ソーシャルトレーディングサービスにおけるトレーダの特徴分析

竹田　創† Chenyi Zhuang† 馬　強†

† 京都大学情報学研究科社会情報学専攻　〒 606–8501 京都市左京区吉田本町
E-mail: †{takeda,zhuang}@db.soc.i.kyoto-u.ac.jp, ††qiang@i.kyoto-u.ac.jp

あらまし　近年，SNS のようにフォローするだけで他のトレーダの売買が複製され，自動的に取引を行うことができるソーシャルトレーディングという投資手法が注目されている．投資商品についての知識は要求されないが，複製するトレーダの選別が重要となる．トレーダの特徴を分析して，複製するフォロイーの適切な選択を支援することが重要である．そのため本稿では，取引履歴とニュース記事を用いた行列因子分解モデルによってトレーダの成績とニュースイベントの関連性を発見する手法を提案する．

キーワード　ソーシャルトレーディング 非負値行列因子分解　情報推薦

1. はじめに

近年ソーシャルトレーディングの人気が高まっており，ZuluTrade(注1) や eToro(注2) など多くのサービスが存在する．ソーシャルトレーディングは，投資家が意見や投資情報を共有できる新しい投資手法である．一般的な投資では，ファンダメンタル分析やテクニカル分析などの手法で市場での需要と供給を分析し，適切な価格を推定して実際に商品を売買する．一方ソーシャルトレーディングでは，ユーザはパフォーマンスの高いトレーダを発見し，そのトレーダをフォローすることによって，フォローしたトレーダの取引が自動にコピーされ，取引を行うことができる．つまりソーシャルトレーディングサービスは，商品ではなく人をベースとした投資戦略を組むことが特徴である．

ソーシャルトレーディングサービスでの利益額は，フォローするトレーダのパフォーマンスに依存する．従って公開されているトレーダの情報を考慮して適切なトレーダを選ぶことが重要である．公開されているトレーダの情報には，過去の取引履歴，評価指標，プロフィールやランキングなどがある．しかしながら，次の二点においてトレーダの選択が難しく，掲示板や質問サイトでも、頻繁に議論されている．

- ソーシャルトレーディングにおいてトレーダを評価する確立された指標が存在しない．そのため多くの評価指標が提示されており，どれに着目すればよいのかが難しい．
- サービスが提供している評価指標は過去の一定期間の結果に基づいて算出されるため，日々の変動が激しい．トレーダのランキングの変動も激しく，ハイリスクハイリターンのトレーダがランキングの上位になりやすい．

そこで，Lee らは，トレーダの取引履歴を分析して，performance，risk と consistency の三つの評価指標を提案し，トレーダの安定的なランキングを得ようとしている [9]．しかしながら，Lee らの手法はトレーダの得意な分野や市場の局面を考慮しておらず，局面が変化した場合に不適切なトレーダを推薦する場合がある．

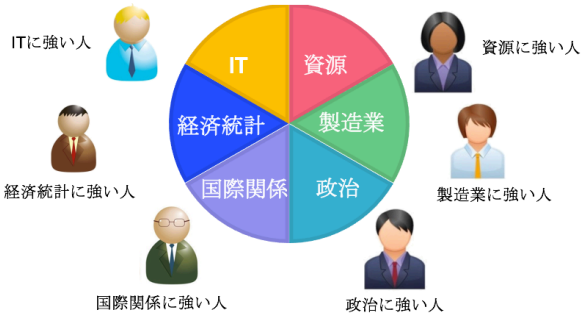

図 1　ソーシャルトレーディングで安定した投資を行うためには，人をベースとしたポートフォリオを考慮する必要がある．図はそのイメージである．

本研究では，主要なソーシャルトレーディングサービスの一つである ZuluTrade のデータを利用する．また，このサービスでは外国為替証拠金取 (FX) の取引が行われており，為替取引でのソーシャルトレーディングを対象とする．

本研究の目標は，各トレーダの得意なニューストピックを抽出することである．それにより，為替相場でそのときどきの変動の要因となっているニューストピックに適した，トレーダの選択に生かすことができる．

為替相場は，経済指標の発表，政治的決定や有事など様々なニュースイベントで大きく変動し，トレーダもニュースイベントを考慮して取引を行っている．(図 2)．ソーシャルトレーディングでは，トレーダごとに精通・着目するニューストピックは異なり，自己紹介文で自分が着目しているニュースイベントを明言するトレーダも存在する．例えば，雇用統計や GDP などの経済指標の発表に着目していると明言するトレーダや，資源の価格に着目していると明言するトレーダを観測している．

