

# 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専門性分析

## Feature Analysis of traders on Social Trading Services

竹田 創<sup>†</sup>  
Hajime Takeda

Chenyi Zhuang<sup>†</sup>

馬 強<sup>†</sup>  
Qiang Ma

### 1. はじめに

#### 1.1 ソーシャルトレーディングサービス

近年、新しい投資プラットフォームとしてソーシャルトレーディングの popularity が高まっており、ZuluTrade\* や eToro<sup>†</sup> などの数多くのサービスが存在する。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を用いて投資商品を実際に売買することで利益を得る。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群を選択してフォローすることにより、フォローしたトレーダの取引に応じて間接的に投資商品を買って利益を得ることができる。(図1)。ソーシャルトレーディングサービスでは、投資に関する知識がない層にとっても参加しやすく、今後も利用者や取引額が拡大していくことが見込まれる。

似たような投資形態として投資信託がある。投資信託は、投資信託ファンドが投資家から資金を集め、投資信託ファンドが株式や債券などに投資・運用する投資形態である。投資信託ファンドは専門家の集まりでありある程度信頼がおける一方、手数料が高い、最低購入金額が1万円程度のもが多く投資ができないなどのデメリットがある。ソーシャルトレーディングは、より少額の資本から投資ができる一方、誰でもトレーダになることができるためパフォーマンスの低いトレーダも多く、トレーダの取捨選択が難しいというデメリットがある。

ソーシャルトレーディングサービスの利用者は二つに大別でき、本論文ではそれらをトレーダとユーザと呼ぶ。

#### トレーダ

実際に投資商品の売買を行う。フォロワーや成果に応じて手数料をもらうことができる。

#### ユーザ

直接は投資商品の取引を行わず、トレーダをフォローすることで間接的に投資商品の売買を行う。サービス利用者の多くがユーザである。

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

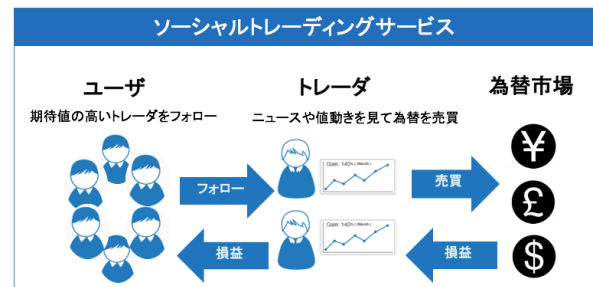


図1: ソーシャルトレーディングサービスの仕組み

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、市場の状況が日々変化の中で、膨大な数のトレーダの情報を整理し適切なトレーダの選択することは難しい。これは掲示板や質問サイトでも、頻繁に議論されておりユーザーにとって悩みの種となっている。サービスもトレーダのランキング機能を提供しているが信頼性が低く十分に活用されているとはいえない。具体的には過去数ヶ月間の成績に基づいて算出されるため、ハイリスクハイリターンなトレーダがランキングの上位になりやすく、ランキングの変動が激しい。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。このサービスでは外国為替証拠金取引 (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。

#### 1.2 研究概要

本研究の目標は、市場に影響のあるニューストピックとそれを得意とするトレーダ群を発見することである。具体的には取引履歴とニュースデータを利用し、トレーダに影響を与えるニューストピックを抽出し、同時に各ニューストピックを得意とするトレーダを発見することである。本研究により、将来起きる可能性のあるニュースイベントやトレンドを予測した場合、得意とするトレーダを発見できる。

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)。また、ソーシャルトレーディングでは自己紹介文で自分

<sup>†</sup> 京都大学, Kyoto University

\*<https://zulutrade.com>

<sup>†</sup><https://www.etoro.com>

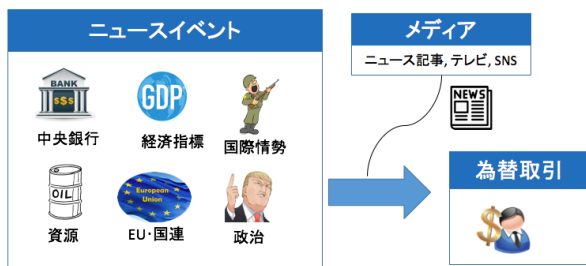


図 2: トレーダは為替市場を変動させるニュースイベントに注目している。

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し、トレーダごとに着目するニューストピックは異なっている。そこで本研究で提案するモデルは、ニュースイベントが為替市場や為替トレーダの取引行動に影響を与えるという仮定に基づく。提案手法では、3つの観測行列を構築し、協調型非負値行列因子分解 (Collective NMF[13]) を利用することで、トレーダ・取引・ニュースの関係性を明らかにすることでトレーダの特徴理解を支援する。

