

賛否表明パターンと行列分解に基づく賛否モデリング

佐々木 彬 埴 一晃 岡崎 直観 乾 健太郎

東北大学

{aki-s, hanawa, okazaki, inui}@ecei.tohoku.ac.jp

1 はじめに

ソーシャルメディアの普及は、人々のニュースとの関わり方を大きく変えた。Pew Research Centerの最新の調査によると、米国の成人の約62%はニュースをソーシャルメディア経由で取得しており、そのうち64%は一つのソーシャルメディアのみに依存している[1]。興味深かったニュースは友人に共有され、エコーチェンバー現象¹[2]やフィルターバブル²[3]を経て、イデオロギー的偏向を形成していく[4]。ソーシャルメディアは人々の意見・興味の宝庫であり、そのユーザの意見を分析することは、世論調査、選挙予測、選挙勧誘、オンラインディベートなど、幅広い応用に直結する。

このような背景の中、スタンス分析(stance detection)と呼ばれる新しいタスクが提案され、多くの研究が発表されている[5, 6, 7, 8, 9, 10]。このタスクでは、対象(選挙の候補者や政策など)とテキストが与えられた時、そのテキストの著者が対象に賛成しているか、反対しているか、どちらとも言えないかを予測する。例えば「憲法96条の改正」という対象と「憲法96条の改正は危ない」というテキストが与えられた時、この著者は対象に反対していると推測する。これは「Aは危ない」という言語パターンが反対の意見表明であり、Aに対象が埋め込まれていることから、比較的容易に推測できる。

しかし、スタンス分析においてこのような簡単な例は稀で、実用的と言えるには程遠いとの指摘もある[11]。これには様々な要因があるが、スタンス分析の既存手法が賛否対象やユーザが限定された訓練事例に依存しており、賛否対象やユーザに関する一般的な知識を活用できていないことに問題があると考えている。例えば「総議員の2/3以上の賛成で発議」というのはおかしい」という発言から「憲法96条の改正」という対象に関するスタンスを認識するには、「総議員の2/3以上の賛成で発議」と「憲法96条の改正」の関連性を理解しておく必要がある。「総議員の2/3以上の賛成で発議」は「憲法96条の改正」の導入前の事象、すなわち「改正」が活性化されない場合であること、「Aというのはおかしい」は反対表明の言語パターンであることを使うと、この発言は対象に賛成していると推定できる。

このように「Aに賛成する人はBにも賛成する」や「Aに反対する人はBにも反対する」のような知識を獲得しておくことは、スタンス分析およびその応用において非常に重要である。そこで、本研究では賛否が

議論の争点となるような対象に関して、世の中の人々の賛否の傾向をモデリングする。まず、ツイートに含まれるハッシュタグを利用して、「Aは認められない」など、人々が賛成・反対を表明するパターンを抽出する。次に、そのパターンをツイートに適用することで、Twitterのユーザがどのような対象に賛成・反対を表明しているのかをマイニングする。Twitterのユーザは全ての対象について賛否を表明するのではなく、ごく一部の対象についてのみ言及するので、マイニングにより獲得した人々の賛否は非常にスパースとなる。そこで、商品推薦と同様の考え方に基づき、ユーザの賛否を表す行列に行列因子分解(matrix factorization)を適用し、対象に対する賛否を表す低次元ベクトルを求める。このベクトルを応用することで、「Aに賛成する人はBにも賛成する」といった対象間の相関関係や、ユーザが賛否を表明していない対象の賛否を予測できることを実証する。

2 賛否表明パターンによる賛否マイニング

2.1 賛否表明パターンの獲得

本節では、Twitter上で「誰が何に賛成/反対を表明したのか」に関する事例を抽出する手法を説明する。抽出元のデータとして、2013年2月6日から2016年9月30日までの42,782,881,167ツイートをを用いた。なお、リツイートは分析の対象から除外した。

まず、「# 賛成」あるいは「# 反対」というハッシュタグを含むツイートを抽出する。以下、これらのハッシュタグをそれぞれ賛成ハッシュタグ、反対ハッシュタグと呼び、両者をまとめて賛否ハッシュタグと呼ぶ。また、「」や「」に入る文字列を賛否対象、もしくは単に対象と呼ぶ。以上の処理により抽出されるツイートの例を示す。