そこで本研究では，ニュース記事に報道されているイベント，取引履歴とニューストピックの関連性を考慮して，協調型非負値行列因子分解 (Collective NMF [15] を利用して，トレー

(注1)：https://zulutrade.com
(注2)：https://www.etoro.com

図 2　ニュースイベントは為替相場が大きく変動する．トレーダはメディアによってニュースイベントを確認する．

ダの特徴分析手法を提案する．提案手法では，ニュース記事と取引履歴の観測を，それぞれ，ニューストピックと単語，およびニューストピックとトレーダの特徴に因子分解することで，トレーダの得意とするニューストピックを明らかにする．これにより，為替相場でそのときどきの変動の要因となっているニューストピックに適した，トレーダの選択に生かすことができる．安定的な投資を行うには，一般的にリスクの分散を目的とした投資ポートフォリオを考える必要がある．ソーシャルトレーディングでは，人をベースとしたポートフォリオを構築する必要があり，本研究はそのサポートができると考えている．(図 1)．本研究の貢献は，次のようにまとめられる．

- 因子行列の制約を考慮した行列因子分解により，テキストデータ（ニュース）と数字データ（取引履歴）を同時に扱うトレーダの特徴分析手法を提案した．
- 協調型非負値行列因子分解に β-divergence を用いた逐次更新アルゴリズムを利用するなどの工夫をし，実際のソーシャルトレーディングにおいて有用であるということを示した．

本論文の構成は次の通りである．2 節では関連研究を示し，3 節では本研究で提案するモデルについて記す．4 節ではモデルの学習，5 節では実験結果について説明する．6 節はまとめである．

2. 関 連 研 究

2.1 ソーシャルトレーディング

Pan らは，ソーシャルトレーディングサービス eToro を対象に調査を行い，一般的な投資に比べてソーシャルトレーディングサービスが高い ROI の期待値を得ることを示している [14]．

Lee らは，一般的な投資で利用される評価指標であるリスクとリターンに加え，ソーシャルトレーディングサービス特有の評価指標として consistency という評価指標を提案している [9]．これは，トレーダの取引の一貫性を意味し，利益などの標準偏差などから計算し，これによりソーシャルトレーディングにおいて適切にトレーダの総合的評価ができることを示した．各指標から線形モデルによって総合的なスコアを算出しトレーダを推薦するシステムも提案している．

しかしながら Lee らの手法は，トレーダの得意なトピックを考慮しておらず，推薦されたトレーダが局面が変化した場合でも適切であるとは限らない．また，トレーダの各評価指標の重み付けがヒューリスティックであるという問題点がある．本研究は，ニュースイベントによって潜在的なトピックを考慮することで，現在の局面に合うトレーダを発見できるという点で異なる．

2.2 エキスパートマイニング

ソーシャルネットワークサービスにおける，オピニオンリーダや専門家など権威のあるユーザを発見することが重要であり，オピニオンマイニング，ソーシャルネットワーク分析が提案されている ([3], [10], [11])．ユーザ関係に基づく手法では，El-Korany [3] は Stackoverflow のようなオンラインコミュニティの専門家を推薦する方法を提案している．オンラインコミュニティは，通常，一般ユーザと専門家で構成されるため，ソーシャルネットワークサービスの評価機能を利用して，他社から高い評価を得ているユーザが専門家であるとみなし，エキスパートユーザを識別している．Li ら [10] は，Twitter において，継続時間マルコフプロセス（IDM-CTMP）に基づく情報拡散モデルと呼ばれる，他のユーザにどの程度影響を与えたかを予測する方法を提案している．具体的には，Twitter の投稿やそれがリツイート，お気に入りされる頻度に基づく分析である．馬ら [11] は，人物検索のために URL やスニペット，検索結果数などから総合的に判断した famousness とよぶランキング指標を提案した．

しかしソーシャルトレーディングは発展途上のサービスであり，上記のような人や投稿を評価する機能やメタデータが備わっていない．そのため，提案された手法を直接適用することはできない．

2.3 非負値行列因子分解

非負値行列因子分解（NMF：Nonnegative Matrix Factorization）は，非負値からなる行列を分解する手法である [7]．近年，画像，音声，文書，購買データなど幅広い分野で注目されている．非負値行列を 2 つの非負値行列に分解することで，次元削減を行うと同時に，もとの行列が持つ潜在的要素を明確に示すことができる．次元削減を行う手法の一つであり，一つの巨大な行列を複数の行列に効率良く分解する．これによりよりよい推薦を行うことができたり，特徴抽出が可能になる．