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モデルを提案した。これにより、市場に影響のあるニューストピックとそれを得意とするトレーダ群を発見することができる。
- 取引履歴を用いることにより、ニュースイベントが為替取引に与える影響の時間的関係性を明らかにした。ニュースイベントと金融市場の関連性を前提とした既存研究は多く存在するが、ニュースイベントが実際に金融市場に与える影響の大きさやその影響期間を定量的に評価した論文は少ない。

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サービス

ソーシャルトレーディングサービスでは、ユーザはパフォーマンスの高いトレーダ群を選択してフォローすることにより、フォローしたトレーダの取引に応じて間接的に投資商品を売買し利益を得ることができる。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、自分で取引を行うユーザーに比べて他のトレーダをコピーするトレーダは高い Return On Investment (ROI) を得ることを示している [12]。

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 [8]。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市場トレンド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、現在の市場トレンド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コミュニティ理解には、オピニオンリーダー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、多くの方法が提案されている ([3], [9], [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 [9] 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 ニュースイベントと為替市場の関係

Luiら [10] によると、為替取引の 85% 以上は、ファンダメンタル分析とテクニカル分析両方を考慮して行われ

ている。為替市場は国や経済の状況をリアルタイムに反映しており、例えば以下のようなニュースイベントは市場を大きく動かす。

- 各国の政府政策や中央銀行の金融政策等の発表
- GDP や雇用統計などの経済指標の発表
- テロなどの有事
- 原油などの資源価格

Bauwens ら [1] は、ニュースイベントがユーロ/アメリカドルの為替市場に与える影響について調査し、為替相場はニュース発生後だけでなくニュースの発生前においても影響を受けていると結論づけた。しかし、影響を与えやすいニュースを著者の判断のもと手動で選択しており、どのようなニュースが影響を与えにくいことやニュースが為替相場に与える影響の期間については議論されていない。本論文では、ニュースイベントが実際に為替市場に与える影響の大きさやその影響期間を評価する。

### 3. 関係データ学習

#### 3.1 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 [6]。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。

$M \times N$  のサイズをもつ非負値行列  $X = [x_1, \dots, x_N] \in R^{M \times N}$  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。

$$X \approx U \times V^T \quad (1)$$

$U \times V$  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。

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

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。

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

#### 3.1.1 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る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 [13]。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、 $\alpha$  は  $[0, 1]$  の重みで  $X$  と  $Y$  の重要度の比を表す。

$$\begin{aligned} L_1(U, V|W) &= \mathbb{D}_{F_1}(UV^T \| X, W) + \mathbb{D}_G(0 \| U) + \mathbb{D}_H(0 \| V) \end{aligned} \quad (4)$$

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

$$\begin{aligned} L(U, V, Z|W, \tilde{W}) &= \alpha L_1(U, V \| W) + (1 - \alpha) (V, Z \| \tilde{W}) \end{aligned} \quad (6)$$

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 [16]。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 3 に示す。

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

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

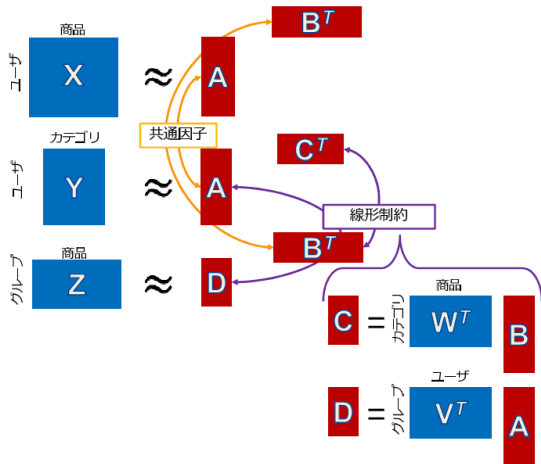


図 3: 共通因子や線形制約を考慮することによって、ユーザ-商品購買行列  $X$ , ユーザ-カテゴリ購買行列  $Y$ , グループ-商品購買行列  $Z$  という 3 つの購買行列を同時に因子分解している。

