

機械学習を用いたアナリストレポート分析と 投資判断レーティング予測

Analyst Reports Analysis And Investment Decision Rating Prediction Using Machine Learning

鈴木章悟¹ 小澤誠一^{1,2}* 渡辺 一男³
廣瀬勇秀³ 池田 佳弘³ 飯塚 正昭³ 西田 大輔³
Shogo SUSUKI¹ Seiichi OZAWA^{1,2} Kazuo WATANABE³
Takehide HIROSE³ Yoshihiro IKEDA³ Masaaki IIZUKA³ Daisuke NISHIDA³

¹ 神戸大学大学院 工学研究科

¹ Graduate School of Engineering, Kobe University

² 神戸大学 数理・データサイエンスセンター

² Center for Mathematical and Data Sciences, Kobe University

³ 三井住友 DS アセットマネジメント株式会社

³ Sumitomo Mitsui DS Asset Management Company, Limited

Abstract: Fund managers at investment trust management companies make investment policy decisions by referring to the results of research on candidate companies for investment compiled by analysts. However, when there are many candidate companies, it is necessary to refer to a considerable number of reports, which is considered difficult to read carefully. Therefore, technology is required to (1) accurately determine the business sentiment of the companies concerned and (2) extract important information for investment decisions from the contents of the reports. In this study, we developed a machine learning model that predicts the rating, which is a rating index for investment decisions, in order to support the work of fund managers, especially for the requirement (1). There are two types of ratings that affect investment decisions: outperform and underperform. Since the number of cases that fall into these two categories is small compared to other ratings, we attempted to expand the data of documents that give the same rating. As an existing data expansion method, there is a method to expand data by synonyms that can be obtained from WordNet. In this study, we propose a method to expand data based on the frequency of occurrence in financial documents. As a result of experiments, we verified the effectiveness of the proposed data expansion method.

1 はじめに

投資信託会社における、投資信託の開発や売買の指示、投資判断といった業務は、ファンドマネージャーと呼ばれる専門家を中心として行われる。ファンドマネージャーは、様々な分野における専門家の分析結果を参考に運用方針を決定する。その専門家の一人はアナリストと呼ばれ、投資候補企業の財務や業績の分析、調査訪問を通じて得られた情報をアナリストレポートにまとめ提出する。ファンドマネージャーは、アナリ

ストレポートやその他の分析結果に基づいて投資決定を行うが、アナリストレポートに記述される文章量は多く、また、数多くの企業に対して定期的に作成されるため、全ての内容を人手で瞬時に把握することは困難である。そこで、アナリストレポートに機械学習を適用することで、投資判断に有用な情報を抽出し、より迅速で広範囲な投資行動を可能にする技術が求められる。

アナリストレポートに機械学習を適用し、投資判断支援に繋げる研究は既に存在する [1][2][3]。小林ら [4]の研究では、「手がかり表現」と呼ばれる、アナリストの予想根拠情報を抽出する際に有効な手がかりとなる表現を自動的に獲得し、それに基づいてアナリスト予想

*連絡先： 神戸大学 数理・データサイエンスセンター
〒 657-8501 神戸市灘区六甲台町 1-1
E-mail: ozawasei@kobe-u.ac.jp

根拠文の抽出が可能であることを示した。そして、抽出された予測根拠文に対して高い精度で極性付与が可能であることを示した。小林らの手法における手がかり表現の抽出では、人手にて少数の手がかり表現（「予想する」、「考える」、「高い」）を設定し、それらの表現とともに頻出する手がかり表現を自動抽出する。自動獲得した手がかり表現を含む文をアナリストの予想根拠文とし極性（ポジティブ・ネガティブ）が付与される。この手法では、手がかり表現を含む文に対してのみ極性付与を行うため、手がかり表現を含まない重要な文を取りこぼす可能性がある。また、単一の文に対して極性判定が行われるため、レポート内の複数の文同士の関係性を考慮した極性判定が行われていない。

