# オンライン学習可能なラベル伝播法の研究

永田　耕平† 　佐々木勇和† 　藤原　靖宏†† 　鬼塚　真†

† 大阪大学大学院情報科学研究科　〒 565–0871 大阪府吹田市山田丘 1-5
†† NTT ソフトウェアイノベーションセンタ　〒 180–8585 東京都武蔵野市緑町 3-9-11
E-mail: †{nagata.kohei,sasaki,onizuka}@ist.osaka-u.ac.jp, ††fujiwara.yasuhiro@lab.ntt.co.jp

**あらまし**　半教師あり学習は，ラベルありデータの割合が極端に少ないような領域の問題に対して，高い精度を得ることができることが知られているため，機械学習の分野において盛んに研究が行われている．ラベル伝播法はグラフベースの半教師あり学習の一種であるが，既存のラベル伝播法はデータのインクリメンタルな追加に対応しておらず，データが更新が頻繁に起こると，ラベルの予測値を再計算するコストが無視できなくなるという問題がある．そこで，本研究では，インクリメンタルなデータ追加に対応するラベル伝播法の再計算コスト削減のため，既存のラベルなしノードへの教師ラベルの追加時と，新規のノードの追加時の 2 つのインクリメンタルな追加を対象とした，ラベル伝播法を提案する．教師ラベルの追加に関しては，ラベルの推定値が，既存のラベルに関する部分と新規のラベルに関する部分に線形に分割できることを利用して差分計算を行い，ノードの追加に関しては，既存の推定値を初期値として用い高速な収束実現することで，インクリメンタルなモデルの更新を実現する．評価実験の結果，提案手法が既存手法に比べ，教師ラベルの追加およびノードの追加に関して最大で 10 倍程度の高速化を示した．

**キーワード**　半教師あり学習, ラベル伝播法, オンライン学習

## 1. 序　　論

情報推薦や情報検索から生物学などの様々な分野において，グラフ構造が利用されている．近年では，情報通信技術の進歩により，ノード数が数億にのぼるような大規模なグラフが登場しており，それらを高速に処理できる技術への需要が高まっている．例えば，2016 年 9 月時点で，Facebook の 1ヶ月当たりのアクティブユーザ数は 17.9 億人であると報告されている (注1)．今後，このようなグラフはますます大規模化し，また応用分野も増大するものと考えられる．したがって，大規模なグラフ構造に対する高速な解析手法の開発が重要な課題となっている．

それらの解析手法の一つにデータの分類がある．分類とは各データについてラベルを割り当てる問題である．分類に用いられる機械学習の分野の一つに，半教師あり学習がある．半教師あり学習は学習データとして，ラベルありデータに加えラベルなしデータも用いることができる．ラベルありデータの割合が少ないような領域の問題に対しては，ラベルありデータのみを用いて学習を行う教師あり学習に比べ，半教師あり学習の方が高い精度を得ることができることが知られている [1]．そのため半教師あり学習は多くの現実世界の問題において有用であり，機械学習の分野において，盛んに研究が行われている．半教師あり学習の手法も，それらの特徴によって Self Training [2] や Generative Models [2] などに分類され，また，多くの半教師あり学習手法がこれまでに提案されている [3] [4].

半教師あり学習の一つに Zhou らによって提案されたラベル伝播法（label propagation）[1] がある．ラベル伝播法の応用領域として，ラベル推定済みのデータに，少量のデータあるいはラベルを逐次的に追加していくような状況が想定される．このような少量のデータの追加が頻繁に発生する場合，ラベル推定値の再計算のコストは無視できなくなる．よって本研究では，ラベル伝播法の応用を考えたときの計算コスト削減のため，保存しておいた既存の計算結果を用いることで，計算機の処理時間を短縮する．具体的には，教師ラベルの追加に関しては，各ノードのラベルの推定値が，既存のラベルに関する部分と新規のラベルに関する部分の線形和に分解することで差分計算を行い，既存のラベルの推定値に新規のラベルの推定値を加算する．ノードの追加に関しては，ラベルの推定値の変化は微小であるとの仮定のもと，既存の推定値を初期値として用い高速な収束を実現する．

