

逆伝播を持つ有向グラフ上でのラベル伝播を用いた半教師付き学習

泉 谷 暁 彦[†] 上 原 邦 昭[†]

グラフを用いた半教師付き学習の殆どは、無向グラフをデータ構造として扱ったものであり、有向グラフを扱ったものは少ない。それは、有向グラフでは、到達性の検証が無向グラフと比べて困難なためである。本研究では、有向グラフに逆伝播を持つ辺を導入し、グラフ上でのラベル伝播を利用した分類手法を提案する。この手法は、グラフが全連結でさえあれば到達性が保障される上、グラフの構築に大きな計算量を必要としないため、データの更新にも強いアルゴリズムである。また、更なる精度向上のため、半教師あり学習で用いられる Co-Training の枠組を、ラベル伝播を用いた本手法に拡張する。公開データセットによる評価実験では、本手法は既存の手法と比較して高い性能を実証でき、Co-Training の利用により精度向上することが確認できた。

Label Propagation on Directed Graph with Backoff Process

AKIHIKO IZUTANI[†] and KUNIAKI UEHARA[†]

Existing Graph-based Semi-Supervised methods mostly use non-directed graph, because directed graphs can produce unreachable nodes. In this paper, we introduce a digraph with propagation of backward direction (backoff process). We also propose the label propagation algorithm on digraph. Our algorithm supports reachability for connected graph and less computational cost for graph construction. Moreover, we extend our method to Co-Training framework for accuracy improvement. In some experiments, we obtained encouraging results.

1. はじめに

通常の機械学習では、数多くのクラスラベル付き訓練事例（ラベルデータ）が必要になる。しかしながら、ラベルデータを得るためには人手によるラベル付けが必要になり、非常に多くのコストがかかる。それに対して、クラスラベルなしの事例（未ラベルデータ）は、インターネットからの自動収集などによって、比較的容易に得られる場合が多い。そこで、近年、ラベルデータと未ラベルデータの両方を利用する半教師付き学習が盛んに研究されている。¹⁾

特に近年、グラフ構造を用いた手法は、半教師付き学習において高い性能を上げている。しかし、データ構造をグラフとして表現するため、構築したグラフに分類精度が依存する場合が多い。そのため、グラフの構築手法は大きな問題であるが、多くの手法では、十分に議論されないまま無向グラフが用いられている。

一方、Wang ら²⁾によって、隣接頂点間の類似度を考慮した二次計画法により最適な有向グラフを構築し、そのグラフ上でのラベル伝播による分類手法が提

案された。この手法は、隣接頂点間でのラベルの伝播という比較的単純な手法でありながら、既存の無向グラフの手法と比べて高い性能を示している。しかし、この手法には問題点がある。第一に到達不可能点（ラベル伝播が不可能な頂点）を生成する可能性があるという点がある。第二にグラフの構築に大きな計算量が必要とされるため、データの追加や変更が困難であるという点である。

本研究では、これらの問題を解決するために、逆伝播を持つ有向グラフを利用した分類アルゴリズムを提案する。提案手法では、有向辺に対して逆方向の伝播を与えて、到達不可能点へのラベル伝播を可能にしている。また、グラフ構築に際して、近傍構造を利用したグラフ構造の最適化計算はせず、訓練事例の追加が起こった場合などに、グラフ構築のコストを減少させることができる。

さらに、メタ学習手法である Co-Training³⁾ の導入について議論する。ラベル伝播アルゴリズムには、Co-Training のような反復学習手法が効果的であるといわれる⁴⁾ が、グラフ上でのラベル伝播手法に有用かどうかは明らかにされていない。そこで、本研究では、Co-Training の枠組を提案手法に適用する方法を提案し、分類精度の変化について検討する。

[†] 神戸大学大学院工学研究科
Graduate School of Engineering, Kobe University

2. 逆伝搬を持つ有向グラフの構築

グラフを用いた手法の多くは、データセットを無向グラフとして表現している。しかし、距離が対称な無向グラフよりも、非対称な有向グラフの方がより柔軟な表現ができる。その反面、到達性を持つグラフの構築は有向グラフの場合には単純ではない。ラベル伝播アルゴリズムなどでは、到達不能点が分類不可能となってしまうため、到達性は重要な条件である。

ここで、Fagin ら⁵⁾によって提案された、逆伝播を持つ有向グラフに注目する。逆伝播は、有向辺に対して逆方向への確率的な伝播を同時に定義するものである。このグラフは連結グラフでさえあれば、到達性が保障されるという利点がある。