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。

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

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のみならずとみなすことができるとしている。

幸島らの手法と本研究の違いは二点ある。一点目、本研究では取引履歴を用いており、観測行列を構築する際に時系列性を考慮している。二点目、本研究では数値データに加えて文書データも用いている。

### 3.2 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) [5] やその拡張である潜在的ディリクレ配分法 (LDA) [2] 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 [15][7], 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 [4]。

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

表 1: データセットの項目と例

項目	例
ニュース記事	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 EUR/USD 利益額 0.1

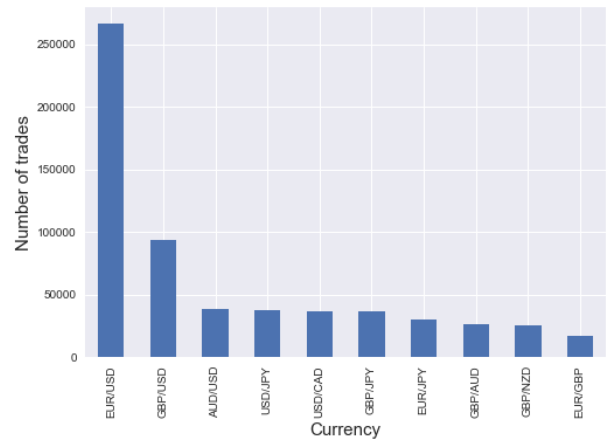


図 4: 取引履歴データの通貨ペアごとの取引高 (上位 10 通貨ペア)

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、ニューストピックの抽出に加えて、各ニューストピックを得意とするトレーダを発見することを試みる。

## 4. トレーダの特徴分析モデル

### 4.1 利用する取引履歴データと経済ニュースデータ

実際の取引履歴と経済ニュースデータを用いてモデルのアイデアについて説明する。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 件の経済ニュースデータを利用する。データセットの例を表 1 にまとめる。

738701 件の取引履歴のうち、約 36% にあたる 265942 件がユーロ/アメリカドルの通貨ペアであった。(図 4) Bauwens らの研究 [1] においても、ニュースイベントがユーロ/アメリカドルの為替価格に影響を与えていると結論づけている。そこで本研究ではユーロ/アメリカドルの通貨ペアのみを利用する。図 5 は、取引履歴データのユーロ/アメリカドルの取引高の取引時間を 1 時間ごとに示した図である。ZuluTrade で取引されるユーロ/



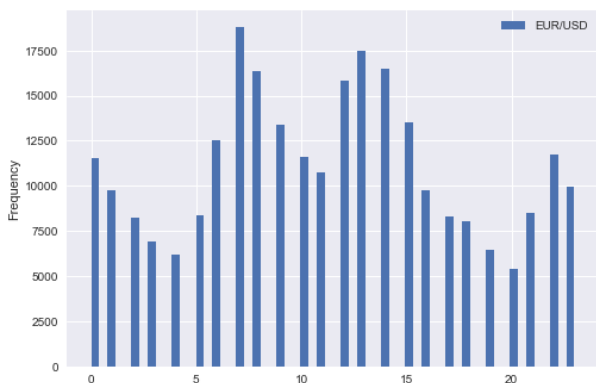


図 5: 取引履歴データの 1 時間ごとのユーロ/アメリカドルの取引高

アメリカドル通貨ペアの取引高は、取引高の増減はあるものの一日を通して取引されていることがわかる。

#### 4.2 提案モデルのアイデア

本研究の目的は、市場に影響のあるニューストピックとそれを得意とするトレーダ群を発見することである。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、トレーダもニュースイベントを考慮して取引を行っている。(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、トレーダによって着目するニューストピックは異なっている。実際に自己紹介文で自分が着目するニュースイベントを明言するトレーダも観測できる。

