The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

経営情報キュレーションのための事前学習付き重要文ランク学習

Neural Ranking with Pretraining for Curation of Important Sentences in Business Operations

是枝祐太 Yuta Koreeda　黒土健三 Kenzo Kurotsuchi　柳瀬利彦 Toshihiko Yanase　柳井孝介 Kohsuke Yanai　間瀬久雄 Hisao Mase　佐藤美沙 Misa Sato

日立製作所研究開発グループ
Research & Development Group, Hitachi, Ltd.

We have developed a machine learning method for ranking sentences based on their importance, to assist decision making in business operations. The proposed method is based on deep learning that is trained end-to-end from user feedbacks to capture semantic importance of sentences without relying on domain knowledge of business operations. We employ pretraining to avoid needs for large training data, which we cannot obtain easily because there are numerous groups of people with different interest in business operations and the each group tends to be small. The proposed method outperformed conventional ranking methods in an automated evaluation. We validated the quality of the ranking with a subjective impression evaluation.

1. 重要性に基づくランキング

営業活動，経営管理，知的財産，マーケティングなどの経営活動において，合理的な意思決定のためには多様で，深い見識が必要である．自らの役割に関係のあるニュース記事，有価証券報告書などの経営情報を日々読むこと理想的であるが，アクセス可能な情報のすべてを読み解くのは現実的ではない．意思決定を支援するために，それぞれのユーザにとって重要度が高い情報を効率よく集める方法が必要とされている．

効率的に重要度の高い情報を集める方法として，情報を重要度に基づきランキングし，ユーザにとって重要と思われる情報のみを提示するキュレーションが注目されている．経営活動に関するキュレーションは以下理由から人手で行うコストが高い．

- ユーザがどのような情報を重要とするかを理解するためには高度なドメイン知識が求められる．
- そのようなドメイン知識はユーザ群の特性により無数に存在する．

本研究の目的は，ユーザにとっての情報の重要度をユーザフィードバックから学習し，情報を自動でランキングできるシステムを実現することである．特に，企業の一部署など類似した嗜好性を持つユーザ群のために，異なる日付，あるいは異なる取引先など，異なる情報のグループごとに，重要な情報を文単位でまとめ，提示する問題に取り組む．

情報のランキングは情報検索のコンテキストで盛んに研究されている（[Liu 09]）．情報検索におけるランキングでは検索クエリと検索対象の関係性を捉えることが主たる研究対象とされてきた．本研究で対象とするキュレーションには検索クエリが存在せず，文の内容自体の重要性に基づくランキングを対象にしている点が異なる．

検索クエリとの関係性に基づかない，あるいは検索クエリとの関係性を補完するために使われる重要性に基づくランキングは，リコメンデーションや情報検索などで使われてきた．リコメンデーションで使われる協調フィルタリングはユーザ間の嗜好の類似性を基礎としているため，新しい情報を他のユーザに薦めるためには，その情報に対して十分なフィードバックを取得する必要がある．経営情報キュレーションでは，各ユーザ群の大きさが限られ，かつ新しい情報が重要とされるため，これらの手法は適用できない．情報検索の枠組みで重要性をモデリングする方法として例えば [Page 99] が知られているが，この手法はウェブサイト間の被リンク関係を用いているため，協調フィルタリングの場合と同様に，本研究の用途には効果が小さい．

ランク学習において，文の内容から重要性をモデリングしようとした試みに [Duan 10, Beigman Klebanov 14] などがあるが，これらの研究ではドメイン依存の重要性を捕らえるために相応の特徴量エンジニアリングを行っている．前述したとおり，経営活動を支援するためには異なる嗜好を有するユーザ群の要求に応える必要があるため，ドメイン知識を前提とした手法は経営支援には適さない．一般的な情報の質について検討した研究には [Lee 02, Zhu 00] があるが，本研究でとらえようとしているような経営判断における意味的な重要性については言及していない．

以上より経営情報キュレーションのための重要文のランク学習には，次の要素が特に重要だと考えられる．

1. メタ情報や，他ユーザの評価からだけではなく，文の意味情報から重要性をとらえ，ランキングを行う必要がある．
2. 重要性の観点はユーザ群ごとに多様であるため，人手で重要性のドメイン知識をモデリングすることは望ましくない．
3. 異なる嗜好を持つユーザ群が多数存在する一方で，各ユーザ群の大きさは限られるため，(2) を補うために大量のデータを取得することが難しい．