- (1) (ユーザA) ありえない! #特定秘密保護法反対
- (2) (ユーザB) それが見たい #憲法9条改正賛成

(1)は「特定秘密保護法」という対象に関して反対ハッシュタグを含み、(2)は「憲法9条改正」という対象に関して賛成ハッシュタグを含む。ハッシュタグ付きの投稿は、強い意見表明と捉えることができるので、ユーザAは「特定秘密保護法」という対象に反対、ユーザBは「憲法9条改正」という対象に賛成であることが分かる。以上の処理により、30,698人のユーザと7,800件の対象に関して、58,917件の賛否表明の事例を収集できた。

このようにして抽出したデータはユーザの賛否を正確に反映すると考えられるが、ハッシュタグ付きの投

¹同じ興味を持つ集団の中で考え方や思想が増幅される現象。

²(ニュース配信などの)システムがパーソナライゼーションを進めた結果、ユーザが自分の意見に合わない情報から隔離される現象。

稿は非常に少ないため、これだけでは賛否表明の事例を十分に収集できない。そこで、ユーザの賛否は一貫しているという仮定に基づき、賛否を強く示唆する賛否表明パターンをマイニングする。例えば、(1)と(2)のユーザがそれぞれ、以下の投稿を行ったとする。

(3) (ユーザ A) 特定秘密保護法 は認められない

(4) (ユーザ B) 憲法 9 条改正 を実施すべきだ

これらのツイートは賛否ハッシュタグを含まないが、ユーザ A が「特定秘密保護法」に反対、ユーザ B が「憲法 9 条改正」に賛成であることは(1)と(2)で把握しているので、これらのツイートも同様の賛否を表明すると期待してもよい。そこで、(3)と(4)のツイート中の賛否対象を変数化し、それぞれ、「X は認められない」という反対パターン、「X を実施すべきだ」という賛成パターンを抽出する。このように、賛否ハッシュタグを付けて投稿したユーザの別の投稿の中で、賛否対象が言及されるパターンを統計的に調べ上げることで、賛成もしくは反対を強く示唆する賛否表明パターンを獲得できる。

ただし、以上の手順だけで賛否表明パターンを獲得すると、「X へ」「X なんです」のような有用ではないパターンが獲得されてしまう。そこで、以下の手順により人手で賛否表明パターンを選別した。

1. 賛成パターン、反対パターンのそれぞれについて、パターンを使ったユーザ数の降順に並び替える。
2. ユーザ数上位のパターンから順に人手で確認し、ノイズと判断されるものを削除する。
3. 賛成パターン、反対パターン各 100 件を取得した時点で終了する。

以上の手順により「X 歓迎」「X は必要」のような賛成パターン、「X させるな」「X 許さない」のような反対パターンが獲得される。

2.2 ユーザの賛否表明事例の収集

2.1 節で獲得した賛否表明パターンをツイートに適用し、ユーザの賛否表明事例を収集する。抽出元のデータとして、2015 年 10 月 1 日から 2016 年 9 月 30 日までの 12,145,619,420 ツイートを用いた。なお、リツイートは分析の対象から除外した。

あるツイートを文に分解し、文末が賛成表明パターンのいずれかにマッチし、その直前に 2.1 節で抽出した賛否対象 v が出現しているとき、そのツイートの投稿者 u は v に賛成を表明したと認定する。同様に、文末が反対表明パターンのいずれかにマッチし、賛否対象 v が直前に出現しているとき、そのツイートの投稿者 u は v に反対を表明したと認定する。以上の処理を全ツイートに適用したところ、311,937 人のユーザと 1,445 件の賛否対象に関して、481,109 件の賛否表明が抽出された。

3 ユーザ・賛否対象行列の行列因子分解

2.2 節の方法により、多くの Twitter ユーザの賛否表明事例を収集することができる。しかし、Twitter のユーザが全ての対象について賛成・反対を表明する訳ではないので、ほとんどの対象に関する賛否は不明と