実際に取引履歴と経済ニュースデータを用いて、ニュースイベントが市場のトレーダの投資行動に与える影響について説明する。為替市場の局面は刻々と変化し、局面に応じてパフォーマンスが高いトレーダも変化している。2015 年 3 月の一ヶ月間において、アメリカ及びヨーロッパ関連のニュースが 40 件発行された。図 6 の青色部分は、ニュース発行時の時刻を 0 とし、その前後 30 分のユーロ/アメリカドルの為替取引回数を表す。ニュース発生直後から取引回数が増え、20 分経過したあたりで取引回数が収まっていることがわかる。本研究では、図 6 の赤線のような、ニュースが取引に与える影響の大きさの時系列推移を考慮する。影響度の形状を変化させてモデルの作成評価を繰り返すことで、尤もらしい影響の推移を発見する。

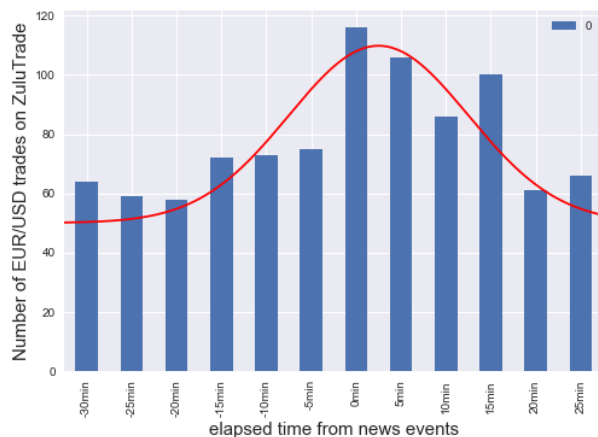


図 6: ニュース発生時の時刻を 0 とした場合のユーロ/アメリカドルの為替取引回数 (2015 年 3 月)

#### 4.3 トレーダの特徴分析モデル

トレーダと取引の関係を表現する行列を  $X$ 、ニュース記事とニュースイベントの関係を表現する行列を  $Y$ 、ニュース記事と単語の関係を表現する行列を  $Z$  とし、取引履歴とニュース記事からこれらの行列を構築する。構築の詳細は章 5.1 で示す。  $x_{ab}$  は取引  $b$  の実行者がトレーダ  $a$  であるかどうか、  $y_{bc}$  はのニュース記事  $c$  が取引  $b$  に与えた影響の強さ、  $z_{cd}$  はニュース記事  $c$  に単語  $d$  が含まれるかを表す。提案モデルは式 9 で表される非負値行列因子分解モデルであり、イメージは図 7 である。

$$\begin{cases} X \approx A \times B^T \\ Y \approx B \times C^T \\ Z \approx C \times D^T \end{cases} \quad (9)$$

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  $B, C$  を用いて制約をつけることである。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 取引数・ニュース記事数・トレーダ数が多く、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。

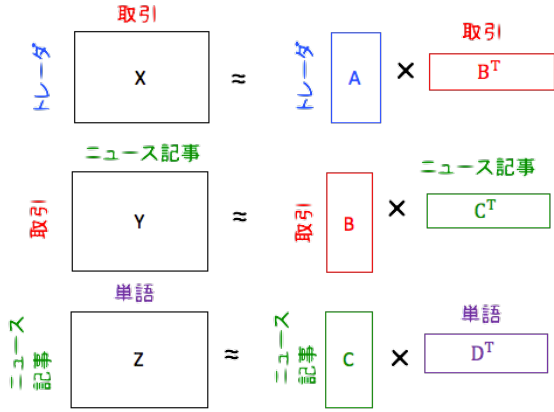


図 7: 提案モデルの概要

損失関数  $L$  をフロベニアスノルムを用いて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。 $L_1, L_2, L_3$  はそれぞれ行列  $X, Y, Z$  の行列分解時の損失関数を表す。

$$\begin{cases} L_1 = \|X - AB^T\|^2 + \lambda \sum \|a_i\|^2 + \lambda \sum \|b_j\|^2 \\ L_2 = \|Y - BC^T\|^2 + \lambda \sum \|b_j\|^2 + \lambda \sum \|c_k\|^2 \\ L_3 = \|Z - CD^T\|^2 + \lambda \sum \|c_k\|^2 + \lambda \sum \|d_k\|^2 \\ L = L_1 + L_2 + L_3 \end{cases} \quad (10)$$