本研究では，ドメイン知識へ依存することなく意味情報の抽出とランキングを end-to-end で学習できる深層学習によるランク学習法を開発する．データが大量に必要であるという深層学習の欠点は経営情報キュレーションにおいて致命的であるため，それ補うために経営文書に特に効果的と考えた [Kiros 15] の事前学習を用いることを提案する．本報告では，自動評価に基づく評価実験で提案手法をベースラインと比較し（3.2 節），実験の妥当性を検討するため

連絡先: 是枝祐太，株式会社日立製作所，〒 185-8601 東京都国分寺市東恋ヶ窪一丁目 280 番地，042-323-1111，yuta.koreeda.pb@hitachi.com

に人手による A/B テストを行う（3.3 節）.

2. 経営の意思決定支援のためのランク学習

2.1 問題設定

本研究における問題設定を定義する.

複数の文から構成される情報のグループが複数個と，類似した嗜好性を持つユーザ群があるとする．ここでグループとは，同一ドメインだが条件が異なる文の集合であり，たとえば異なる日付ごとのニュース記事や，異なる会社の有価証券報告書である．あるユーザが 1 つのグループを選択すると，ランク関数によりグループ内の文がランキングされ，所定の数の上位文がユーザに提示される．ユーザは提示された文に対して，ユーザの考える重要性の観点から，*Like* ボタンのような形式で各文にフィードバックを与える．1 回の提示で得られた文とフィードバックのペアが複数個集まったものをセッションと呼ぶこととする．前述した提示とフィードバックを異なる情報のグループ，異なるユーザで行い，得られた複数のセッションからランク関数を更新する.

2.2 意味表現エンコーダ

文を入力に，文の意味を表現する固定長のベクトル出力する関数をエンコーダと呼称する．そこで，本研究ではディープラーニングを用い，文から end-to-end でランキングを行う.

セッション i の文 j が $L_{i,j}$ 個の形態素から構成されるとする.

$$\mathbf{x}_{i,j} = \{\mathbf{x}_{i,j,t}\}_{1 \leq t \leq L_{i,j}} \tag{1}$$

ただし $x_{i,j,t}$ はセッション i の文 j の t 番目の形態素の one-hot 表現である．各行に one-hot 表現に対応する分散表現を持つ行列 $\mathbf{W}_E$ を用い，各形態素を分散表現 $\mathbf{u}_{i,j}$ に埋め込む.

$$\mathbf{u}_{i,j} = \mathbf{x}_{i,j,t} \cdot \mathbf{W}_E \tag{2}$$

Long Short-Term Memory([Hochreiter 97]) を用い，形態素の分散表現列から，LSTM に $\mathbf{u}_{i,j,L_{i,j}}$ を入力したときの LSTM の状態を，文の意味表現ベクトル $\mathbf{e}_{i,j}$ として取得する.

$$\mathbf{e}_{i,j} = \mathrm{LSTM}\left([\mathbf{u}_{i,j,1}, \mathbf{u}_{i,j,2}, \cdots]\right) \tag{3}$$

2.3 経営文書に適した事前学習

ビジネス文書はルールや慣習により定型文が頻繁に用いられる. 例えば有価証券報告書では, 企業のとった具体的な施策のあとにはその施策が経営指標に与えた影響が書かれることが多い．したがって，似たような意味合いの文で挟まれた文は同じようなトピック，意味的な特徴を共有していると考えられる．このようなビジネス文書の特徴を踏まえ，RNN を用い前後の文を予測する事前学習 ([Kiros 15]) を活用し，得られるデータが少ないことを補う．エンコーダの出力 $\mathbf{e}_{i,j}$ を入力に，前後の文を予測するデコーダを導入し，エンコーダとデコーダを同時に学習する.

2.4 文の意味表現を用いたランク学習

ランク学習においては，デコーダは破棄し，エンコーダの出力のみを使用する．文ごとに全結合層を使い，文の意味表現ベクトル $\mathbf{e}_{i,j}$ から予測スコア $y_{i,j} \in \mathbb{R}$ を取得する.

$$\mathbf{y}_i = \sigma(\mathbf{e}_{i,j} \cdot \mathbf{W}_0 + \mathbf{b}_0) \cdot \mathbf{W}_1 + \mathbf{b}_1. \tag{4}$$

ただし，σ は各要素に対する正規化線形関数 $\sigma(x) = \max(x, 0)$，$\mathbf{W}_0$，$\mathbf{W}_1$，$\mathbf{b}_0$，$\mathbf{b}_1$ はモデルパラメータである.

