

# 感情によるマルチモーダル AI を利用した IPO 株価推定

## Prediction of IPO stock prices by using multimodal emotion AI

河合 継<sup>1</sup>, 新田 翔<sup>1,2</sup>, 大川 堯郁<sup>1,3</sup>, 西山 昇<sup>4</sup>

Kei Kawai<sup>1</sup>, Sho Nitta<sup>1,2</sup>, Takafumi Okawa<sup>3</sup>, and Noboru Nishiyama<sup>4</sup>

<sup>1</sup> クリスタルメソッド株式会社

<sup>1</sup> Crystal method co.ltd<sup>[1]</sup><sub>SEP</sub>

<sup>2</sup> 東京理科大学工学研究科経営工学専攻

<sup>2</sup> Department of Management Science, Graduate School of Engineering,  
Tokyo University of Science

<sup>3</sup> 東京大学工学系研究科物理工学専攻

<sup>3</sup> Department of Applied Physics, Graduate School of Engineering, The University of Tokyo

<sup>4</sup> 千葉商科大学会計大学院 客員教授

<sup>4</sup> MBA Program, Graduate School of Accounting & Finance, Chiba University of Commerce

**Abstract:** IPO 及び New Stage (市場変更) 時に放映される StockVoice TV 出演社の社長や他出演者の表情・声の特徴量・テキストを基にした感情特徴量が株価に与える影響について、機械学習を通じた検証を行った。価格データについては、株式会社 Quick 様から提供して頂いたデータを利用し、検証対象として番組全体の感情特徴量と翌日の終値、及び番組中の一分ごとの感情特徴量と次点の一分足の終値の学習を行った。本研究では、目・耳・言葉の三系統のマルチモーダルな特徴抽出を行う事で予測の精度を上げることが出来るのではないか、という仮説のもと検証を行った。日足に関しては、SVM・ロジスティック回帰に特徴量を入力し二値検証を行い、1分足は5秒ごとの特徴量を入力とし、XGBoost 及び DNN を用いた3クラス分類による検証を行った。

## 1. はじめに

古典的な経済学において、人間の売買行動は需要と供給によって決定される価格に基づいて行われるものだと考えられている。しかしながら、近年の様々な研究によると、人間は感情によって売買行動を起こすという研究結果も報告されている。実社会においては、店員が顧客と会話を進めている際、「顧客が頻繁にうなづくような演出」をすることにより、購買意欲が上昇し、一方で首を振るような演出をすると購買意欲が低下するという報告がされている。これは、傾きによって人間は無意識に良い感情を抱くためであるとされている。同時に、プレゼンテーションの良し悪しも購買意欲や売上に大きな影響を及ぼすと考えられ、人々は何かものを買う際に論理的に買う理由を考える一方で、買うタイミングにおいては何かしらの感情の動きがあり、それに基づいて売買を行うことが多いとされている。

同様に、IPO 及び市場変更の際に企業の社長

や従業員が出演し会社の事業計画をスピーチする映像においても、人々が受ける印象や感情によって売買行動への影響があるのではないかと考え、検証を行った。映像による企業説明においても、一般的な購買行動と同様にセールスが店頭において行う営業活動と共通する部分があるのではないかと考えた。

このような IPO・市場変更に伴うプレゼンテーション映像は、株式会社 StockVoice 様によって上場・公開日当日の後場開始時に放映されたものを用い、番組では主に自社の事業計画について説明が行われる。(表1参照)

本映像は、キャストと二人で話をするものや、プレゼンテーションの資料を基に事業内容の説明を行う13分程度のインターネット・CS 放映である。これらの映像に基づく感情の特徴量が株価の値動きにどのような影響を与えるのか、また影響があるのかということについて、機械学習を用い、検証を試みた。

2017年			コード	
	統合	AOI TYO Holdings (株)	3975	★①
	統合	(株)FCホールディングス	6542	★JQ
127	1030	(株)シャノン	3976	★M
206	1430	(株)CDG※質問順番変更	2487	②→①
207	1330	森トラスト・ホテルリート投資法人	3478	★REIT
210	1030	(株)安江工務店(やすえこうむてん)	1439	★JQ
216	1430	(株)日宣(にっせん)	6543	★JQ
217	1330	(株)TOKYO BASE	3415	M→①
223	1330	(株)レノバ	9519	★M
223	1430	ユナイテッド&コレクティブ(株)	3557	★M
227	:	(株)ハブ	3030	JQ→②
302	1330	イフジ産業(株)	2924	②→①
303	1030	(株)やまぜんホームズ	1440	★PRO
306	:	(株)エイチワン	5989	②→①
307	1030	(株)ロコンド	3558	★M