M × N のサイズをもつ非負値行列 $X = [x_1,\ldots, x_N] \in R^{M \times N}$ が与えられたとき，NMF は積がもとの行列 X と近くなる二つの非負値行列，$U = [u_{ik}] \in R^{M \times K}$ と $V = [v_{jk}] \in R^{N \times K}$ を推定する．

$$X \approx U \times V^{\mathrm{T}} \tag{1}$$

U × V と X の差が最小となるように推定を行い，ユークリッド距離を利用した損失関数を最小化するように定式化される．式 (2) は距離関数にフロベニウスノルムを利用した場合の損失関数の例である．

$$O = \left\| X - UV^T \right\|^2 = \sum_{ij} \left(x_{ij} - \sum_{k=1}^{K} u_{ik} v_{jk} \right)^2 \tag{2}$$

また，損失関数に正則化項を加える研究も数多くなされている．最も一般的に利用されるのは式 (3) のような L2 正則化項を加える手法である．

$$R = \lambda_u \sum_{i=1}^{M} \|u_i\|^2 + \lambda_v \sum_{j=1}^{N} \|v_j\|^2 \tag{3}$$

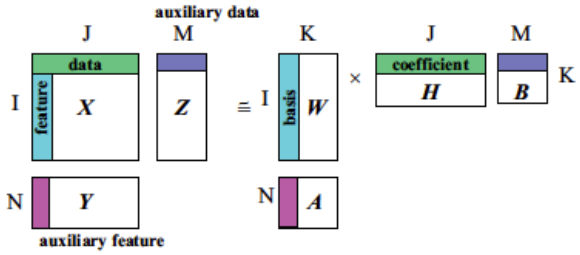

図 3　NMMF のモデル 観測行列 X を基底行列 W と係数行列 H に分解する．その際補助行列 Y と Z を同時に分解することでより性能の高い行列分解を可能とする．

2.3.1　協調型非負値行列因子分解

Singh らは，複数の行列を組み合わせて因子分解を行う高次元のフレームワークである，協調型非負値行列因子分解 (Collective Matrix Factorization) を提案している [15]．二つの行列 X と Y を因子分解するための損失関数 L は式 (6) ように定義される．ただし，$\mathbb{D}$ は Bregman divergences，W は観測行列 X と Y の重み行列，α は [0，1] の重みで X と Y の重要度の比を表す．

$$L_1\left(U, V|W\right) = \mathbb{D}_{F_1}\left(UV^T \parallel X, W\right) + \mathbb{D}_G\left(0 \parallel U\right) + \mathbb{D}_H\left(0 \parallel V\right) \quad (4)$$

$$L_2\left(V, Z|\tilde{W}\right) = \mathbb{D}_{F_1}\left(VZ^T \parallel Y, \tilde{W}\right) + \mathbb{D}_H\left(0 \parallel V\right) + \mathbb{D}_I\left(0 \parallel Z\right) \quad (5)$$

$$L\left(U, V, Z|W, \tilde{W}\right) = \alpha L_1\left(U, V \parallel W\right)\left(1-\alpha\right)\left(V, Z \parallel \tilde{W}\right) \quad (6)$$

Takeuchu らは Collective Matrix Factorization の一種として，インデクスの対応が取れる複数の行列形式データを同時分解する複合非負値行列因子分解 (NM2F: Non-negative Multiple Matrix Factorizaion) を提案している [16]．NM2F は，図 3 のように複数の行列間に共通の因子を仮定することで欠損値の多いスパースな行列を分解する際に欠損値推定の精度が向上する性質を持ち，スパースなデータの行列因子分解でも高い精度がでている．

幸島らは，因子行列の共有と因子行列間の線形制約導入により，属性情報との対応関係を考慮できる非負値多重行列因子分解手法を提案し，調査パネルデータを用いた消費者行動パターン抽出を行っている [19]．ユーザとグループとの対応関係を与える行列 V，商品とカテゴリとの対応関係を与える行列 W と，ユーザ-商品購買行列 X, ユーザ-カテゴリ購買行列 Y，グループ-商品購買行列 Z という 3 つの購買行列を観測行列としたとき，行列 X，Y，Z について，式 (7) のように因子分解を行っている．手法の概要を図 4 に示す．