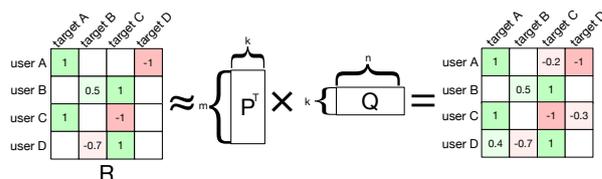


図 1: ユーザ・賛否対象行列に対する行列因子分解。

なる。また、Twitter のデータに含まれていないユーザの賛否を予測することは本質的に不可能である。そこで、商品推薦と同様の考えに基づき、ユーザの賛否を表す行列に行列因子分解を適用し、ユーザや対象に関する特徴ベクトルを求める。対象に関する特徴ベクトルを用いることで、「A に賛成する人は B にも賛成する」といった対象間の相関関係や、賛否を表明していない対象の賛否を予測できるようになる。

2.2 節で賛否表明が収集されたユーザの総数を m 、賛否対象の総数を n とし、 m 行 n 列の疎行列 R を用意する。あるユーザを $u \in \{1, \dots, m\}$ 、ある賛否対象を $v \in \{1, \dots, n\}$ というインデックスで表す。ユーザ u が対象 v について、賛否表明を行っていた場合は、疎行列 R の u 行 v 列の要素 $r_{u,v}$ を次式で求める。

$$r_{u,v} = \frac{\#(u,v,+1) - \#(u,v,-1)}{\#(u,v,+1) + \#(u,v,-1)} \quad (1)$$

ここで、 $\#(u,v,+1)$ と $\#(u,v,-1)$ は、それぞれユーザ u が対象 v に賛成、反対を表明した回数である。したがって、行列 R の要素 $r_{u,v}$ はユーザ u が対象 v について賛成に傾いている場合は $+1$ に近づき、反対に傾いている場合は -1 に近づく。なお、ユーザ u が対象 v について賛否表明を行っていない場合は、式 1 の計算を行わず、疎行列 R の要素 $r_{u,v}$ は存在しないこと(欠損値)とする。行列 R はユーザ・賛否対象行列と呼ばれる。

このように定義した疎行列 R に対して、行列因子分解を適用し、低次元行列 $P \in \mathbb{R}^{k \times m}$ と $Q \in \mathbb{R}^{k \times n}$ の積に分解する。行列 P と Q は以下の目的関数を最小化することにより求める。

$$\min_{P,Q} \sum_{(u,v) \in R} (r_{u,v} - p_u^T q_v)^2 + \lambda_P \|p_u\|^2 + \lambda_Q \|q_v\|^2 \quad (2)$$

ここで、 $(u,v) \in R$ は疎行列の中で埋められている要素のインデックス (u 行 v 列) に対する繰り返し、 $p_u \in \mathbb{R}^k$ と $q_v \in \mathbb{R}^k$ は、それぞれ P の u 列ベクトル、 Q の v 列ベクトルである。また、 $\lambda_P \geq 0$ と $\lambda_Q \geq 0$ は正則化項の係数である。以下、 p_u をユーザ u の特徴ベクトル、 q_v を賛否対象 v の特徴ベクトルと呼ぶ。このようにして求めた特徴ベクトルを用いると、ユーザ u の賛否対象 v に対する賛否スコアの推定値 $\hat{r}_{u,v}$ は、次式で計算できる(図 1)。

$$\hat{r}_{u,v} \simeq p_u^T q_v \quad (3)$$

なお、式 2 の最小化には、行列因子分解の実装である libmf³ [12] を用いた。

³<https://github.com/cjlin1/libmf>

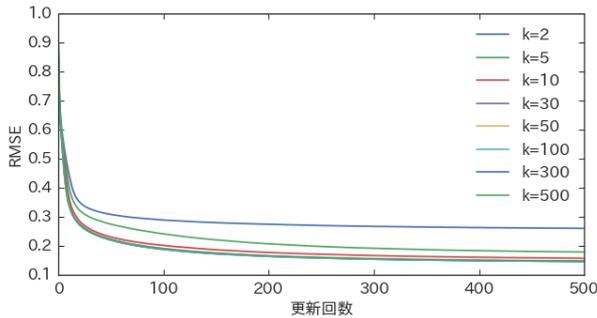


図 2: 次元 k を変化させたときの RMSE の推移 .