本論文では、特定のルールによって文のフィルタリングを行うことなく、アナリストレポート内で共通して用いられる観点に基づき、レポートに付与された極性を予測する。具体的には、アナリストレポート1件の文章を入力とし、機械学習を用いてそれぞれの観点の文章を特徴ベクトルに変換し、レポートに付与されたレーティングを予測するモデルを開発する。アナリストやファンドマネージャーが注目する観点に基づいた推論構造を組み込むことで、将来的にモデルの判定根拠を可視化する際、その根拠の出处が明確化されると共に、レーティングには反映されていない企業の注目箇所を情報として提示することが可能となり、投資判断支援に繋がれると考える。また、学習データ量やレーティングの出現頻度の偏りによる、レーティング予測性能への影響を考慮し、訓練データに対してデータ拡張を適用する。

2 提案システム

本論文では、投資信託運用会社に蓄積されたアナリストレポートを用いて、アナリストレポートの文章情報から、深層学習を用いて対象企業の景況感を予測するシステムを提案する。

構築した深層学習モデルは、アナリストレポートで共通して用いられる観点別に記述された文章をそれぞれ特徴変換し、レーティングを予測するという構造をとる。また、訓練データとするアナリストレポートにデータ拡張を適用することで、予測性能の改善を図る。

2.1 投資判断レーティング予測モデル

アナリストレポートの文章情報から対象企業のレーティングを予測する深層学習モデルの構造を決定するにあたり、本研究では、アナリストレポートの文章の構成に着目した。アナリストレポートでは、アナリストやファンドマネージャーが共通して着目するいくつか

の観点に基づいて、対象企業の景況感が観点別に文章で記述される。各観点の名称は、レポートの作成者や作成時期によって異なるが、以下の6つに大別される。

- **投資判断**：企業のレーティングと大まかな根拠。
- **業績動向**：直近の業績に対する事実やその要因、今後の推移に対する見解。
- **バリュエーション**：株価収益率や株価純資産倍率などの投資尺度に関する事実や見解。
- **中長期カタリスト**：相場の変動の端緒となることが予想される要因。
- **マネジメント**：組織運営や経営管理の状況。
- **リスク**：業績に悪影響を及ぼすと予想される要因。

アナリストレポートでは、これらの各観点に基づいて分析された企業の状況を総合的に評価し、その企業に投資すべきかどうかの判断がレーティングという指標として付与される。本提案システムにおける深層学習モデルは、レポート中の各観点別の文章の情報を総合的に評価し、レーティングを予測する構造を採用した。

提案システムにおける深層学習モデル（以下「提案モデル」）の具体的な説明を行う。提案モデルの主な構成要素は、双方向の時間の流れを学習可能な Long Short-Term Memory[5] (BiLSTM) や Transformer[6] のエンコーダ部分である。入力として、アナリストレポート内の各文毎に、文中の単語を Word2vec[7] の単語ベクトルで置き換えたものを使用し、出力はアナリストレポートに付与されたレーティングである。

提案モデルの計算フローを図1に示す。まず、BiLSTM（図1中の Sentence-BiLSTM）により、レポートの各文中の出現単語の特徴が考慮された文ベクトルが計算される。次に、Transformer エンコーダ（図1中の Transformer-Encoders）により、各文同士の関係性を考慮した文ベクトルが計算される。そして、BiLSTM（図1中の Document-BiLSTM）により、各観点別に文ベクトルの情報が、観点別の文章ベクトルに畳み込まれる。最後に、各観点の文章ベクトルは連結され、階層型ネットワークを経てレーティングの予測値が出力される。

2.2 金融文書に対するデータ拡張の適用

投資候補企業に対して各レーティングが付与される頻度は、必ずしも均等ではない。特に、「O（アウトパフォーム）」や「U（アンダーパフォーム）」と言ったレーティングは、その銘柄の売買が推奨される際に付与され、ニュートラルなレーティング（「N-」や「N+」）