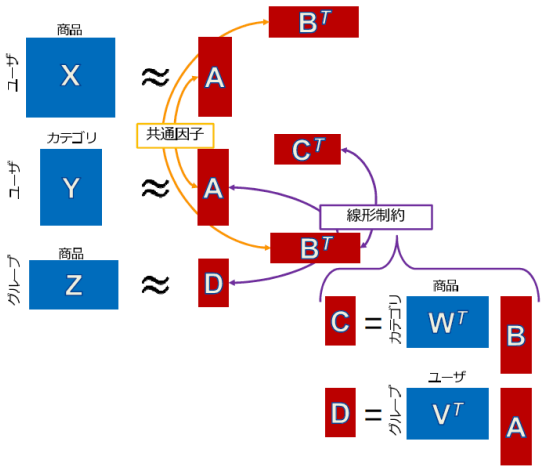

図 4　共通因子や線形制約を考慮することによって，ユーザ-商品購買行列 X, ユーザ-カテゴリ購買行列 Y，グループ-商品購買行列 Z という 3 つの購買行列を同時に因子分解している．

$$\begin{cases} X \approx A \times B^{\mathrm{T}} \\ Y \approx A \times C^{\mathrm{T}} \\ Z \approx B \times D^{\mathrm{T}} \end{cases} \quad (7)$$

式 (7) の行列分解を考えるだけでは，ユーザ，商品というミクロな情報とカテゴリやグループといったマクロな属性情報の関係性が考慮されない．そこで，商品とカテゴリの対応関係を示す行列 W とユーザ-グループの対応関係を示す行列 V を用いて行列間に式 (8) のように線形制約を導入している．

$$\begin{cases} C = W^{\mathrm{T}} \times B \\ D = V^{\mathrm{T}} \times A \end{cases} \quad (8)$$

これにより，解釈が可能な形で複数の行列因子分解を同時にでき，ユーザと商品，カテゴリ，グループの同時クラスタリングを行っているとみなすことができるとしている．

幸島らの手法は数値データの行列因子分解であり，本研究では，数値データと文書データを用いている点で異なる．

2.4　文書データのトピック抽出

文書から潜在的なトピックを獲得するには，確率的潜在意味解析法 (PLSA) [6] やその拡張である潜在的ディリクレ配分法 (LDA) [1] が広く知られており，文書と文書中の単語から，トピックを抽出と個々の文書における各トピックの現れやすさを表す確率を計算することができる．例えばニュースデータの場合，個々のニュースを経済，政治やスポーツなどの潜在的なトピックに分類できる．

文書と単語の非負行列を文書とトピック，トピックと単語に因子分解することによっても，潜在的なトピック抽出や文書クラスタリングを行うことができることが知られており [18] [8]，PLSA と比べても同様の最尤推定結果となることも示されている [5]．

LDA を用いてトピックの抽出ができるが，そのトピックの意味を解釈するのが難しい．本研究では，協調型非負値行列因

子分解を用いて，投資履歴を制約条件として加えることで解釈性を高め，局面を抽出することを試みる．

3. トレーダの特徴分析モデル

本研究では，各トレーダの得意なニューストピックを抽出することを目的とし，ニュースイベント，ニューストピック，取引の関連性を考慮して，協調型非負値行列因子分解を用いた統合分析モデルを提案する．それにより，為替相場でそのときどきの変動の要因となっているニューストピックに適した，トレーダの選択に生かすことができ，人をベースとしたポートフォリオの構築を支援することができる．本研究の提案するモデルは以下の二つの観測に基づいている．

(観測 1) 為替相場は，経済指標の発表，政治的決定や有事など様々なニュースイベントで大きく変動し，トレーダもメディアを通じてニュースイベントを確認しながら取引を行っている (図 2)．特に ForexFactory などの経済ニュースでは，為替に関係する情勢について記述されており，多くのトレーダが経済ニュースイベントを考慮して取引を行っている．

(観測 2) 特定のニューストピックに着目しているトレーダが一定数存在し，実際に自己紹介文で自分が着目しているニュースイベントを明言するトレーダも確認できる．例えば，雇用統計や GDP などの経済指標の発表に着目していると明言するトレーダや，資源の価格に着目していると明言するトレーダを観測している．

モデル中の変数の説明を表 1 で示す．X をユーザの取引結果とニュースイベントを表現する行列，Y をニュースイベントとそのニュースに出現する単語の関係を表現する行列と定義する．うな

$$\begin{cases} X \approx U \times V^{\mathrm{T}} \\ Y \approx V \times W^{\mathrm{T}} \end{cases} \tag{9}$$