4 実験

4.1 次元パラメータ k と行列因子分解の誤差

ユーザ・賛否対象行列 R を $P \in \mathbb{R}^{k \times m}$ と $Q \in \mathbb{R}^{k \times n}$ の積に分解するとき、次元パラメータ k はどのくらいに設定すればよいのか？ この問いに答えるため、次元パラメータ k を変化させたときの行列因子分解の誤差を調査する。誤差を測定する尺度として、Root Mean Squared Error (RMSE) を用いる。

$$RMSE = \sqrt{\frac{\sum_{(u,v) \in R} (p_u^T q_v - r_{u,v})^2}{N}} \quad (4)$$

ここで、 N は疎行列 R の要素数（既知の値の数）である。なお、正則化の係数は $\lambda_P = 0.1$, $\lambda_Q = 0.1$ とし、それ以外のパラメータは `libmf` のデフォルト値を採用した。

$k = 2, 5, 10, 30, 50, 100, 300, 500$ と変化させたとき、`libmf` が P と R を更新するにつれて RMSE がどのように推移したのかを図 2 に示した。この結果より、 $k = 100$ で元の行列 R を十分に再現できていると判断し、以降の実験では $k = 100$ に固定した。

4.2 欠損している賛否の予測実験

本研究で求めた特徴ベクトルは、ユーザの賛否の欠損値をどのくらい正確に予測できるのか？ これを調査するため、以下の手順で欠損値の予測精度を評価した。

1. ユーザ・賛否対象行列で t 個以上の賛否対象に賛否を表明しているユーザの集合を求める。
2. これらのユーザ集合を 5 個のグループに分割する。
3. 5 分割交差検定のように、4 個のグループの賛否表明を学習データと見なし、ユーザ・賛否行列の行列分解により特徴ベクトルを求め、残りの 1 グループの賛否を予測するというプロセスを 5 回繰り返す。このとき、賛否を予測するグループでは、各ユーザの賛否対象のうち 80% の賛否が与えられ、残りの 20% を欠損値として予測し、その精度を測定する。欠損値の予測では、予測されたスコア $\hat{r}_{u,v}$ が非負であれば賛成、負であれば反対とした。
4. 各ユーザ集合に対する予測精度の平均を算出する。

図 3 に、 t の値を変化させた時の予測精度を示した。比較対象として、各賛否対象毎に学習データの賛成・

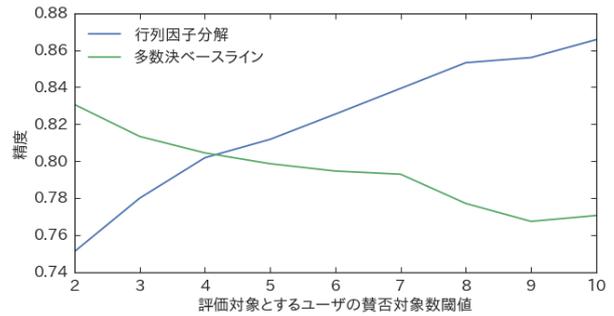


図 3: t 回以上賛否表明をしたユーザに対する賛否予測の精度（5 分割交差検定）。

反対が優勢だった方を選ぶという多数決ベースラインの結果を掲載した。この実験結果より、 $5 < t$ の時は特徴ベクトルによる欠損値補完が多数決ベースラインの性能を上回ること、より多くの賛否表明を行っているユーザほど、欠損している賛否の予測が高精度に行えることが分かった。なお、多数決ベースラインは t が大きくなるにつれて精度が緩やかに低下している。これは賛否表明が少ないユーザの賛否は世の中の多数派の賛否から予測しやすいが、賛否表明が多いユーザは政治信条を持っているため、世の中の多数派の賛否から乖離していくためだと考えられる。

表 1 に、特徴ベクトルにより予測される賛否の例を示した（ユーザ名は匿名化してある）。この表では、ユーザ A と B に関して、表明されていた賛否からどのような対象の賛否が強く予測されるのかを示している。この結果より、例えば「強制送還」に賛成して「難民受け入れ」に反対しているユーザ A は、「通名廃止」に賛成、「外国人参政権」「移民受け入れ」「外国人生活保護」に反対と予測されるなど、世の中のイデオロギーを反映した形で賛否の予測が行われていることが確認できる。