本研究では、その概念をグラフを用いた半教師付き学習手法に拡張し、各有向辺に逆方向への確率的伝播が存在すると定義する。

まず、データセット $\mathbf{D}$ の i 番目 ($1 \leq i \leq n$) の事例を $\mathbf{x}_i = [a_1 \cdots a_d]^T$ と定義する。また、ラベルデータの集合を $\mathbf{D}_l$ 、未ラベルデータの集合を $\mathbf{D}_u$ とする。 x_i が表現する頂点を v_i とし、 v_i から v_j への有向辺を e_{ij} と表す。また、辺 e_{ij} に付加された重みを w_{ij} として表す。 w_{ij} の計算にはガウス関数 $w_{ij} = \exp \left\{ -\sum_{k=1}^d \frac{(x_{ik} - x_{jk})^2}{2\sigma^2} \right\}$ を用いる。

次に、有向近傍グラフ (k -NN グラフ) を構築する。有向近傍グラフとは、頂点 v_i から頂点 v_j の類似度が、上位 k 以内である場合のみ、辺を張るグラフである。なお、構築後に $\sum_j w_{ij} = 1$ と正規化する。

さらに、各有向辺について確率 ρ での逆伝播構造を導入することを提案する。本手法では、逆伝播を元の辺の重みと逆伝播を足し合わせた値とする。これによって、 $(\mathbf{W})_{ij} = w_{ij}$ で定義されるグラフの隣接行列 $\mathbf{W}$ は更新され、

$$\hat{\mathbf{W}} = \mathbf{W} + \rho \mathbf{W}^T \quad (1)$$

となる。ただし、元の重みより大きな逆伝播は起こさないとし、 $0 \leq \rho \leq 1$ とする。以上の手法で逆伝播を持つ有向グラフを構築し、このグラフ上でのラベル伝播を利用して分類する。

3. ラベル伝搬アルゴリズム

逆伝播を持つ有向グラフを用いたラベル伝播による分類手法を提案する。提案手法は、Wang ら²⁾の伝播アルゴリズム (LNP) を改良したもので、辺の重みの最適値計算の手順を取り除き、計算量を削減している。提案手法は、逆伝播を持つ有向グラフの特性を継

承し、到達不能点を考慮する必要がないという利点がある。以下で、提案アルゴリズムの説明をする。

ラベルデータ $\mathbf{x}_i$ が表現する頂点 v_i に与えられるクラスの数を C とする。各頂点にクラスラベルを示す C 次元ラベルベクトルを与え、 $\mathbf{f}_i$ と定義する。クラス c に属するラベルデータは $\mathbf{f}_{ik} = 1$ ($k = c$)、 $\mathbf{f}_{ik} = 0$ ($k \neq c$) とし、未ラベルデータは、 $\mathbf{f}_i = [0, 0, \dots, 0]$ とする。分類アルゴリズムは、辺によって連結された頂点間 v_i から v_j に対し、ラベルベクトル $\mathbf{f}_i$ を伝播させる。

本研究では、ラベルの伝播アルゴリズムを、逆伝播を用いたグラフに適用できるように拡張する。まず、隣接した頂点の $\mathbf{f}_i$ に、辺の重み $\hat{w}_{ij} = w_{ij} + \rho w_{ji}$ を掛け合わせたものを伝播の係数とする。そして、伝播されたラベルと元のラベルとの加重平均をとって、未ラベルデータのラベル $\mathbf{f}_i^t$ ($i \in \mathbf{D}_u$) を更新し、 $\mathbf{f}_i^{t+1}$ を求める。全ての頂点のラベルベクトルを $\mathbf{F} = [\mathbf{f}_1, \dots, \mathbf{f}_n]^T$ と行列表示すると、アルゴリズム全体は、以下のよう

$$\mathbf{F}^{t+1} = \alpha(\hat{\mathbf{W}} - \hat{\mathbf{W}}\mathbf{A})\mathbf{F}^t + (1 - \alpha)\mathbf{F}^0 \quad (2)$$

ただし、 $\mathbf{A}$ は $n \times n$ 対角行列で $x_i \in \mathbf{D}_l$ のとき $(\mathbf{A})_{ii} = 1$ 、 $x_j \in \mathbf{D}_u$ のとき $(\mathbf{A})_{jj} = 0$ となる行列とする。また、 α は、このステップで伝播されてきたラベルベクトルと、初期のラベルベクトルの重ね合わせ係数であり、本研究では $\alpha = 0.5$ としている。