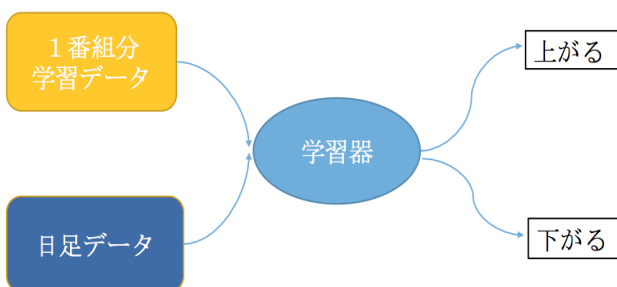
表 1. IPO・市場変更に伴う放映一覧の例

## 2. 前提・検証概要

今回の研究では、人間の感情特徴量をマルチモーダルに抽出し、機械学習を用いて行った。マルチモーダルに特徴量を抽出する理由としては、例えば、「さようなら」という単語の意味を扱う際、まず一次的には単語の意味と話し方、そして二次的には文脈によってその意味合いが大きく左右されると考えられ、どのような意味合いで「さようなら」と言っているのかをより精度良く判断したいと考えたからである。今回は、目・耳・言葉（テキスト）の三つのモーダルチャネルを用いることによって、感情のそれぞれにおける特徴量を抽出して学習を行った。

検証対象は、

- ・番組中全体の評価を受けて翌営業日の値動き（日足）
- ・番組中の値動き（一分足）を用い、一分足は歩み値から生成を行った。
- ・日足



- ・一分足

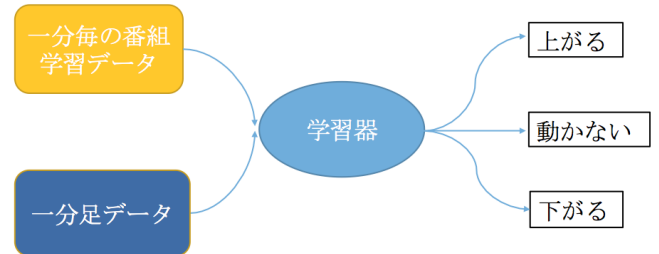


図 1. 日足と分足の二値・三値予測

### 2.1 目の特徴量（映像の顔表情）

視覚による感情特徴量抽出では、StockVoiceTVの5秒ごとの顔表情特徴量を取得した。

特徴量の抽出は、MicrosoftのEmotion APIを利用した。この際の特徴量のデータは、以下のようなベクトル量である。

```
emotion_indices=['disgust','happiness','surprise','anger','fear','neutral','sadness','contempt']
```

複数人が画面上に表れている場合は、全体の平均、プレゼンテーションの場合、画面上に顔が現れないため特徴量は neutral を表す要素のみが 1 となる One-Hot ベクトルとなる。日足、分足については同様のデータを利用した。

### 2.2 耳の特徴量（感情特徴量）

音声処理によって感情認識を行う場合、音声のパワー、周波数、MFCC が大きな影響を及ぼすとされる。これらの特徴は感情特徴量と呼ばれる。[1]

今回は、200次元のMFCCとパワーの特徴量変数を学習に利用し、OpenSmileを用いて5秒ごとに情報抽出を行った。

### 2.3 言葉の特徴量（テキスト）

言葉による特徴量抽出では、StockVoiceTVを一分ごとに区切り、音声ファイルを作成した後、その音声ファイルをGoogle Speech APIに送信し発話内容のテキストを受け取った。そして、Google Natural Language APIによって感情極性を取得した。

また、日足検証用のテキストに関しては、Google

ストレージにマップした後、Google Speech APIによって日本語化を行った。一分足のデータとの相違点としては、途中で文が切れること無く、文が区切られて表現されている。特徴量は、-1.0から+1.0の数値によって表現される。

また、東北大学で収録した感情音声コーパス (JTES) についても同様にテキスト起こしを行い、Google Natural Language API に渡した。この際の評価値は、以下の通りであり、感情について Google Natural Language API で判定ができると考えている。