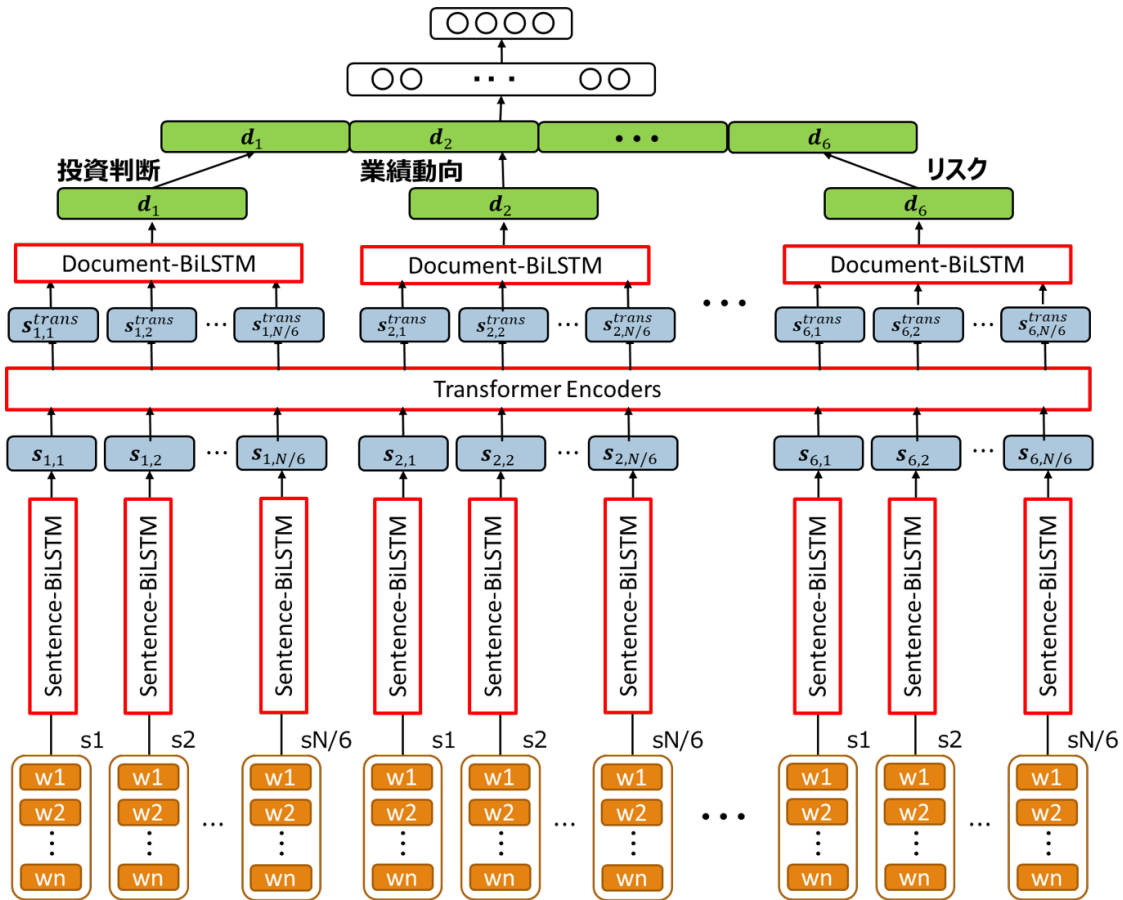


図 1: 深層学習を用いたレーティング判定モデル

に比べ運用成績への影響度は大きくなる。従って、その銘柄への投資判断を促す確信的な要因がない場合には、これらのレーティングは付与されにくくなるという側面がある。また、多数の企業が、長期的な著しい株価上昇率や下落率を示すという状況も考えにくい。これらの要因により、「O」や「U」のレーティングが付与されるアナリストレポートの件数は「N-」や「N+」が付与されたレポートの件数に比べ少数となり、提案モデルの学習に悪影響を及ぼす。

この問題を解決するために、訓練データに対してデータ拡張を適用して提案モデルを学習した。適用するデータ拡張手法は、テキストデータに対する既存のデータ拡張手法 [8] を応用した。既存研究では、訓練データ中の各文に対し、次の処理からランダムに一つの処理を選択し実行する。