本モデルの特徴は，X と Y の二つの観測行列の非負値行列因子分解を行う際に，ニュースの特徴を表現する行列 V を用いて制約をつけることである．U が目的とする行列であり，最適なパラメータを求めた場合に U は局面を考慮したユーザの特徴を表す． 行列 Y を考慮することで，行列 X のみを用いた行列因子分解よりも妥当性の高い行列因子分解を行うことができる．本モデルにおいて，協調型非負値行列因子分解を用いる利点は以下の二点である．

スパースな高次元データへの対応 ニュースイベント数やトレーダ数が多く，かつトレーダの取引数は限られているため，データが非常にスパースである．スパースなデータからでも安定して特徴抽出ができる．

結果の解釈性 共通因子を用いて二つの非負値行列因子分解を同時に行うことで，因子分解後の結果が解釈しやすくなる．具体的には，観測行列 X のみの因子分解ではトレーダのクラスタリングやトレーダが得意とするニューストピックの妥当性を判断することは難しいが，ニューストピックと単語の関係を表す行列 W を考慮することで，その妥当性を判断したり解釈したりすることができる．

表 1 提案モデルの行列と変数の説明

記号	サイズ	内容
X	M × N	ユーザ – ニュースイベントの行列
Y	N × O	ニュースイベント – 単語の行列
U	M × k	ユーザの latent space
V	N × k	ニュースの latent space
W	O × k	単語の latent space
d	1	ニュースイベントの影響期間

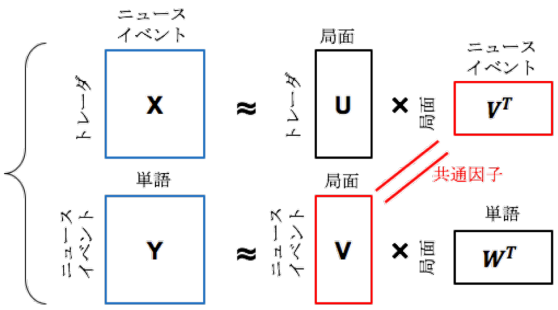

図 5 提案モデルの概要

式 9 に対応する本モデルのイメージは，5 である．

観測行列 X の値は，ユーザ i がニュースイベント j の影響期間において行った取引の結果を正規化した値を表す． 観測行列 Y の値は，ニュース記事 j に含まれる単語 k の回数を正規化した値を表す．そのとき，損失関数 L を以下のように定義し，L が最小となるようなパラメータを求める．L1 と L2 はそれぞれ行列 X，Y の損失関数を表す．

$$L_1 = \left\| X - UV^T \right\|^2 + \lambda_u \sum_{i=1}^{M} \|u_i\|^2 + \lambda_v \sum_{j=1}^{N} \|v_j\|^2 \tag{10}$$

$$L_2 = \lambda_y \left\| Y - VW^T \right\|^2 + \lambda_v \sum_{j=1}^{M} \|v_j\|^2 + \lambda_k \sum_{k=1}^{N} \|w_k\|^2 \tag{11}$$

$$L = L_1 + L_2 \tag{12}$$

提案モデルのグラフィカルモデルを図 6 に示す． 観測行列は X，Y であり黒く表している．X の行列因子分解により U と V が，W の行列因子分解により V と W が求められる．因子分解して導かれる行列 U，V，W には，バイアス項としてそれぞれ λ_u,λ_v,λ_w が影響する．λ_y， λ_u,λ_v,λ_w は 5 節で示すように，パラメータの学習により決定する．

4. モデルの学習

4.1 観測行列の構築

ニュースの日時とタイトルより，ニュースと単語の行列を構築する．その行列を正規化し観測行列 Y とする．

ユーザ i がニュースイベント j の影響期間においてだした取引結果を $profit_{i,\,j}$ とすると，以下のように [0，1] に正規化できる．$x_{i,\,j}$ を要素として持つ行列を観測行列 X として構築する．本研究では，ユーザが得意とする局面の特徴を取得することに着目するため，利益額については考慮せず，正の利益を上