- ang (怒り) : -0.312
- joy (喜び) : 0.516
- sad (悲しみ) : -0.114
- neu (中立) : 0.198

### 3. 入力データとモデル

#### 3.1 価格

株式価格の入力データとしては、日足分 2015 年 1 月から 2017 年 12 月の分について株式会社 Quick 様より提供して頂いたものを用いた。一分足データについては、2017 年の歩み値を同様に提供していただき、2015、2016 年分については、JPX データクラウドのサービスからダウンロードしたものを PostgreSQL にロードし、今回検証分のみ抽出した。

またデータの欠損等により、最終的に分類器の作成に用いた企業数は 201 社となった。

#### 3.2 分類器について

日足の分類器作成には、SVM とロジスティック回帰を用いた。これらは、多クラス分類で一般的に用いられているアルゴリズムである。

1 分足の分類器作成には、XGBoost および DNN を使用した。XGBoost は高速かつ予測精度が高いアルゴリズムとして注目されており、機械学習のコンペティションでよく使用されている。DNN は昨今注目されている Deep Learning の手法の 1 つである。

### 4. 検証方法・結果

#### 4.1 番組全体の値動きについて

日足データに関して、IPO (新規株式公開) のみの場合 (91 社) と New Stage (上場変更) を含む場合 (196 社) に分けて検証を行った。SVM、ロジスティック回帰それぞれにおいて分類器を作成した。

使用した特徴量は顔データ・音声・テキストの特徴量を会社ごとに一つのファイルにまとめ、当日の

株価と翌営業日の株価の差がプラスであれば0、マイナス及び動きがない場合は1のクラスを設定して学習を行った。

7 : 3	LOGISTIC	SVM
1	0.6923	0.5128
2	0.4872	0.3333
3	0.4615	0.5128
4	0.5385	0.5128
5	0.4872	0.5385
6	0.4103	0.4359
7	0.4615	0.4872
8	0.5128	0.4872
9	0.5385	0.5641
10	0.5385	0.5128
Max	0.6923	0.5641
Min	0.4103	0.3333
Avg	0.5128	0.4897

表★. 新規上場企業のみ (Train : Validation = 7 : 3)

8 : 2	LOGISTIC	SVM
1	0.7692	0.5385
2	0.5385	0.6923
3	0.6538	0.3846
4	0.5385	0.5385
5	0.5769	0.5000
6	0.4615	0.5000
7	0.3846	0.3846
8	0.3846	0.5385
9	0.4615	0.6154
10	0.5385	0.5000
Max	0.7692	0.6923
Min	0.3846	0.3846
Avg	0.5308	0.5192

表★. 新規上場企業のみ (Train : Validation = 8 : 2)

7 : 3	LOGISTIC	SVM
1	0.4576	0.3390
2	0.4746	0.5593
3	0.4915	0.5254
4	0.5085	0.5424
5	0.4915	0.4407
6	0.4576	0.4407
7	0.4407	0.4746
8	0.5254	0.5085
9	0.4915	0.5085
10	0.5424	0.5424
Max	0.5424	0.5593
Min	0.4407	0.3390
Avg	0.4881	0.4881

表★. すべての企業 (Train : Validation = 7 : 3)

8:2	LOGISTIC	SVM
1	0.5128	0.3590
2	0.4872	0.5385
3	0.5897	0.6154
4	0.4615	0.5641
5	0.4359	0.4359
6	0.3846	0.3590
7	0.4872	0.3846
8	0.5641	0.4872
9	0.4872	0.4872
10	0.4359	0.6667
Max	0.5897	0.6667
Min	0.3846	0.3590
Avg	0.4846	0.4897

表★. すべての企業 (Train : Validation = 8 : 2)

## 4.2 リアルタイム中の値動きについて

IPO の開始時点で公募価格は決定しており、リアルタイムの値動きに変動があった場合にのみ配信される歩み値を基にデータを構築しているため、番組開始時点で価格配信がされていない場合もありえる。201 社中 91 社が番組開始時点で価格配信を受けていたため、まずその 91 社について検証を行った。(約定がない等、リアルタイムでのデータがとれないことが、公募価格を取り入れたことで、約定がない時間帯に、価格が上昇しているのか、下降しているのかがわかるようになった。) また、公募価格についても株式会社 Quick 様のご協力のもとデータを受け取り、価格配信がなされていない 40 社程度に

についても検証を行えるようにした。(合計 138 社)