実際のデータを用いた評価実験の結果，提案手法が既存手法に比べ，教師ラベルの追加及びノードの追加に関して最大で 10 倍程度の高速化を示す結果を確認した．

(注1)：http://newsroom.fb.com/company-info/

## 2. 事 前 知 識

グラフ伝播法は，入力データをノード，ノード間の類似度をエッジによって表したグラフと，ノードの属性としての教師ラベルを入力とし，ラベルの推定値を出力する．

以下でラベル伝播法のアルゴリズムについて説明する．データ数とラベル数をそれぞれ $n, c$ として，ノードの集合とラベルの集合をそれぞれ $\mathcal{X} = \{x_1, \ldots, x_l, x_{l+1}, \ldots, x_n\} \subseteqq \mathbb{R}^m$, $\mathcal{L} = \{1, \ldots, c\}$ とおく．$l$ はラベルありデータ数であり，$x_1, \ldots, x_l$ および $x_{l+1}, \ldots, x_n$ はそれぞれラベルありデータ，ラベルなしデータに対応するノードである．ラベルありノード $x_i (1 \leqq i \leqq l)$ に対する教師ラベルを $y_i \in \mathcal{L}$ とする．また，$\mathcal{F}$ を，各成分が非負の実数である $n \times c$ 型行列の集合とする．$F_0, F_1, F_2, \cdots \in \mathcal{F}$ は，$F_0$ を初期値とする，後述

のアルゴリズムでの反復計算 (5) により求められた行列とする．$F_t = [F_{ij}^{[t]}], t = 0, 1, 2, \ldots$ とおいたとき，$F_{ij}^{[t]}$ は反復回数が $t$ のとき，ノード $i$ がラベル $j$ であると推定される度合いを表す．反復回数が $t$ のときのノード $i$ のラベルの推定値 $\hat{y}_i^{[t]}$ を $\hat{y}_i^{[t]} = \arg\max_j F_{ij}^{[t]}$ と定める．ここで，初期値 $F_0$ にかかわらず $F_t$ は収束することが分かっている [1] ので，$\lim_{y\to\infty} F_t = F^* = [F_{ij}^*]$ とすると，アルゴリズムの出力は $F^*$ を用いたラベルの推定値 $\hat{y}_1, \hat{y}_2, \ldots$ である．$\hat{y}_i$ はノード $x_i$ のラベルの推定値であり，$\hat{y}_i = \arg\max_j F_{ij}^*$ である．

ラベル伝播法のアルゴリズムは以下のようになる．

（1） 教師ラベルに関する行列 $Y = [Y_{ij}] \in \mathbb{R}^{n\times c}$ を構築する．ラベルありデータ $x_i$ に対して $y_i = j$ ならば $Y_{ij} = 1$，それ以外は $Y_{ij} = 0$ とする．

（2） グラフの構造に関する，類似度行列 $W = [W_{ij}] \in \mathbb{R}^{n\times n}$ を構築する．ただし，ループはないものとする．すなわち，$i = j$ のとき $W_{ii} = 0$ とする．

（3） 行列 $S = D^{-1/2}WD^{-1/2} \in \mathbb{R}^{n\times n}$ を計算する．ここで $D = [D_{ij}] \in \mathbb{R}^{n\times n}$ は対角行列であり，$D_{ii} = \sum_j W_{ij} = \sum_j W_{ji}$ とする．

（4） 次の式を $F_t \in \mathbb{R}^{n\times c}, t = 0, 1, 2, \ldots$ が収束するまで反復する．収束結果を $\lim_{y\to\infty} F_t = F^*$ とする．

$$F_{t+1} = \alpha S F_t + (1-\alpha)Y \tag{1}$$

ここで $\alpha$ は $\alpha \in (0, 1)$ を満たす任意の実数である．

収束結果 $F^*$ は，反復計算の初期値 $F_0$ に依存しないことが分かっている [1] ので，$F_0$ は任意であるが，一般的には教師ラベルの信頼性より，$F_0 = Y$ が用いられる．