1. 同義語置換 (SR): 文中のストップワードでない単語をランダムに n 個選ぶ。選択した単語をランダムに選んだ同義語のうちの 1 つに置き換える。
2. ランダム挿入 (RI): 文中のストップワードでない

ランダムな単語の同義語を選択し、その同義語を文中のランダムな位置に挿入する。これを n 回行う。

3. ランダムスワップ (RS): 文中の 2 単語をランダムに選び、その位置を入れ替える。これを n 回行う。
4. ランダム削除 (RD): 文中の各単語を確率 p でランダムに削除する。

ただし、 n は、文の単語数 l と任意の割合 α によって $n = \alpha l$ と設定する。また、確率 p は $p = \alpha$ と設定する。また、各訓練データの拡張数は n_{sug} 回とし、既存の訓練データそれぞれに対して n_{sug} 個の拡張データが得られる。同義語は、WordNet [9] と呼ばれる概念辞書から選択される。WordNet とは、各英単語の定義や同義語のグループが記述された公開データベースである。

本研究では、この既存研究をアナリストレポートに適用する。訓練データとするアナリストレポート一件中の各文に対し、上の処理からランダムに一つの処理を選択し実行する。この処理を各レポートに n_{aug} 回繰り返す。

表 1: 各アナリストレポートの件数

レーティング	件数
O (アウトパフォーマンス)	476
N+ (中立よりややポジティブ)	2105
N- (中立よりややネガティブ)	1538
U (アンダーパフォーマンス)	43

繰り返し行うことで訓練データの拡張を行う。同義語選択には、日本語版の WordNet を使用する。

3 実験設定

3.1 実験に使用したデータ

本研究で使用したデータは、1998 年から 2016 年のアナリストレポート 21547 件である。このアナリストレポートのうち、「O」、「N+」、「N-」、「U」のいずれかのレーティングが付与されたアナリストレポートを使用した。また、記述されている各段落の文章に対して、2.1 節で説明したどの観点についての文章であるかが自動的に識別可能なアナリストレポートを解析の対象とした。

これらの条件を満たすアナリストレポートとして、4162 件が得られた。また、それぞれレーティングに該当するレポートの件数を表 1 に示した。

3.2 レーティング予測モデルの実験設定

提案モデルの学習において、入力データに使用する Word2vec の事前学習は、アナリストレポート並びに、TIS 株式会社が公開するデータセット CoARiJ (Corpus of Annual Reports in Japan)¹から取得した 2014 年から 2018 年の 5 年分の有価証券報告書を使用して行なった。単語ベクトルの次元数は 300 次元とした。Word2vec の実装には gensim² と呼ばれるオープンソースライブラリを使用した。また、形態素解析器として MeCab³ を使用し、MeCab の辞書として、Neologism dictionary (NEologd)⁴ を使用した。

さらに、1 文あたりの最大の単語数を 20 単語とし、1 つの観点の文章中の最大の文の数を 15 文とした。

その他、提案モデルの各構成要素における入出力の次元を表 2 に示した。学習時のバッチサイズは 64 とし、学習率は 10^{-4} とした。また、最適化アルゴリズムとして Adam を使用した。深層学習モデルの実装には、

表 2: 提案モデルの構成要素と各入出力の次元

モデルの構成要素	次元
モデルの入力 ⁷	(6,15,20,300)
Sentence-BiLSTM(入力) ⁸	(20,300)
Sentence-BiLSTM(出力)	(1,600)
モデル全体の出力 ⁹	(90,600)
Transformer-Encoders(入力) ¹⁰	(90,600)
Transformer-Encoders(出力)	(90,600)
Document-BiLSTM (入力) ¹¹	(15,600)
Document-BiLSTM (出力)	(1,600)
モデル全体の出力 ¹²	(6,600)
連結後の出力・全結合層の入力	(1,3600)
全結合層の出力	(1,300)
モデルの出力 (ソフトマックス) ¹³	(1,4)