会社ごとに 5 秒ごとの顔の表情・声・テキストの特徴量を入力としてデータを構築した。

その後 DNN を使い、分類器を作成した。ここで、3 クラス分類のラベルは以下のように設定した。

0 : 0.01% 以内の値動き or 約定なし

1 : 0.01% 以上の上昇

2 : 0.01% 以上の下落

検証の結果は以下に示す通りである。

## 4.3 91 社のデータを用いた学習・分析

### 4.3.1 ランダムフォレスト

次に、学習が高速であることや説明変数が多数でも上手く働くという理由から、ランダムフォレストを用いて分類器を構築した。学習した分類器のテストデータでの正答率を以下に示す。

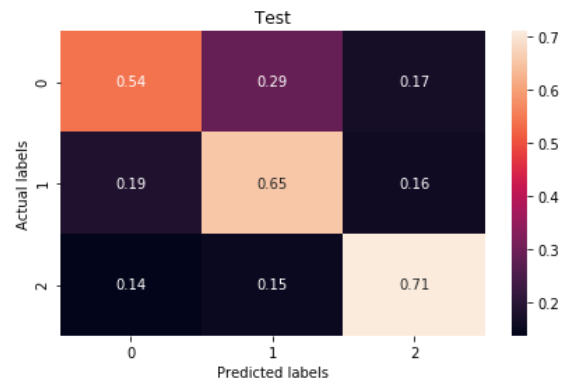


図 3. ランダムフォレストによる予測

### 4.3.2 XGBoost

さらに、予測精度の向上のため、XGBoost を用いて分類器の構築を行った。データはランダムフォレストと同じものを用いた。学習した分類器のテストデータでの正答率を以下に示す。結果は、ランダムフォレストと比較して予測精度が向上した。

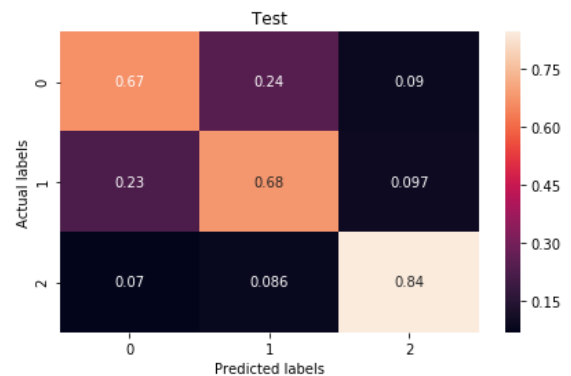


図 4. XGBoost による予測

また、XGBoost では、説明変数の重要度（寄与度）の算出を行った。結果を以下に示す。

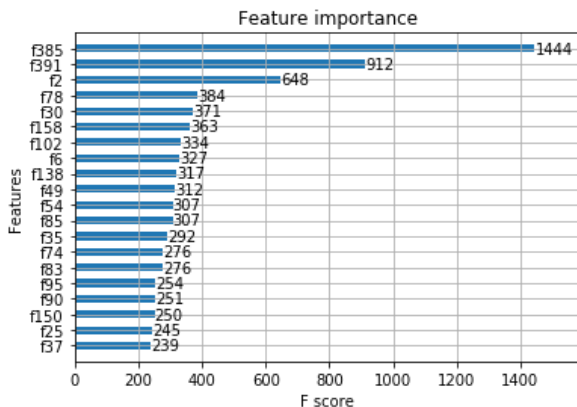


図 5. 説明変数の算出

この分析において、最も重要度が高い「f385」は、テキストの感情極性を示す。また、二番目の「f391」は、中間を表す顔表情感情特徴量である。一方で、テキストの感情極性は一分間同じ値を保持しているため、テキストの感情極性を除いて、再度勾配ブースティング決定木を用いた予測器の構築と説明変数の重要度（寄与度）の算出を行った。テストデータでの正答率と重要度算出の結果を以下に示す。

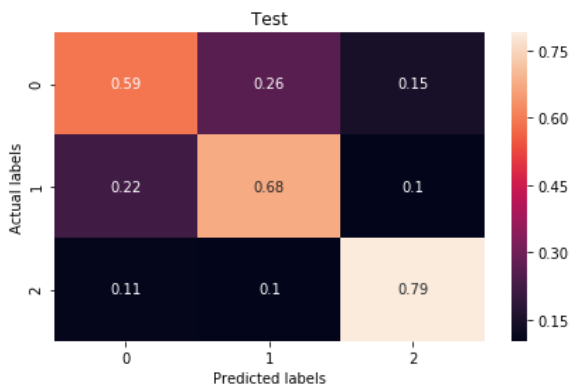


図 6. テストデータの正答率と説明変数の重要度

最も重要度の高い「f390」は中間を示す顔表情特徴量であり、二番目の「f2」は音量の二乗平均平方根値のデータ中の最小値である。三番目の「f30」は、二次メル周波数ケプストラム係数の算術平均を示す。「中間」以外の顔表情特徴量で高い重要度を示したものは無かった。