提案手法では、以上をラベルが変化しなくなるまで実行し、ラベルデータから未ラベルデータへラベルベクトルの値を伝播させる。最終的に、各頂点に与えられたラベルベクトルの中で最高の値を持つ要素 $\arg_i \max \mathbf{f}_i$ を推定クラスとして出力する。

4. 評価実験

性能評価のため、公開データセットを用いた実験を行う。なお、特に記述しない場合、作成するグラフは 5-NN グラフとし、ガウス関数のパラメータは $\sigma = 1.0$ としている。また、アルゴリズムの収束の閾値を $1e-6$ と設定し、全てのラベルベクトルが閾値以上の変化しなくなったときに収束したとみなしている。実験結果は、各試行における未ラベルデータの平均分類精度を用い、差の検定には 5% の t -検定を用いる。

4.1 既存手法との比較実験

ここでは、提案手法の性能を評価する。UCI データセット^{*}より、breast, bupa, ionosphere, pima, ecoli,

^{*} <http://www.ics.uci.edu/mllearn/MLRepository.html>

表 1 UCI データセットを用いたラベル伝播アルゴリズムの評価。
Table 1 Result for Effectiveness of Proposal Method on UCI Datasets.

Dataset	Labels	Harmonic	SGT	TSVM	SVM	kNN	Proposal
breast	10	74.1 (± 12.2)	83.8 (± 5.7)	94.5 (± 2.0)	92.0 (± 9.6)	80.2 (± 16.8)	90.6 (± 5.3)
	50	76.1 (± 10.6)	86.5 (± 1.9)	94.8 (± 1.5)	96.3 (± 0.5)	95.7 (± 0.7)	94.6 (± 1.2)
bupa	10	59.2 (± 0.4)	56.9 (± 4.4)	57.1 (± 6.4)	53.7 (± 6.0)	55.1 (± 4.8)	57.9 (± 4.1)
	50	64.2 (± 1.2)	61.8 (± 3.8)	62.3 (± 4.0)	55.5 (± 5.4)	62.2 (± 3.8)	63.9 (± 2.6)
ionosphere	10	65.1 (± 0.4)	74.9 (± 5.8)	74.4 (± 7.6)	64.1 (± 11.8)	65.2 (± 6.7)	72.7 (± 8.3)
	50	69.4 (± 0.8)	84.4 (± 2.7)	84.5 (± 1.8)	82.4 (± 3.9)	78.2 (± 6.3)	80.6 (± 2.0)
pima	10	65.6 (± 1.9)	56.8 (± 4.3)	62.3 (± 7.3)	66.2 (± 3.7)	64.1 (± 6.4)	63.2 (± 4.9)
	50	67.3 (± 4.2)	59.2 (± 3.5)	68.3 (± 5.1)	69.2 (± 3.0)	72.1 (± 1.9)	68.7 (± 2.3)
ecoli	10	62.0 (± 9.1)	N/A	N/A	58.6 (± 10.9)	54.8 (± 10.2)	67.5 (± 8.6)
	50	76.2 (± 6.8)	N/A	N/A	72.4 (± 6.5)	82.0 (± 2.1)	81.1 (± 2.2)
glass	10	18.7 (± 4.5)	N/A	N/A	38.5 (± 6.7)	38.4 (± 7.1)	46.4 (± 8.1)
	50	46.8 (± 1.4)	N/A	N/A	47.6 (± 5.1)	68.5 (± 2.7)	70.0 (± 3.8)
iris	10	76.2 (± 11.7)	N/A	N/A	73.5 (± 10.2)	71.1 (± 12.5)	91.1 (± 1.8)
	50	93.5 (± 1.8)	N/A	N/A	94.8 (± 2.4)	96.4 (± 1.5)	95.3 (± 1.0)
wine	10	69.4 (± 14.9)	N/A	N/A	73.2 (± 14.2)	65.6 (± 11.3)	84.4 (± 5.8)
	50	95.6 (± 2.1)	N/A	N/A	97.2 (± 1.3)	96.1 (± 1.5)	93.8 (± 2.1)

glass, iris, wine を用いた。

比較対象は、無向グラフを用いた手法として、Zhu ら⁶⁾による Harmonic 関数を用いる手法、Joachim によるグラフのスペクトルを用いた手法である SGT⁷⁾、また、グラフを用いない半教師あり学習として、SVM を半教師付き学習に拡張した TSVM⁸⁾、通常の教師あり学習として SVM, k-NN である。