オープンソースライブラリの TensorFlow を基礎とした API である Keras⁵ を使用した。また、Transformer は、公開されている実装済みの Transformer⁶ のエンコーダ部分のみを使用した。

比較として、観点毎に文章の特徴変換を行う構造 (図 1 の「Document-BiLSTM」) と、文ごとに単語ベクトルの特徴変換を行う構造 (図 1 の「Sentence-BiLSTM」) それぞれに対して評価を行うために、以下の 2 種類のレーティング予測モデルを構築し、性能評価を行なった。

観点毎に文章の特徴変換を行う構造を評価する比較モデルとして、表 3 に示す構造を持つ深層学習モデル (比較モデル 1) を構築した。変更点として、表 2 の Document-BiLSTM ではなく、レポート内の全文のベクトルを入力とし文章ベクトルに変換する BiLSTM (表 3 の「All-Document-BiLSTM」) を導入した。

さらに、文ごとに単語ベクトルの特徴変換を行う構造を評価する比較モデルとして、表 4 に示す構造を持つ深層学習モデル (比較モデル 2) を構築した。変更点として、表 2 の Sentence-BiLSTM ではなく、レポート内の全文のベクトルを入力とし文章ベクトルに変換する BiLSTM (表 4 の「All-Words-BiLSTM」) を導入し

⁵<https://www.tensorflow.org/guide/keras?hl=ja>

⁶<https://github.com/CyberZHG/keras-transformer>

⁷各次元数において 6 はアナリストレポートで使用される 6 つの観点に対応し、15 は各観点到記述された文の数、20 は各文の単語数、300 は単語ベクトルの次元数に対応する。

⁸「モデルの入力」における、各文の単語ベクトルに相当する特徴量 (次元は (20,300)) に対して、一文ずつ 600 次元のベクトルに特徴変換し出力される (Sentence-BiLSTM(出力))。

⁹「Sentence-BiLSTM」によって $90 (= 6 \times 15)$ 文のベクトルが算出される。

¹⁰「Transformer-Encoders」により、レポートの各文同士の関係性を考慮した文のベクトルが算出される。

¹¹各観点的文章に対応する 15 個の文ベクトルを入力 (Document-BiLSTM (入力)) とし、それぞれを 600 次元のベクトルとして算出 (Document-BiLSTM (出力)) する。

¹²レポートで使用される 6 つの観点の文章のベクトルが計算される。

¹³4 つのレーティングのいずれかが算出される。

¹<https://github.com/chakki-works/CoARiJ>

²<https://radimrehurek.com/gensim/models/word2vec.html>

³<https://taku910.github.io/mecab/>

⁴<https://github.com/neologd/mecab-ipadic-neologd>

表 3: 比較モデル 1 (観点毎の文章の特徴変換を行わないモデル) の構成要素と各入出力の次元

モデルの構成要素	次元
モデルの入力	(6,15,20,300)
Sentence-BiLSTM(入力)	(20,300)
Sentence-BiLSTM(出力)	(1,600)
モデル全体の出力	(90,600)
Transformer-Encoders(入力)	(90,600)
Transformer-Encoders(出力)	(90,600)
All-Document-BiLSTM (入力)	(90,600)
All-Document-BiLSTM (出力)	(1,600)
全結合層の出力	(1,300)
モデルの出力 (ソフトマックス)	(1,4)

表 4: 比較モデル 3 (文の特徴変換を行わないモデル) の構成要素と各入出力の次元

モデルの構成要素	次元
モデルの入力	(6,15,20,300)
All-Words-BiLSTM(入力)	(1800,300)
All-Words-BiLSTM(出力)	(1800,600)
Sentence-Average(出力)	(90,600)
Transformer-Encoders(入力)	(90,600)
Transformer-Encoders(出力)	(90,600)
Document-BiLSTM (入力)	(15,600)
Document-BiLSTM (出力)	(1,600)
モデル全体の出力	(6,600)
連結後の出力・全結合層の入力	(1,3600)
全結合層の出力	(1,300)
モデルの出力 (ソフトマックス)	(1,4)