4.3 賛否対象の特徴ベクトルの類似度

行列因子分解で求めた賛否対象の特徴ベクトルは、「A に賛成する人は B にも賛成する」とような情報を捉えられているのだろうか？ この評価のための正解データを作ることは非現実的であるため、ここでは定性的な事例確認を行う。表 2 は、「TPP」「原発」「夫婦別姓」「消費税」「集団的自衛権」の特徴ベクトルとコサイン類似度が高い賛否対象を求めた結果を示した。特徴ベクトル「TPP」と「TPP」（半角と全角）、「消費税」と「消費税増税」といった同義語が得られるだけでなく、「原発」と「核燃サイクル」、「消費税」と「金融引き締め」など、関連しそうな賛否対象の類似度が高くなっている。したがって、本研究で求めた賛否対象の特徴ベクトルが、世の中の人々の賛否の傾向を反映していることが確認できた。

5 おわりに

本論文は、「A に賛成する人は B にも賛成する」や「A に反対する人は B にも反対する」とような知識を獲得するため、世の中の人々の賛否の傾向をモデリングする方法を紹介した。ツイートに賛否表明パターンを適用して獲得したユーザ・対象行列に対して、行列因子分解を適用することで、ユーザや賛否対象に関する

表 1: ユーザごとに補完した賛否対象の例 (カッコ内は賛否の予測スコア $\hat{r}_{u,v}$) .

ユーザ	種別	対象
A	賛成 (表明)	空爆/強制送還
	反対 (表明)	難民受け入れ/移民難民受け入れ/反日ヘイト/難民移民受け入れ/外国人生保/偏向報道/鎖国/日韓スワップ協定/通貨スワップ/舛添
	賛成 (予測)	死刑制度 (0.6461)/通名廃止 (0.6182)/パチンコ廃止 (0.6047)/電波停止 (0.5962)/パチンコ規制 (0.5880)/偽装難民 (0.5595)/制裁強化 (0.5317)/ラーメン (0.5249)/食べて応援 (0.5108)/靖国参拝 (0.4925)
	反対 (予測)	外国人参政権 (-0.7620)/移民受け入れ (-0.7258)/外国人生活保護 (-0.7032)/移民政策 (-0.6663)/沖縄独立 (-0.6518)/児童ポルノ (-0.6459)/外国人労働者受け入れ (-0.6366)/外国人生活保護支給 (-0.6251)/消費税引き上げ (-0.6093)/石破 (-0.6065)
B	賛成 (表明)	政権交代
	反対 (表明)	最終処分場/安倍総理/消費税/マイナンバー/マイナンバーカード/マイナンバー制度/集団的自衛権/新国立競技場/安保/安保法案
	賛成 (予測)	国民連合政府 (0.8520)/火事場飯 (0.8338)/ハッシュタグ荒らし (0.8321)/棄民 (0.8270)/ノーベル賞受賞 (0.8069)/どみの (0.7833)/通名廃止 (0.7760)/雇用創出 (0.7637)/小遣い制 (0.7620)/国連脱退 (0.7554)
	反対 (予測)	安全保障関連法案 (-0.9957)/特定秘密保護法 (-0.9783)/強行判決 (-0.9710)/電磁波攻撃 (-0.9507)/原発事故 (-0.9480)/精神攻撃 (-0.9464)/帰還政策 (-0.9459)/安保法案絶対 (-0.9405)/未成年飲酒 (-0.9367)/安倍政権 (-0.9334)

表 2: 賛否対象の特徴ベクトルとのコサイン類似度が高い賛否対象 (括弧内はコサイン類似度) .

賛否対象	コサイン類似度が高い賛否対象
TPP	戦争ビジネス (0.8691)/移民難民 (0.8304)/TPP 交渉参加 (0.8267)/日韓スワップ再開 (0.8050)/T P P (0.7831)/民泊規制緩和 (0.7689)/安全保障関連法 (0.7672)/水道事業民営化 (0.7638)/残業代ゼロ法案 (0.7615)/T P P 批准 (0.7607)
原発	戦争法案成立 (0.9691)/核燃サイクル (0.9527)/伊方原発 (0.9513)/原発稼働 (0.9442)/伊方原発再稼働 (0.9275)/強行採決 (0.9266)/戦争法案絶対 (0.9265)/川内原発再稼働 (0.9258)/ テクノロジー犯罪 (0.9179)/風評被害 (0.9178)
夫婦別姓	戸籍廃止 (0.8336)/男女共同参画 (0.6435)/外国人参政権 (0.6285)/朝鮮学校 (0.6085)/日韓通貨スワップ (0.5734)/負担増 (0.5718)/移民受入 (0.5714)/移民受け入れ (0.5623)/移民難民 (0.5466)/内政干渉 (0.5404)
消費増税	金融引き締め (0.8064)/法人税引き下げ (0.6272)/消費税増税 (0.6215)/AIIB 参加 (0.5933)/自民改憲草案 (0.5903)/遺伝子組み換え作物 (0.5790)/民泊規制緩和 (0.5714)/日中韓 FTA(0.5409)/ニャンタイコ (0.5385)/外国人参政権付与 (0.5330)
集団的自衛権	米軍基地移設 (0.7760)/安全保障法制 (0.7308)/自民改憲草案 (0.7080)/核発電 (0.7072)/法人税引き下げ (0.6798)/集団的自衛権行使 (0.6760)/基地建設 (0.6725)/子宮頸がんワクチン接種 (0.6544)/防衛装備移転三原則 (0.6385)/チート (0.6190)