（5） 得られた $F^*$ より，$\hat{y}_i = \arg\max_j F_{ij}^*$ となるよう各データ $x_i$ にラベル $\hat{y}_i$ を割り当てる．

また，式 (1) を $F_t = F^*$，$F_{T+1} = F^*$ として変形すると，

$$F^* = \lim_{t\to\infty} F_t = (1-\alpha)(I - \alpha S)^{-1}Y \tag{2}$$

が得られる．また，アルゴリズムのステップ 5 より，$F^*$ に正の数を乗じてもラベルの推定結果は変わらない，すなわち，$\arg\max_j F_{ij}^* = \arg\max_j (1-\alpha)F_{ij}^*$ であるから，$F^*$ を，

$$F^* = (I - \alpha S)^{-1}Y \tag{3}$$

と定義し直してもよい．また，反復計算の打ち切り誤差を無視すると，ラベル伝播法のステップ 4 を式 (3) の計算に置き換えても同一の結果が得られることがわかる．[1] では計算速度を向上させるためにべき乗法を用いて計算を行なっている．

ラベル伝播法においてその基礎となる仮定は，入力データをノードとして構築したグラフにおいて，比較的エッジが密な部分グラフの各ノードに対応するデータは，同じラベルを持つ傾向にある，ということである．

ラベル伝播法の入力であるグラフの構造と教師ラベルはトレードオフの関係にあり，このトレードオフはパラメータ $\alpha \in (0, 1)$ で制御される．$\alpha$ は，0 に近いほど正解ラベル，1 に近いほどグラフの構造への依存度を大きくする．本研究では，[1] を参考に $\alpha = 0.99$ を用いる．

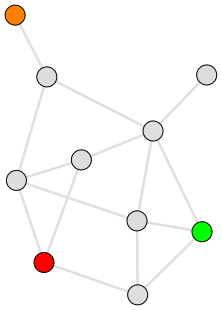

図 1　教師ラベル追加前

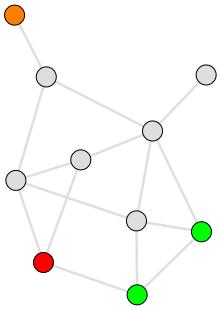

図 2　追加後 (下，緑のラベル)

図 3　教師ラベルの追加

## 3. 提 案 手 法

本研究では，逐次的なデータの追加に対応する，インクリメンタル更新可能なラベル伝播法の提案を行う．手法の設計方針は，保存しておいた既存の計算結果を再利用することで，計算機の処理時間を短縮することである．また，データの追加としては，教師ラベルの追加と，新しいラベルなしノードの追加を想定する．

### 3.1 教師ラベルの追加

教師ラベルの追加について考える．図 3 では，各ラベルを異なる色で表現している．ただし，灰色のノードはラベルなしノードである．この図 3 で言うところのラベルとは，推定ラベルではなく教師ラベルであることに注意する必要がある．この図 3 では，一番下にあるノードに，緑の正解ラベルが追加されている場合を示している．

教師ラベルの追加の場合，グラフの隣接構造自体は変化しない．すなわち，ラベル伝播法のアルゴリズムにおけるラベルの隣接状態を表す行列 $W$，$D$，および $S$ は変化することがない．よってこれらの行列の再計算の必要はなく，あらかじめ計算して保存しておくことによって計算コストを削減することができる．

ここで，ラベル伝播法の式 (3) に着目する．ラベルなしデータ $x_{l+1}, \ldots, x_{l+m}$ の教師ラベルが新たに判明したとする．これら $x_{l+1}, \ldots, x_{l+m}$ に対応するスコアの行列を $Y$ と同様の方法で構築し，それを $\Delta Y = [\Delta Y_{ij}]$ とする．新たなラベル追加後の推定値を $F'^*$ とすると，ラベルの追加によって $S$ は変わらないので，式 (3) より，

$$\begin{aligned} F'^* &= (I - \alpha S)^{-1}(Y + \Delta Y) \\ &= (I - \alpha S)^{-1}Y + (I - \alpha S)^{-1}\Delta Y \\ &= F^* + (I - \alpha S)^{-1}\Delta Y \end{aligned} \tag{4}$$