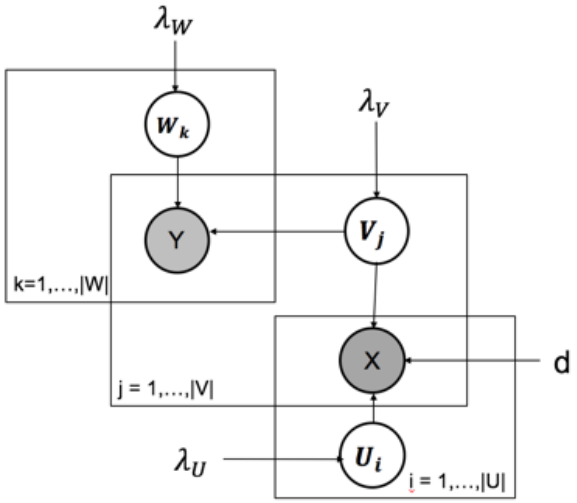

図 6　提案モデルのグラフィカルモデル

げていれば 1，そうでなければ 0 としている．

$$x_{i,\,j} = \begin{cases} 0 & profit_{i,\,j} \leqq 0 \\ 1 & profit_{i,\,j} > 0 \end{cases} \tag{13}$$

ニュース記事 j に含まれる単語 k の回数を $term\_count_{j,\,k}$ とすると，次のように正規化できる．$y_{j,\,k}$ を要素として持つ行列を観測行列 Y として構築できる．

$$y_{j,\,k} = \frac{term\_count_{j,\,k}}{max(term\_count)} \tag{14}$$

この結果，$X \in R^{1214\,\times\,24110}$，$Y \in R^{24110\,\times\,462}$ の行列として構築される．行列 X の非ゼロ要素の割合は，ニュースイベントの影響期間 d を 2 週間とした場合は 7.6%，12 時間とした場合は 3.1%，6 時間とした場合は 1.6% であり，非常に疎な観測行列である．

4.2　逐次更新アルゴリズム

式 (12) を最小化する非負値のを求めるのが目的であるが，非負制約つき非線形最適化問題であり，解析的に解くことはできない．Lee らは，目的関数の上限となる補助関数を反復的に降下させることで目的関数を間接的に降下していく方法を考案し，制約付き非線形最適化問題の解を見通し良く探索することができるとしている [7]．実装上，β-divergence を規準とした逐次更新を行い，収束を効率化させるアルゴリズムを [2] [4] を用いて学習を行う．β-divergence とは，距離尺度の一つであり式 (15) のように表現される [12]．

$$D_\beta\,(y|x) = \frac{y^\beta}{\beta\,(\beta-1)} + \frac{x^\beta}{\beta} + \frac{yx^{\beta-1}}{\beta-1} \tag{15}$$

$D_\beta\,(y|x)$ は y の x に対する疑距離を表し，$\beta = 0$ のときに Itakura-Saito divergence，$\beta = 1$ のときに generalized Kullback-Leibler divergence，$\beta = 2$ のときにユークリッド距離の x，y 間の距離尺度をとる．β-divergence を用いた式 (1) の非負値行列因子分解の目的関数は式 (16) のように表される．

$$E_\beta\,(\theta) = \sum_{i,\,j} D_\beta\left(X_{i,\,j} | \sum_k U_{i,\,k} \cdot V_{k,\,j}\right) \tag{16}$$

表 2　データセットの項目と例

	項目	例
ニュース	タイムスタンプ	2015-05-29 10:00:00
	本文	Revised UoM Inflation Expectations..,
取引履歴	取引の日時	2015-05-29 10:00:00
	通貨	USD
	利益額	0.1

ここで，$X_{i,\,j}$, $U_{i,\,k}$, $V_{k,\,j}$ は行列 X，U，V の要素，θ は，最適化したい全てのパラメータである．この $E_\beta\,(\theta)$ が最小となるように $U_{i,\,k}$, $V_{k,\,j}$ を推定する．β-divergence を用いた目的関数に対する乗算型更新式は p を反復更新のインデックスとして式のように表される [13]．分解後の非負値行列に初期設定し，それぞれの行列がある程度収束するまで式 (17) を繰り返すことにより，効率のよい因子分解を行う．

$$\begin{cases} U_{i,\,k} \leftarrow U_{i,\,k}\left(\dfrac{\sum_j \left(\sum_k U_{i,\,k}V_{k,\,j}\right)^{\beta-2} X_{i,\,j}V_{k,\,j}}{\sum_j \left(\sum_k U_{i,\,k}V_{k,\,j}\right)^{\beta-1} V_{k,\,j}}\right)^{\varphi(\beta)} \\ V_{k,\,j} \leftarrow V_{k,\,j}\left(\dfrac{\sum_i \left(\sum_k V_{i,\,k}U_{k,\,j}\right)^{\beta-2} X_{i,\,j}U_{i,\,k}}{\sum_i \left(\sum_k V_{i,\,k}U_{k,\,j}\right)^{\beta-1} U_{i,\,k}}\right)^{\varphi(\beta)} \end{cases} \tag{17}$$