試行回数は 20 回とし、harmonic は無向 50-NN グラフ、k-NN は $k=5$ を用いた。また、SGT と TSVM は二値分類器であるため、二値クラス問題についてのみ実験した。ラベル数が 10, 50 の場合の、実験結果を表 1 に示す。また、最良の結果と、それらと有意差が無い結果を太字で示している。

表 1 から、多くのデータセットにおいて、既存の手法と比べて最良か最良のものとはほぼ同等であることが分かった。特に、ラベル数が 10 といった少ない状況で、高い性能を示すことが出来た。これらの結果から、提案手法は、全てのデータセットで極端に性能が下がることなく、既存手法と比較して、様々なデータセットに対応可能なことが分かる。

4.2 LNP との比較実験

LNP の最適化手法はグラフ全体の隣接構造を利用するため、訓練データが変化するたびにグラフを全て更新する必要がある。そのため、データが追加されるような場合には、あらかじめ行った訓練事例上でのラベル伝播の結果を用いて推定ラベルを計算する LNP induction という手法で解決が図られている²⁾。一方、本手法は、逆伝播を用いた有向グラフであるため、追加したデータからの有向辺を追加するだけでグラフ更新が可能である。

よって、未ラベルデータが追加される条件で、LNP induction との性能比較実験をする。なお、2) に基づき、

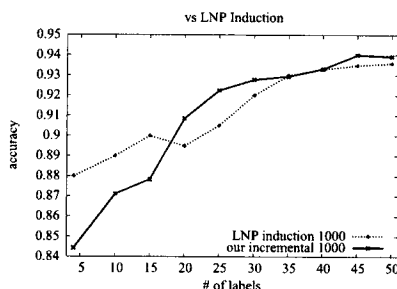


図 1 USPS データセットにおける追加学習時の分類精度の評価
Fig. 1 Incremental Classification Results on USPS.

USPS データセット^{*} から、クラス '1', '2', '3', '4' のみを利用した。実験は、予め与えられた 1000 個のデータで学習し、追加される残りのデータでの分類精度を比較する。ラベルデータは 4 – 50 個まで変化させ、各設定について 20 回試行した。また、グラフの近傍数は 20 としている。なお、LNP の実験結果は、2) で示された値を利用した。実験結果を図 1 に示す。

グラフより、LNP induction と比べてほぼ同等の性能であることが分かる。また、提案手法はグラフ更新により、追加で与えられたデータの構造の学習ができるため、初期に与えられる未ラベルデータが少なくなるとつれて有利になると考えられる。

4.3 逆伝播の有効性

本研究で導入した逆伝播プロセスが到達性の問題に有効であることを示すために、逆伝播を行わない場合との比較実験をする。設定は、4.1 節に従うとした。20 ラベル時の実験結果を表 2 に示す。また、有意差が確認できた場合、分類精度が高い方を太字で強調した。

実験結果より、すべてのデータセットにおいて、逆伝

^{*} <http://kernel-machines.org/data.html>

表 2 逆伝播プロセスの有効性評価.

Table 2 Result for Effectiveness of Backoff Process.

	nobackoff	backoff
breast	83.6 (± 6.8)	92.7 (± 3.6)
bupa	57.4 (± 3.4)	58.8 (± 2.9)
iono	58.1 (± 6.3)	78.5 (± 4.2)
pima	62.7 (± 3.3)	66.5 (± 3.1)
ecoli	70.9 (± 4.8)	76.2 (± 4.5)
glass	52.5 (± 6.2)	55.1 (± 4.9)
iris	87.0 (± 3.8)	92.0 (± 2.9)
wine	82.9 (± 3.7)	90.7 (± 3.1)

播を導入した場合の分類精度が上回り、分散値も低下していることが分かった。逆伝播の導入により到達不可能点の問題が解消されたことによって、未分類点が発生しなくなり、精度向上ができたと考えられる。以上より、逆伝播プロセスの有効性を示すことができた。

5. Co-Training による拡張

半教師付き学習では、2つの独立した学習器を利用したメタ学習手法である Co-Training が広く知られている。ラベルデータが少なく、特徴選択などの手法の利用が難しい半教師付き学習において、分類精度やロバスト性の向上に成果を挙げている⁹⁾。そのため、本研究では、Co-Training を提案手法に適用する方法を提案し、精度の向上が可能か検討する。