が成立する．ここで，$F^*$ はすでに計算済みであり，さらに，$(I - \alpha S)^{-1}$ も計算済みである．よって，保持する必要のある行列は $F^*$，$(I - \alpha)$ と，教師ラベルの情報をもつ $Y$ である．この 3 つの行列を保持することで逐次的な再計算の高速化が期待できる．ただし，グラフの構造が変わらない限り $(I - \alpha S)^{-1}$ は更新する必要がないが，$F^*$，$Y$ は再計算するごとに更新する必要がある．

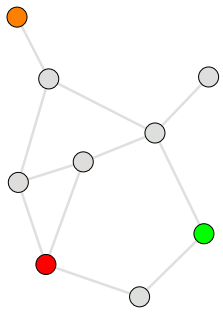

図 4　ノード追加前

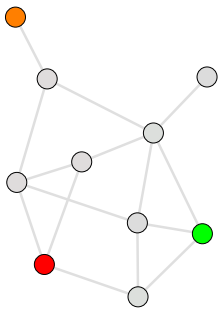

図 5　追加後 (中央右下)

図 6　ノードの追加

したがって，教師ラベルの追加の場合のアルゴリズムは以下のようになる．

（1） 行列 $F^*, Y, (I-\alpha S)^{-1}$ を取得する．

（2） 追加する教師ラベル $y_i = j$ （すなわち $\Delta Y_{ij} = 1$）ごとに，$F'^*$ の $j$ 列に $(I-\alpha S)^{-1}$ の $i$ 列を追加する．

（3） 得られた $F^*$ より，$\hat{y}_i = \arg\max_j F^*_{ij}$ となるよう各データ $x_i$ にラベル $\hat{y}_i$ を割り当てる．

（4） $Y$ を更新し，$F, Y$ を保存する．

### 3.2　ノードの追加

新しいラベルなしノードの追加について考える．図 6 では，右側の図の中央右下に，ラベルなしノードが追加されている．図 6 を見てわかる通り，ノードの追加が，教師ラベルの追加と異なっている点は，グラフの構造そのものが変化してしまうことと，ノード数の増加により，各行列のサイズも大きくなってしまうと言うことである．$k$ NN グラフ [10] の場合は新しいデータと既存の各データとの距離を計算し，新しいデータの $k$ 近傍のデータを調べるだけではなく，既存のすべてのデータについて，新しいデータが既存の $k$ 近傍であるかどうかを調べて，エッジの重み行列 $W$ を更新する必要がある．完全グラフを用いる場合は，$W$ の更新は新しいデータと既存の各データとの距離を計算し，$W$ のサイズを拡張して格納するだけでよいが，行列 $S$ の計算式は $S = D^{-1/2} W D^{-1/2}$ であるから，結局，新しいノードの追加の影響が，新しい行列 $S$ の全体に及んでしまう．したがって，ノードの追加については，教師ラベルの追加の場合のように，計算式を線形な形に分割することが困難である．

そこで，本研究では，反復計算の式 (5) と，前回に計算した行列 $F^*$ に着目する．少数のノードの追加では，新しいノードがグラフ全体に及ぼす結果はわずかだと考えられる．よって，$F^*$ の更新はごくわずかであり，$Y$ よりも反復計算における前回の $F^*$ の方が収束値として得られる $F'^*$ を近似していることから，収束が早い事が期待できる．したがって，初期値としてこの前回の $F^*$ のサイズを拡張してを用いる手法を提案する．

アルゴルズムは次のようになる．ここで，$n$ は既存のデータ数，$n'$ は追加するデータ数，$c$ はラベル数である．

（1） $Y = [Y_{ij}] \in \mathbb{R}^{(n+n')\times c}$ を構築する．ラベルありデータ $x_i$ に対して $y_i = j$ ならば $Y_{ij} = 1$，それ以外は $Y_{ij} = 0$ とする．

（2） $W = [W_{ij}] \in \mathbb{R}^{(n+n')\times(n+n')}$ を構築する．ただし，ループはないものとする．すなわち，$i = j$ のとき $W_{ii} = 0$ とする．