## 5. モデルの学習

### 5.1 観測行列の構築

トレーダ  $a$  と取引  $b$  の関係  $x_{a,b}$  は式 11 のように定義し行列  $X$  を構築する。

$$x_{a,b} = \begin{cases} 1 & \text{if トレーダ } a \text{ が取引 } b \text{ の実行者である} \\ 0 & \text{if トレーダ } a \text{ が取引 } b \text{ の実行者でない} \end{cases} \quad (11)$$

取引  $b$  とニュース記事  $c$  の関係  $y_{b,c}$  はニュース記事  $c$  が取引  $b$  に与えた影響の強さを表し、式 12 のように定義し行列  $Y$  を構築する。

$$y_{b,c} = \begin{cases} f(\text{delta}(b, c)) & \text{if 取引 } b \text{ の利益が正} \\ 0 & \text{if 取引 } b \text{ の利益が負} \end{cases} \quad (12)$$

ただし、 $\text{delta}(b, c)$  は取引  $b$  の時刻とニュース記事  $c$  の時刻の差を表し、 $f(x)$  はニュースの影響度の関数を表し、 $f_1(x)$  や  $f_2(x)$  を用いる。

- $f_1(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$  (図 8)
- $f_2(x) = \frac{1}{\omega\pi} e^{-\frac{(x-\xi)^2}{2\omega^2}} \int_{-\infty}^{\alpha\left(\frac{x-\xi}{\omega}\right)} e^{-\frac{t^2}{2}} dt$  (図 9)

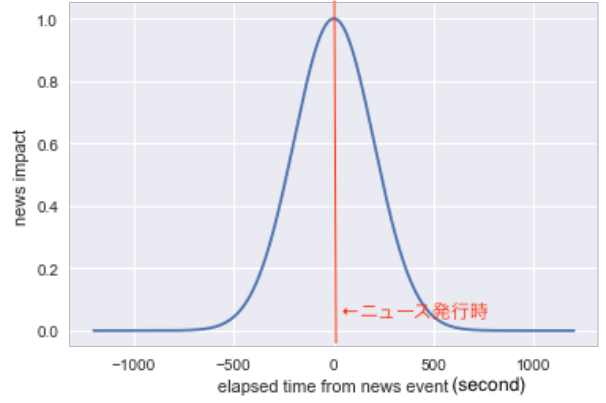


図 8: 正規分布を用いた場合のニュースの影響度の関数

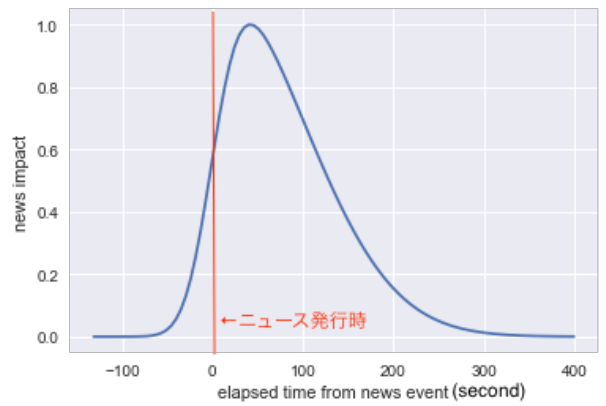


図 9: 歪正規分布を用いた場合のニュースの影響度の関数

ニュース記事  $c$  と単語  $d$  の関係  $z_{c,d}$  は式 13 のように定義し行列  $Z$  を構築する。ニュース記事  $c$  に含まれる単語  $d$  の回数を  $\text{term\_count}(c, d)$  とすると、

$$z_{c,d} = \frac{\text{term\_count}(j, k)}{\max(\text{term\_count})} \quad (13)$$

2015年3月の、アメリカ及びヨーロッパ関連のニュースとユーロ/アメリカドルの為替取引データを利用して行列を構築したところ、行列のサイズは  $X \in R^{1188 \times 27358}$ ,  $Y \in R^{27358 \times 40}$ ,  $Z \in R^{40 \times 462}$  となった。