ただし，$\varphi\,(\beta)$ は以下のように表す．

$$\varphi\,(\beta) = \begin{cases} 1/(2-\beta) & (\beta < 1) \\ 1 & (1 \leqq \beta \leqq 2) \\ 1/(\beta-1) & (2 < \beta) \end{cases} \tag{18}$$

5.　実　　験

データセットとして，ZuluTrade における 2012 年から 2015 年までの 1217 トレーダの取引履歴と ForexFactory による 2007 年から 2015 年までの 37018 件のニュースデータを利用する．データセットの例を表 2 にまとめる．

実験では, 与えられたデータセットをトレーニングセットとテストセットに分割した後，トレーニングセットに対して評価指標が高くなるようにパラメータの学習を行う．

関連研究 [17] のように，評価指標として recall を採用する．recall を採用した理由は,4.3 節でのべたように，観測行列においてゼロの値は取引をしたが利益がでなかった場合もあれば，取引が行われなかった場合もありうる．そのため，適合率を評価することは難しく，非ゼロの要素の推定精度を計測するために再現率を計測する．観測行列 X の非ゼロ要素の一部を欠損させた行列に対して通常の行列因子分解と提案手法で行列因子分解を行う．その結果，欠損させた値を推定し，その recall を測定し評価する．

モデルを実験する際には，パラメータの学習が必要となる．パラメータはバイアス項 λ_u,λ_v,λ_w，ニュースの影響範囲 d，X と Y の重要度の重み α であり，事前のパラメータの学習により，以下のパラメータを設定した．

- 影響範囲 d は，ニュースイベントの前後 12 時間

表 3　評価 (ただし，λ_u,λ_v,$\lambda_w = 0.1, \alpha = 0.5$, ニュースの影響範囲は 6 時間とした場合)

特徴次元 k	提案手法	非負値行列因子分解
10	0.423	0.406
20	0.461	0.456
50	0.491	0.492
100	0.493	0.493

表 4　トピック分類された単語

トピック 1	sales retail core home new italian vehicle german monitor wholesale brc motor pending existing total realized hia cbi manufacturing loans
トピック 2	unemployment rate claims change statement overnight spanish official cash italian monthly german utilization capacity rba minimum bid mpc rbnz boc
トピック 3	bank holiday german official french italian lending votes rate mpc auction bond 10 test results stress prelim wpi consumer spending

- λ_u，λ_v，λ_w は 0.3
- α は 0.5

5.1 評　　価

行列 X に対して通常の行列因子分解を行った場合との提案手法での reall を計測した結果を表 3 に示す．一般的に非負値行列因子分解では，特徴の次元数 k が大きいとより高精度な推定ができ，特徴の次元数 k が小さいとより複雑性の少ないモデルとなるというトレードオフの関係がある．特徴の次元数 k が大きい場合はベースラインと変わらない精度だが，特徴の次元数 k が小さい場合において，良い精度になった．

5.2 ケーススタディ

提案手法により，行列因子分解の解釈性を向上させることができる．生成される単語と特徴 (トピック) の行列 W の定性評価を行い，単語とニュースイベント行列 Y の因子分解の妥当性を確認する．4 に特徴次元が 10 の場合のトピックに分類の一例を示す．トピック 1 は製造業や売上に関するトピック，トピック 2 は雇用に関するトピック，トピック 3 は，銀行や選挙に関するトピックであり，潜在的なトピックが抽出されていることがわかる．上記のようなニュースとトピックの関係と推定されたユーザーとニュースイベントの行列 X を用いることで，特定ユーザーが，どのトピックに強く，どのトピックで弱いもしくは取引を行わないということが判断できる．

6. おわりに

本研究では，ソーシャルトレーディングサービスの取引履歴とニュース記事を用いた，行列因子分解モデルを提案した．具体的には，ユーザの取引結果，ニュースイベント，単語の関係から二つの非負値行列因子分解を制約の下で同時に行うことで，局面を考慮したトレーダの特徴抽出をを可能とするモデルを提案した．