（3） 行列 $S = D^{-1/2} W D^{-1/2} \in \mathbb{R}^{(n+n')\times(n+n')}$ を計算する．ここで $D = [D_{ij}] \in \mathbb{R}^{(n+n')\times(n+n')}$ は対角行列であり，$D_{ii} = \sum_j W_{ij} = \sum_j W_{ji}$ とする．

（4） 次の式を $F_t \in \mathbb{R}^{(n+n')\times c}, t = 0, 1, 2, \ldots$ が収束するまで反復する．初期値 $F_0$ は前回の $F^*$ に，$[0, \ldots, 0]_{1\times c}$ を $n+1 \sim n+n'$ 行目として拡張したものを用いる．収束結果を $\lim_{y\to\infty} F_t = F^*$ とする．

$$F_{t+1} = \alpha S F_t + (1-\alpha) Y \tag{5}$$

（5） 得られた $F^*$ より，$\hat{y}_i = \arg\max_j F^*_{ij}$ となるよう各データ $x_i$ にラベル $\hat{y}_i$ を割り当てる．

## 4.　実　　験

実験では，提案手法による高速性を評価する．指標として計算時間を用い，教師ラベルの追加に関してはベースラインとして，式 (5)，初期値 $Y$ を用いて反復計算するラベル伝播法，式 (3) を用いるラベル伝播法と比較した．ノードの追加に関しては，式 (5)，初期値 $Y$ を用いるラベル伝播法と比較した．パラメータは，[1] [10] を参考に，$k$ -NN グラフを $k = 3$，ラベル伝播法を $\alpha = 0.99$，正解ラベルの割合を 1% とした．

### 4.1　データセット

本実験で使用したデータセットについて説明する．

- Reuters-21578

ロイター通信社により配信されている文書データ (注2)．データ数，ラベル数はそれぞれ 8293，65 である．データ間の類似度には tf-idf を用いる．

- Columbia University Image Library (COIL-100)

様々な 100 個の物体を，5 度づつ回転させながらそれぞれ 72 枚づつ撮影した画像データ (注3)．このため，角度が近い画像同士の類似度が高くなる傾向がある．データ数，ラベル数はそれぞれ 7200，100 である．また， 1 枚の画像は RGB の $32 \times 32$ 画素であるから，特徴量は $32 \times 32 \times 3 = 3072$ 次元のベクトルである．

### 4.2　教師ラベルの追加の評価

通常の反復計算を用いる手法と，逆行列を用いる手法，そして小節 3.2 で説明した提案手法の 3 つについて，計算時間を測定した．逆行列は Python の Numpy モジュールで計算した．ただし，逆行列の計算時間は事前に計算済であるとして，計算時間に含めていない．

結果を図 7 に示す．提案手法が最も高速であるという結果になった．また，逆行列の計算が最も低速であった．ラベル推定値 $F^*$ を得るために既存手法が反復計算を行なっているところを，提案手法 1 回で済ますことができているためだと考えられ