#### 4.3.3 DNN

91 社の学習データ、および、138 社の学習データを用い、ニューラルネットワークワークの分類器構築を行った。今回の検証ではハイパーパラメータの設定が難しく、試行錯誤を行った。無作為にデータ増幅した場合学習率は伸びたが、結果が伴っていない（オーバーフィッティング）。3 月 20 日の研究発表までに解決する見通し。

#### 4.3.4 91 社のデータを用いた学習・分析

保有するデータのうち、Stock Voice TV の開始時刻から 10 分以内に約定を一度でもしている企業の数が 91 社であった。これらの企業のデータを用い、DNN による分類器を作成した。結果は以下に示す通りである。

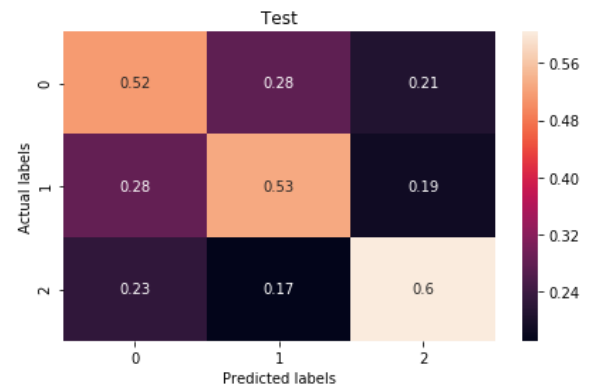
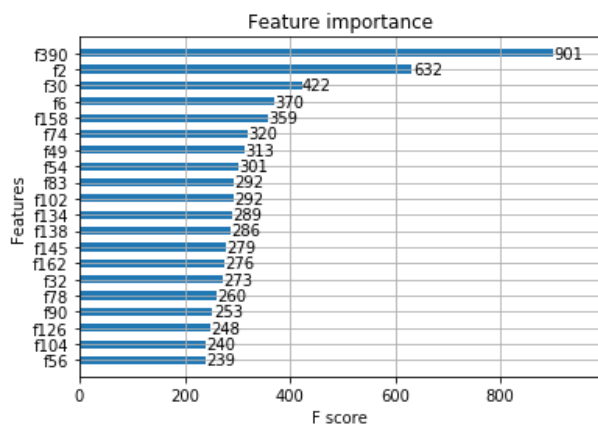
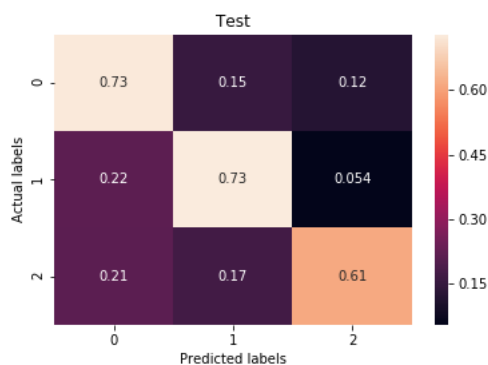
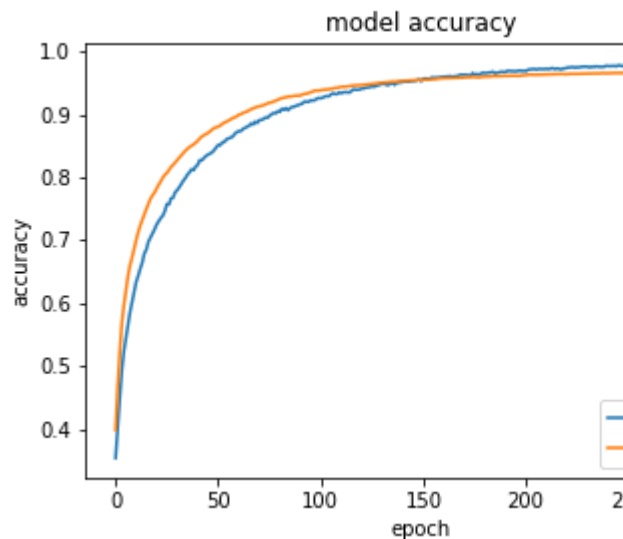
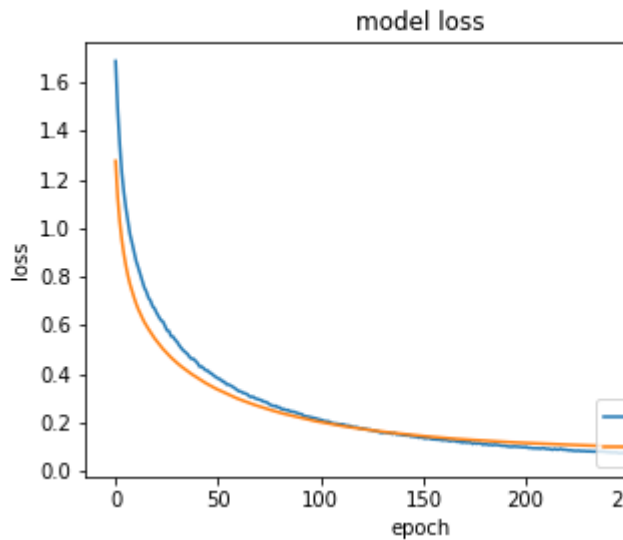