本研究では、2種類の異なる k -NN グラフを構築することで、ラベル伝播アルゴリズムへ適用する。

まず、データセット D から、異なる回数からなるグラフ G_g, G_h を構築する。そして、データセットに与えられるラベル L を初期ラベルとし、グラフ G_g 上でラベルの伝播式 (2) を収束させ分類する。次に、 G_g での分類結果 L_g の中から、最も推定ラベルの値の大きい未ラベルデータを、各クラス 1 個ずつラベルデータとして扱う。今度は、グラフ G_h 上で、 L_g を初期ラベルとして同様に操作する。以上を T 回繰り返すことで、Co-Training を実現する。

Co-Training の導入について、UCI データセットを用いた実験で評価する。Co-Training の繰り返し回数は $T = 10$ とし、異なる k -NN グラフを構築するための k 値の組み合わせは 5, 20 としている。なお、各実験は 50 回試行した。20 ラベル時の実験結果を表 3 に示す。それぞれのデータセットについて、差が有意であると検定された実験結果を太字とした。

実験結果より、多くのデータセットにおいて、精度が向上するか、ほぼ同等であることが分かった。よって、グラフを用いたラベル伝播アルゴリズムにおいても、Co-Training が有効であることが示せた。

表 3 Co-Training の性能評価.

Table 3 Result for Effectiveness of Co-Training.

	Normal	Co-Training
breast	91.4 (± 5.2)	94.2 (± 4.0)
bupa	59.2 (± 2.9)	58.8 (± 3.6)
iono	72.5 (± 6.2)	79.5 (± 6.6)
pima	66.7 (± 2.7)	65.6 (± 3.5)
ecoli	72.7 (± 4.0)	74.1 (± 5.2)
glass	53.5 (± 4.7)	49.8 (± 5.3)
iris	91.5 (± 2.8)	94.6 (± 1.9)
wine	89.4 (± 3.5)	92.5 (± 1.9)

6. おわりに

本研究では、データセットを逆伝播を持つ有向グラフとして表現し、そのグラフ上でラベル伝播を行うという学習アルゴリズムを提案した。また、Co-Training をラベル伝播を用いた手法に適用した。

本手法は、はずれ値などが存在すると誤った伝播経路が出来てしまうため、逆伝播を付与する辺を限定することなどが今後の課題として考えられる。また、提案手法では 2 頂点間の関係である辺の接続のみを利用して分類しているが、グラフのフローといったようなグラフ全体の情報を利用することで分類精度が向上するのではないかと考えられる。

参考文献

- 1) Zhu, X.: Semi-Supervised Learning Literature Survey (2005). http://www.cs.wisc.edu/~jerryzhu/pub/ssl_survey.pdf.
- 2) Wang, F. and Zhang, C.: Label Propagation Through Linear Neighborhoods, *Proc. of the 23rd International Conference on Machine Learning*, pp. 985-992 (2006).
- 3) Blum, A. and Mitchell, T.: Combining Labeled and Unlabeled Data with Co-Training, *Proc. of the 11th Annual Conference on Computational Learning Theory*, pp. 92-100 (1998).
- 4) Galstyan, A. and Cohen, P.: Inferring Useful Heuristics from the Dynamics of Iterative Relational Classifiers, *Proc. of the 19th International Joint Conference on Artificial Intelligence*, pp. 708-713 (2005).
- 5) Fagin, R., Karlin, A. R., Kleinberg, J., Raghavan, P., Rajagopalan, S., Rubinfeld, R., Sudan, M. and Tomkins, A.: Random Walks with Back Buttons, *ACM Symposium on Theory of Computing*, pp. 484-293 (2000).
- 6) Zhu, X., Ghahamani, Z. and Lafferty, J.: Semi-Supervised Learning using Gaussian Fields and Harmonic Functions, *Proc. of the 20th International Conference on Machine Learning*, pp. 912-919 (2003).
- 7) Joachims, T.: Transductive Learning via Spectral Graph Partitioning, *Proc. of the 20th International Conference on Machine Learning*, pp. 290-297 (2003).
- 8) Joachims, T.: Transductive Inference for Text Classification using Support Vector Machines, *Proc. of the 16th International Conference on Machine Learning*, pp. 200-209 (1999).
- 9) Nigam, K. and Ghani, R.: Analyzing the Effectiveness and Applicability of Co-training, *Proc. of Information and Knowledge Management Systems*, pp. 86-93 (2000).