグループがまとめて 0 になるという特徴がある．

埋め込みベクトルを並べた行列 U に対して，行方向にまとめたベクトルを考える．すなわち，U の第 p 行に対するベクトルを，

$$U_{:p} = (u_{1p} \ \cdots \ u_{np})^\mathsf{T} \in \mathbb{R}^n$$

とする．同様に，V の第 p 行に対するベクトルを，

$$V_{:p} = (v_{1p} \ \cdots \ v_{np})^\mathsf{T} \in \mathbb{R}^n$$

とする．

本研究ではグラフ埋め込みに Group Lasso を適用する際に，行列の列を一つのグループとしてまとめる．これにより列がまとめて 0 になり，埋め込みベクトルの非ゼロの次元数を減らすことができる．これによってより低次元なベクトルを得ることを期待できる．正則化関数は，正則化パラメーター $\lambda = (\lambda_{U_1} \cdots \lambda_{U_d}\ \lambda_{V_1} \cdots \lambda_{V_d})^{\mathsf{T}} \in \mathbb{R}^{2d}$ を用いて

$$g(U, V, \lambda) = \sum_{p=1}^{d} \lambda_{U_p} \|U_{:p}\| + \sum_{p=1}^{d} \lambda_{V_p} \|V_{:p}\| \qquad (1)$$

と表現できる．

4 MDL-RS

記述長最小原理 [4] に基づいて正則化パラメーターを最適化することを考える．記述長最小原理ではしばしば正規化最尤符号長 (**NML code length**) とよばれる量が考察される：

$$\mathrm{NML}(x) = \inf_{\theta\in\Theta} f_x(\theta) + \log Z,$$
$$Z = \sum_{y\in X} \sup_{\theta\in\Theta} e^{-f_y(\theta)}.$$

ここで，$x \in X$ は，データとそれの取り得る集合を表し，$\theta \in \Theta$ は，パラメーターとそれの取り得る集合を表す．$f_x(\theta)$ は損失関数とし，$\sum_{x\in X} e^{-f_x(\theta)} = 1$ を満たすとする．

NML code length は ミニマックスリグレットを達成するという性質を持つ：

$$\sup_{x\in X} \mathrm{NML}(x) = \inf_{L} \sup_{x\in X} \left\{ L(x) - \inf_{\theta\in\Theta} \log \frac{1}{e^{-f_x(\theta)}} \right\}.$$

NML code length の拡張として，事前分布を考慮した場合の Luckiness NML (**LNML**) code length がある：

$$\mathrm{LNML}(x, \lambda) = \inf_{\theta\in\Theta} \{f_x(\theta) + g(\theta, \lambda)\} + \log Z(\lambda),$$
$$Z(\lambda) = \sum_{y\in X} \sup_{\theta\in\Theta} \exp\{-f_y(\theta) - g(\theta, \lambda)\}.$$

$Z(\lambda)$ を直接計算することは困難であるが，MDL-RS[5] は，その解析的上界 $\bar{Z}(\lambda)(\geq Z(\lambda))$ を導出し，代わりに $\bar{Z}(\lambda)$ を最小にする λ を求めることで，最適な正則化パラメーターを導出するという手法である．

本研究においてこの手法を適用するにはいくつかの導出が必要であるが，紙面の都合で省略する．

5 実験結果

本研究で行なった実験結果の一部を紹介する．CORA[6] と呼ばれる論文の引用ネットワークに対してグラフ埋め込みを適用した．正則化パラメーターを単純にグリッドサーチする場合と比べて，性能を高く保ちつつ，より低次元なベクトルが得られることを示した．

表 1. 再現率と実効的次元の評価．データセットは CORA.

手法	平均再現率 (%)	実効次元
Proposed	**82.3 ± 0.5**	**15.0 ± 1.0**
Group Lasso	80.0 ± 1.0	23.5 ± 1.5
L2	81.2 ± 0.5	24.1 ± 1.2
Simple	78.2 ± 1.0	24.0 ± 2.0

6 まとめ

本研究では，グラフ埋め込みに対して Group Lasso の正則化を適用し，さらにその正則化パラメーターを MDL-RS に従って最適化する方法を提案した．性能を高く保ちつつ，より低い次元のベクトルが得られることを数値実験で確認した．

参考文献

[1] Hongyun Cai, Vincent W Zheng, and Kevin Chen-Chuan Chang. A comprehensive survey of graph embedding: Problems, techniques, and applications. *IEEE Transactions on Knowledge and Data Engineering*, 30(9):1616–1637, 2018.

[2] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.

[3] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. A note on the group lasso and a sparse group lasso. *arXiv preprint arXiv:1001.0736*, 2010.

[4] Andrew Barron, Jorma Rissanen, and Bin Yu. The minimum description length principle in coding and modeling. *IEEE Transactions on Information Theory*, 44(6):2743–2760, 1998.

[5] Kohei Miyaguchi and Kenji Yamanishi. High-dimensional penalty selection via minimum description length principle. *Machine Learning*, 107(8-10):1283–1302, 2018.

[6] Hinrich Schütze, Christopher D Manning, and Prabhakar Raghavan. Introduction to information retrieval. In *Proceedings of the international communication of association for computing machinery conference*, volume 4, 2008.