### 5.2 実験

5.1 節で述べた方法で観測行列の構築し、実験を行う。 $f(x) = f_1(x)$  の場合は、観測行列  $Y$  を構築する際の  $\mu, \sigma$ , 損失関数のバイアス項の重み  $\lambda$  がモデルのパラメータである。 $f(x) = f_2(x)$  の場合は、観測行列  $Y$  を構築する際の  $\xi, \omega, \alpha$ , 損失関数のバイアス項の重み  $\lambda$  がモデルのパラメータである。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 [14] のように、評価指標として RMSE を採用する。具体的には、観測

表 2: RMSE

特徴次元 k	$f_1(x)$ の場合	$f_2(x)$ の場合
10	0.70	0.64
20	0.54	0.53
30	0.48	0.41

表 3: トピック分類された単語

消費・物価指数関連 のトピック	Core, Sales, Retail, m/m, CPI(消費者物価指数), Goods, Orders, Durable
製造業に関するトピ ック	Manufacturing, PMI(購買担当者 景気指数), ISM(製造業景況感指 数), Flash, German, French
雇用に関するトピッ ク	Unemployment, Claims, Rate, Change

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い、推定した値と本来の値の RMSE(root-mean-square error) を測定する。なぜなら、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、適合率を評価することは難しいためである。学習の結果、パラメータは以下のように定まった。

- $f(x) = f_1(x)$  の場合  $\mu: 0, \sigma: 200, \lambda: 0.1$
- $f(x) = f_2(x)$  の場合  $\xi: 0, \omega: 1000, \alpha: 4, \lambda: 0.1$

### 5.3 評価

行列  $X$  に対して  $f(x) = f_1(x), f_2(x)$  それぞれの影響度関数を用いてモデルを構築し、値の欠損・再推定を行い RMSE を計測した。その結果を表 2 に示す。正規分布を用いた影響度関数  $f_1(x)$  の場合より歪正規分布を用いた影響度関数  $f_2(x)$  の方が良い精度になっており、あてはまりがよいことがわかる。

### 5.4 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。3 に特徴次元が 10 の場合のトピックに分類の一例を示す。消費・物価指数関連のトピック、製造業に関するトピック、雇用に関するトピックなどニュースイベント潜在的なトピックが抽出されていることがわかる。

## 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、協調型非負値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共通因子の制約の下で同時に行うことで、トレードに影響を与えるニューストピックの抽出や各ニューストピックを得意とするトレーダの発見を可能とするモデルを提案した。今後も引き続き、モデルの改良に取り組む。以下の事項を優先すべき今後の課題と考えている。

- ニュースが取引に与える影響の時系列推移の形状は、今回は正規分布を利用した。歪度を導入したりフォンミーゼス・フィッシャー分布を利用することを検討している。
- 今回はユーロ/アメリカドルの通貨ペアに限定して実験を行った。通貨ペアを増やして通貨間関係や通貨ごとの特徴を分析する。

## 参考文献

- [1] Luc Bauwens, Walid Ben Omrane, and Pierre Giot. News announcements, market activity and volatility in the euro/dollar foreign exchange market. *Journal of International Money and Finance*, Vol. 24, No. 7, pp. 1108–1125, 2005.
- [2] 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.
- [3] Abeer El-Korany. Integrated expert recommendation model for online communities. *arXiv preprint arXiv:1311.3394*, 2013.
- [4] 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.
- [5] 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.
- [6] Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factor-

- ization. *Nature*, Vol. 401, No. 6755, pp. 788–791, 1999.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] Yu-Hon Lui and David Mole. The use of fundamental and technical analyses by foreign exchange dealers: Hong kong evidence. *Journal of International Money and Finance*, Vol. 17, No. 3, pp. 535–545, 1998.
- [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] 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 Conference on Social Computing (SocialCom)*, pp. 203–209. IEEE, 2012.
- [13] 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.
- [14] Hao Wang, Binyi Chen, and Wu-Jun Li. Collaborative topic regression with social regularization for tag recommendation. In *IJCAI*, 2013.
- [15] 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 information retrieval*, pp. 267–273. ACM, 2003.
- [16] 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.