図 7. ニューラルネットワークを用いた正答率



#### 4.3.5 138 社のデータを用いた学習・分析

138 社のデータを使用した場合、データ量が少ないため、学習に使用するデータでは、異なる時点の特徴量ベクトルを無作為に入れ替え、同じラベルをつけた。これらのデータにより学習した分類器を用い、無作為に入れ替えていないデータのクラス分類を行った。結果は以下に示す通りである。データ量をランダムに 100 倍増幅をした結果正答率は 67.7%となった。



・3月20日の本会で発表いたします。

## 5. 課題と本技術の応用先について

### 5.1 発話区間・コンテキストについて

何らかの映像作品において、今回のように人が話をしているような場合、発話区間という概念があり、コンテキストとしては、発話区間内で整合性が取れているが、一分ごとという形を取ったことにより、このような発話区間の概念が失われている点に課題があると考えられる。

また、現在では、音声合成ソフトで利用する形で実装し、実用化を目指し開発を行っている。

### 5.2 音声の特徴量について

今回学習したスペクトログラムによる DNN の感情認識エンジンや、東北大学と共同で行っているエンジンも約 50%程度の精度となっており、今回は感情特徴量と呼ばれる値を取り出したが、今後はエンジンの精度をより上げる方向で研究を進めたいと考えている。

### 5.3 テキスト部分について

テキスト学習については、より精度を上げるため、決算短信や、有価証券報告書などを学習したものをエンジンとして価格推定を行うことを検討している。また、Google の自然言語処理 API は極性判定であるため、個別感情について分類できるものを目指している。

### 5.4 映像について

現在は表情推定を行っているが、色や OCR で読み取った文字列なども同時に学習評価が出来る仕組みの構築を目指している。

### 5.5 本技術の活かし方

本技術は、金融情報への応用だけでなく、例えばコミュニケーションロボットなどで、マルチモーダルな AI を用いることで、コンテキストを読み取り、今までより精度の高い認識を行うことが出来るロボットの開発に応用出来るのではないかと考えている。

現在開発を行っているものも、感情を認識し発話を行う段階には至っているが、コンテキストを読み取ることは難しいため、本技術を応用することで、より精度の高い製品を作ることを目標としている。

## 4.3.6 LSTM での予測

## 6. 謝辞

本研究は、コンセプトをお話する中で、株式会社 Quick 様に株式会社 StockVoice 様をご紹介頂き、実現致しました。

また、株式会社 Quick 様より日足、歩み値などのレート情報と公募価格についてご提供を頂きました。東北大学の伊藤・能勢研究室に JTES のデータ提供を受けております。共著に入っていない方で数多くの学生さん、先生の協力を得て実現しました。本当にありがとうございました。

## 参考文献

- [1] 鈴木基之：音声に含まれる感情の認識，日本音響学会誌，Vol. 71，No. 9，pp.484-489，(2015)
- [2] デイビッド・ルイス（著），武田玲子（翻訳）（2013）『買ったがる脳』日本事業出版社
- [3] 伊勢隆一郎 『人は感情で物を買う』（2015）フォレスト出版
- [4] 日 経 新 聞 ニ ュ ー ス  
<https://www.nikkei.com/markets/kigyo/ipo/public-price/>（20180305）