特徴ベクトルを求めた。評価実験では、欠損しているユーザの賛否表明を約 0.75 (2 件の賛否表明だけを手がかりに予測する場合) から 0.87 (10 件の賛否表明を手がかりに予測した場合) の精度で予測できることを示した。また、賛否対象の特徴ベクトルのコサイン類似度を計算することで、「TPP に賛成する人は戦争ビジネスにも賛成する」といった対象間の相関関係が得られることを示した。

本研究をより一般化すると、「A に関心がある人は B に賛成する」とか「A を行った人は B に賛成する」「A に賛成する人は B に投票する」など、賛否対象だけでなく興味や行動などの異種のシグナルを横断した分析が可能になるため、言語処理の新しい応用を開拓できると考えている。また、本研究で求めた特徴ベクトルを分散表現として見なし、畳み込みニューラルネットワークなどで分類器を構築したとき、賛否対象間の知識 (相関関係など) が自然に取り込まれたスタンス分析になるのか、実験を進めていきたい。

謝辞

本研究は、文部科学省科研費 15H01702, 15H05318, および JST, CREST の支援を受けたものである。

参考文献

[1] J. Gottfried and E. Shearer, “News use across social media platforms 2016,” Pew Research Center, Tech. Rep., May 2016.

[2] K. H. Jamieson and J. N. Cappella, *Echo Chamber: Rush Limbaugh and the Conservative Media Establishment*. Oxford University Press, 2008.

[3] E. Pariser, *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. Penguin Books, 2011.

[4] E. Bakshy, S. Messing, and L. A. Adamic, “Exposure to ideologically diverse news and opinion on facebook,” *Science*, vol. 348, no. 6239, pp. 1130–1132, 2015.

[5] S. Somasundaran and J. Wiebe, “Recognizing stances in online debates,” in *Proc. of ACL-IJCNLP 2009*, 2009, pp. 226–234.

[6] A. Murakami and R. Raymond, “Support or oppose?: classifying positions in online debates from reply activities and opinion expressions,” in *Proc. of COLING 2010*, 2010, pp. 869–875.

[7] K. S. Hasan and V. Ng, “Why are you taking this stance? identifying and classifying reasons in ideological debates,” in *Proc. of EMNLP 2014*, 2014, pp. 751–762.

[8] A. Rajadesingan and H. Liu, “Identifying users with opposing opinions in twitter debates,” in *Social Computing, Behavioral-Cultural Modeling and Prediction*. Springer, 2014, pp. 153–160.

[9] D. Sridhar, L. Getoor, and M. Walker, “Collective stance classification of posts in online debate forums,” in *Proc. of the Joint Workshop on Social Dynamics and Personal Attributes in Social Media*, 2014, pp. 109–117.

[10] A. Ramesh, S. H. Kumar, J. Foulds, and L. Getoor, “Weakly supervised models of aspect-sentiment for online course discussion forums,” in *Proc. of ACL-IJCNLP 2015*, 2015, pp. 74–83.

[11] R. Cohen and D. Ruths, “Classifying political orientation on Twitter: It’s not easy!” in *Proc. of the Seventh International AAAI Conference on Weblogs and Social Media*, 20013, pp. 91–99.

[12] W.-S. Chin, Y. Zhuang, Y.-C. Juan, and C.-J. Lin, “A fast parallel stochastic gradient method for matrix factorization in shared memory systems,” *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 6, no. 1, p. 2, 2015.