た。そして、各文の単語ベクトルに対応する時刻における、「All-Words-BiLSTM」の隠れ層の出力値を平均し、文ベクトルが計算される構造を導入した。

性能評価は 5-fold の交差検証で行い、評価指標として macro-F1 スコアを使用した。

3.3 データ拡張の実験設定

同義語選択に用いる日本語の WordNet としては、NLTK により実装された WordNet¹⁴を使用し、データ拡張のパラメータは既存研究の推奨値に設定した。具体的に、 α と p は 0.1 とし、 n_{aug} は 4 とした。さらに、データの拡張回数である n_{aug} は 2, 4, 6, 8 と変化させて比較実験を行なった。

拡張された訓練データを用いて、レーティング予測モデルを学習した。レーティング予測モデルの構造や

表 5: レーティング予測における F 値の比較

比較手法	macro-F1
提案モデル	0.413 ± 0.064
比較モデル 1	0.375 ± 0.035
比較モデル 2	0.510 ± 0.035*

学習条件は、3.2 節の表 2 と同様の設定を採用した。

性能評価は 5-fold の交差検証で行い、評価指標として macro-F1 スコアを使用した。

4 性能評価

4.1 レーティング予測モデルの性能評価

3.2 項で説明した実験設定に基づくレーティング予測モデルの比較実験の結果を表 5 に示す。提案モデルは、比較モデル 1 に比べて macro-F1 スコアが上昇しているが、比較モデル 2 と比較すると低い性能を示した。ただし、サンプルサイズ 5 の Welch の t 検定によると、提案モデルと比較モデル 1 との比較では、平均値に差がない可能性が高い結果となった。また、提案モデルと比較モデル 2 との比較では、有意水準 5% で平均値に差がある結果となった。このことから、提案モデルにおいて、有意差はなかったものの、観点別に文章のベクトルを計算するアーキテクチャの有効性を概ね示すことができた。これは、提案モデルにおいて、レーティング予測に必要な情報が、観点別の文章ベクトル計算することにより効率よく学習されたことが原因だと考えられる。しかし、提案モデルにおける、文ごとに単語ベクトル情報を畳み込む BiLSTM の有効性は示すことができなかった。この原因として、提案モデルにおいて、一文の中のみで単語ベクトルの特徴変換を行なったために、文同士の単語間の情報共有が不足してしまったことが考えられる。

提案モデルの性能をより改善するには、文ごとに単語ベクトルを変換し文ベクトルを計算する処理 (図 1 の sentence-BiLSTM) の前に、比較手法のようなレポート内の全単語間で特徴変換を行う処理 (表 4 の「All-Words-BiLSTM」) や、観点別の文章内で文の特徴変換を行う処理を追加が必要なのではないかと考察している。

4.2 データ拡張適用時の性能評価

3.3 で説明した実験設定に基づき、既存研究を応用したデータ拡張方法の評価実験を行なった。データ拡張適用時と、未適用時の訓練データを用いて学習したときの比較実験の結果を表 6 に示す。また、各データの

¹⁴<https://www.nltk.org/howto/wordnet.html>

表 6: データ拡張時と未適用時の比較結果

比較手法	macro-F1
データ拡張あり	0.564 ± 0.031**
データ拡張なし	0.413 ± 0.064

表 7: データ拡張数を変更したときの比較結果

比較手法 ($n_{aug} =$)	macro-F1
2	0.548 ± 0.039
4	0.564 ± 0.031
6	0.565 ± 0.049
8	0.547 ± 0.040

データの拡張数 n_{aug} を変化させたときの比較実験の結果を表 7 に示す。

表 6 の結果から、提案するデータ拡張手法を適用することによって、レーティング予測の macro-F1 スコアが 0.15 改善し、有意水準 1% で平均値に差がある検定結果となった。