以下の事項を今後の課題と考えている．

- 非負値行列因子分解では，特徴の次元数 k が大きいとより高精度な推定ができ，特徴の次元数 k が小さいとより複雑性の少ないモデルとなるというトレードオフの関係があり，どの次元数が適切かどうかはデータの中身や問題設定を考慮する必要がある．適切な特徴の次元数 k を見つける必要がある．
- トレーダが得意とする局面 (ニューストピック) について，解釈事例を増やして妥当性を検証する．また，ニュースイベントの影響期間を変化させた場合に，どのようにトレーダが得意とする局面の出力結果に影響がでるかを調べる．
- 現在のモデルは，実際の為替の数値を考慮していない．実際の為替の時系列推移もモデルに組み込む．

7. 謝　　辞

本研究の一部は，科研費（課題番号 25700033) による．

文　　献

[1] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, Vol. 3, No. Jan, pp. 993–1022, 2003.

[2] Andrzej Cichocki and PHAN Anh-Huy. Fast local algorithms for large scale nonnegative matrix and tensor factorizations. *IEICE transactions on fundamentals of electronics, communications and computer sciences*, Vol. 92, No. 3, pp. 708–721, 2009.

[3] Abeer El-Korany. Integrated expert recommendation model for online communities. *arXiv preprint arXiv:1311.3394*, 2013.

[4] Cédric Févotte and Jérôme Idier. Algorithms for nonnegative matrix factorization with the β-divergence. *Neural computation*, Vol. 23, No. 9, pp. 2421–2456, 2011.

[5] Eric Gaussier and Cyril Goutte. Relation between plsa and nmf and implications. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 601–602. ACM, 2005.

[6] Thomas Hofmann. Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 50–57. ACM, 1999.

[7] Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, Vol. 401, No. 6755, pp. 788–791, 1999.

[8] Ju-Hong Lee, Sun Park, Chan-Min Ahn, and Daeho Kim. Automatic generic document summarization based on non-negative matrix factorization. *Information Processing & Management*, Vol. 45, No. 1, pp. 20–34, 2009.

[9] Woonyeol Lee and Qiang Ma. Whom to follow on social trading services? a system to support discovering expert traders. In *Digital Information Management (ICDIM), 2015 Tenth International Conference on*, pp. 188–193. IEEE, 2015.

[10] Jingxuan Li, Wei Peng, Tao Li, Tong Sun, Qianmu Li, and Jian Xu. Social network user influence sense-making and dynamics prediction. *Expert Systems with Applications*, Vol. 41, No. 11, pp. 5115–5124, 2014.

[11] Qiang Ma and Masatoshi Yoshikawa. Ranking people based on metadata analysis of search results. In *International Conference on Web Information Systems Engineering*, pp. 48–60. Springer, 2008.

[12] Minami Mihoko and Shinto Eguchi. Robust blind source separation by beta divergence. *Neural computation*, Vol. 14,

No. 8, pp. 1859–1886, 2002.

[13] Masahiro Nakano, Hirokazu Kameoka, Jonathan Le Roux, Yu Kitano, Nobutaka Ono, and Shigeki Sagayama. Convergence-guaranteed multiplicative algorithms for non-negative matrix factorization with β-divergence. In *Machine Learning for Signal Processing (MLSP), 2010 IEEE International Workshop on*, pp. 283–288. IEEE, 2010.

[14] Wei Pan, Yaniv Altshuler, and Alex Pentland. Decoding social influence and the wisdom of the crowd in financial trading network. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*, pp. 203–209. IEEE, 2012.

[15] Ajit P Singh and Geoffrey J Gordon. Relational learning via collective matrix factorization. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 650–658. ACM, 2008.

[16] Koh Takeuchi, Katsuhiko Ishiguro, Akisato Kimura, and Hiroshi Sawada. Non-negative multiple matrix factorization. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, pp. 1713–1720. AAAI Press, 2013.

[17] Hao Wang, Binyi Chen, and Wu-Jun Li. Collaborative topic regression with social regularization for tag recommendation. In *IJCAI*, 2013.

[18] Wei Xu, Xin Liu, and Yihong Gong. Document clustering based on non-negative matrix factorization. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pp. 267–273. ACM, 2003.

[19] 幸島匡宏, 松林達史, 澤田宏. 属性情報を考慮した消費者行動パターン抽出のための非負値多重行列因子分解法. 人工知能学会論文誌, Vol. 30, No. 6, pp. 745–754, 2015.