(注2)：http://www.daviddlewis.com/resources/testcollections/reuters21578/

(注3)：http://www.cs.columbia.edu/CAVE/software/softlib/coil-100.php

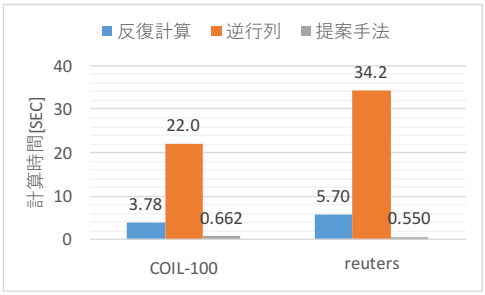


図 7　ラベル追加時の計算時間

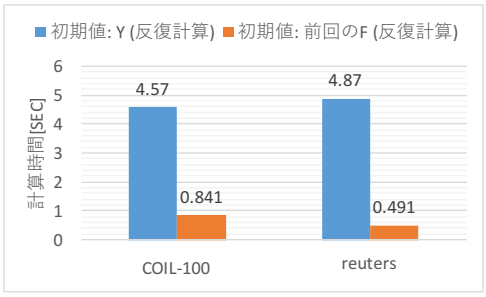


図 8　ノード追加時の計算時間

る．また逆行列の計算は規模の大きな行列ではコストが高いため，最も低速になるとなると考えられる．

### 4.3　ノードの追加の評価

追加するデータを除いたデータでまずラベル伝播法を実行し，既存のデータにラベルなしデータを 1 つ追加して，再度ラベル伝播法を実行した．精度の指標としては，$||F'^* = F^*||_\infty$ で定義した誤差を用いた．新たに追加するデータは，ラベルなしデータとした．

結果は図 8 のようになった．どちらのデータセットにおいても，前回の $F^*$ を初期値として用いた方が高速であった．前回の結果 $F^*$ よりも $Y$ の方が，$F'^*$ をよく近似しており，収束が速いためだと考えられる．

## 5.　関 連 研 究

ラベル伝播法は，グラフ上でのランダムサーファのふるまいのモデル化とも見なすことができ，その点で，Google の検索エンジンにおいて使用されている Pagerank アルゴリズム [5] とも関連のある [6] アルゴリズムである．よって，PageRank の高速化の手法をラベル伝播法の高速化に応用することもできる．[7] では，ノードを既存のものと変更されるものに分割し，それぞれの集合について，集約処理を用いる事で，PageRank 値の近似を行う．[8] では，モンテカルロ法を用いて PageRank 値の近似を行う．[9] では，グラフの一部分を逐次的にクロールすることで PageRank 値の近似を行う．

## 6.　結　　論

本研究では，グラフベースの半教師あり学習の一手法であるラベル伝播法をインクリメンタルなデータの追加に対応させる研究を行った．提案手法ではデータの追加を，教師ラベルの追加とノードの追加に分けて，それぞれに対して手法を提案した．教師ラベルの追加については，グラフの構造が変化しないため，前回の推定スコアとグラフグラフの構造に関する行列の項を線形に分割できることを用いた．それぞれの行列のデータをあらかじめ保存しておき，追加時に取り出して計算を省略する手法を提案した．また，ノードの追加についても，前回の推定スコアを保存しておく手法を提案した．提案手法についての実験では，どの手法についても良好な結果を得ることができた．

今後として，他のグラフベースの半教師あり学習手法や PageRank 等のランダムウォークに関する手法などを調査し，それらを様々なデータセットに適用して，推定精度や計算時間を比較し，インクリメンタルなデータの追加に対応する手法を考案したい．

## 文　　献

[1] Zhou, Dengyong, et al. ”Learning with local and global consistency.” Advances in neural information processing systems 16.16, 2004.
[2] Xiaojin Zhu, Andrew B. Goldberg. Introduction to Semi-Supervised Learning, Morgan and Claypool Publishers, 2009.
[3] Zhu, Xiaojin, Zoubin Ghahramani, and John Lafferty. ”Semi-supervised learning using gaussian fields and harmonic functions.” ICML. Vol. 3. 2003.
[4] Goldberg, Andrew B., and Xiaojin Zhu. ”Seeing stars when there aren’t many stars: graph-based semi-supervised learning for sentiment categorization.” Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing. Association for Computational Linguistics, 2006.
[5] Page, Lawrence, et al. ”The PageRank citation ranking: bringing order to the web.”, 1999.
[6] Zhou, Dengyong, et al. ”Ranking on data manifolds.” Advances in neural information processing systems 16, 2004.
[7] Langville, Amy Nicole, and Carl Dean Meyer. ”Updating pagerank with iterative aggregation.” Proceedings of WWW, 2004.
[8] Bahmani, Bahman, Abdur Chowdhury, and Ashish Goel. ”Fast incremental and personalized pagerank.” Proceedings of the VLDB Endowment 4.3, 2010.
[9] Bahmani, Bahman, et al. ”Pagerank on an evolving graph.” Proceedings of the 18th ACM, 2012.
[10] Von Luxburg, Ulrike. ”A tutorial on spectral clustering.” Statistics and computing 17.4 (2007).