これは、データ拡張を適用することによって、アナリストレポートの話題やレーティングを変えない程度に訓練データ数が増えたことが原因だと考えられる。例えば、「営業利益」や「経常利益」と言った単語は、同義語とはいかないまでも、一見すると似た意味の単語と認識される可能性がある。しかし、これらは、金融分野においては、全く別の話題や概念として捉えられ、置き換わると文の意味が大きく変わってしまう。本実験では、一般的な単語が登録されている概念辞書である WordNet を使用したため、上で挙げた単語が置き換わることはない。これにより、アナリストレポートの話題やレーティングを変えない程度にデータ拡張が適用されたことで、予測性能が改善されたと考えられる。

さらに、表 7 より、データの拡張数を変更したとき、 n_{aug} が 4 や 6 に設定した場合に最も性能が改善し、 n_{aug} が 8 の場合は n_{aug} の場合よりも macro-F1 スコアが低下する結果となった。これは、データ拡張をし続けたとしても、サンプリングされる訓練データの質が上限に達し、モデルの汎化性能の向上に繋がらなかったためだと考えられる。

5 結論

本論文では、ファンドマネージャーの業務支援に向けて、投資信託運用会社に蓄積されているアナリストレポートの文章から、当該企業の景況感を正確に判定するシステムを提案した。具体的には、アナリストレポートの文章の記述方法に基づき、各観点の文章をそれぞれ別々にベクトル化し、レーティング予測を行う深

層学習モデルを構築した。また、学習に使用したデータ量やレーティングの出現頻度の偏りを補填するために、訓練データに対してデータ拡張を適用した。

実験では、提案したレーティング予測モデルとデータ拡張の性能評価を行なった。提案した機械学習モデルの性能評価実験において、レポート内の全文を同時に特徴変換する構造に変更したモデルとの比較実験では、提案手法がより高い macro-F1 スコアを示した。ただし、レポート内の文ごとに特徴変換を行う構造を廃した比較手法との比較においては、比較手法がより高い macro-F1 スコアを示し、提案モデルには一部改善の余地があることが分かった。

金融文書に対するデータ拡張を適用した評価実験では、データ拡張未適用時との比較実験と最適な拡張数の探索を行い、選択する同義語と元の単語との類似度の比較実験を行なった。データ拡張未適用時との比較実験では、データ拡張時が有意に高い macro-F1 スコアを示し、データ拡張数が 4 または 6 で最高性能を示した。

実験結果より、レーティング予測のための深層学習モデルとデータ拡張方法においては、提案モデルにおける観点別に文章の特徴変換を行う構造の有効性を示すことができた。また、特にデータ拡張を適用することにより、データ拡張未適用時に比べ大きく性能を改善することができた。

今後は、機械学習モデルの構造やデータ拡張の方法に変更を加え、レーティングの予測性能のさらなる向上を目指すとともに、本研究の元々の目的である、判定根拠の可視化のための有効な方法の開発及び検証を行い、投資判断に繋げていきたいと考えている。

参考文献

- [1] 坂地泰紀, 平松賢士, 酒井浩之. 深層学習を用いて極性付与されたアナリストレポートと株式リターンとの関連性. 第 20 回 人工知能学会 金融情報学研究会資料, 2018.
- [2] 北島良三, 酒井浩之, 上村龍太郎, and 栗田昌孝. 坂地泰紀, 平松賢士. アナリストレポートと企業業績の関係解析 (第一報). 第 22 回 人工知能学会 金融情報学研究会資料, 2019.
- [3] 高山将丈, 小澤誠一, 廣瀬勇秀, 飯塚正昭, 渡辺一男, and 逸見龍太. アナリストレポートにおけるキーワード関連文の抽出と景況感推移観測への応用. 第 24 回 人工知能学会 金融情報学研究会資料, 2020.
- [4] 小林和正, 酒井浩之, 坂地泰紀, and 平松賢士. アナリストレポートからのアナリスト予想根拠情報の抽

出と極性付与. 第 19 回 人工知能学会 金融情報学
研究会資料, 2017.

- [5] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [6] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- [7] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.
- [8] Jason Wei and Kai Zou. Eda: Easy data augmentation techniques for boosting performance on text classification tasks. *arXiv preprint arXiv:1901.11196*, 2019.
- [9] George A Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.