ユーザフィードバック $\tilde{\mathbf{y}}_i = \{y_{i,j}\}$ と，予測間の差が小さくなるように学習を行う．本研究では，ユーザは 1 回に提示された文の相対評価を行っているものと考えられるため, 2 つの並び順の差異を比較する permutation probability loss [Cao 07] を採用する. 本手法では, $k = 1$ の permutation probability distribution function $f_P : \mathbb{R}^N \mapsto \mathbb{R}^N$ を使用する.

$$f_P(\mathbf{y}_i) = \left[\frac{\exp(y_{i,j})}{\sum_n^N \exp(y_{i,n})}\right]_{j \in \{1,2,\ldots,N\}}. \tag{5}$$

最終的な損失関数は，2 つの分布の Kullback-Leibler divergence D_{KL} を用い，式 (6) とする.

$$D_{KL}\left(f_P(\tilde{\mathbf{y}}_i) \| f_P(\mathbf{y}_i)\right) + \lambda \sum_{\mathbf{W} \in \{\mathbf{W}_0, \mathbf{W}_1\}} \|\mathbf{W}\|_2 \tag{6}$$

ただし，$\| \bullet \|_2$ はフロベニウスノルムである.

損失関数から求めた勾配をもとに $\mathbf{W}_E$ とエンコーダを含むモデルパラメータをミニバッチ確率勾配法を用い学習する．パラメータの更新には Adam（[Kingma 14]）を用いる．学習データへの過剰適合を防ぐために 0.1 の Dropout（[Srivastava 14]）を式 (4) の全結合層に適用し，L2 正則化の係数 λ を 0.005 とした.

3. 評価実験

3.1 対象とする経営活動

本報告では，様々な経営活動の中でも，営業部署が顧客の動向を洗い出す工程を対象とする. 2.1 節のグループが各顧客企業の動向を示す文に対応しており，顧客名称を選択するとグループ内の文がランキングされ，所定の数の上位文が結果としてユーザに提示される.

下記に本実験において対象とした文の例[*1]を示す.

この表示方法の変更を反映させるため，前事業年度の財務諸表の組替えを行っている.

企業の行動ではあるが，会計上の手続きであり，財務部署の役には立っても営業部署にとって役に立つ情報とはいえない．一方で，下記の例では，適当な具体性を有しており，顧客が生産管理に関して注力していることが読み取れる.

製品の品質当社グループは優れた品質の製品を提供するため，開発から生産まできめ細かい管理体制を敷き最善の努力を傾けている.

3.2 自動評価に基づく評価実験

本実験の目的は，経営情報キュレーションのための重要文ランク学習において，提案手法をベースラインと比較評価することである．人手で付与したフィードバックのアノテーションからオフラインで学習を行い，テストデータ上で自動評価を行う.

*1 日産車体株式会社，2014 年度有価証券報告書より抜粋

表 1: 提案手法とベースラインの性能比較

手法		ランク性能 (nDCG)	提案手法との差
提案手法 (ランク学習	+ 意味表現エンコーダ)	**0.873**	–
SVR	+ 意味表現エンコーダ	0.854	−0.019
ランク学習	+ tf-idf	0.829	−0.044
SVR	+ tf-idf	0.843	−0.030
ランダム		0.682	−0.191

表 2: 異なる事前学習手法の比較

事前学習方法	ランク性能 (nDCG)	提案手法との差
[Kiros 15]	**0.873**	
オートエンコーダ	0.852	−0.021
事前学習なし	0.772	−0.101

[Yanai 17] を用い企業の活動を示す述語の辞書を作成し，構文解析に基づいたルールによりマッチングを行い，有価証券報告書，新聞データから各顧客企業の動向を示す文を抽出した．各企業ごとに，復元法でランダムに抽出した 4 文を 1 セッションとして，セッションの各文に +1(相手企業のことを知るうえで役にたつ)，−1(相手企業のことを知るうえで役にたたない)，0(どちらともいえない) のラベルを付与した．自動車業界の会社の 16 つの企業に対し，それぞれ 5 つのセッションを 4 人のアノテータで独立にアノテーションした (のべ 5120 文)．本実験で用いた 16 社は EDINET で有価証券報告書が閲覧できる自動車業界の会社である．訓練データとテストデータで企業が重複しないように，企業を単位とした k-分割交差検証（$k = 8$）にて各手法を評価した．

学習器の評価を行う際，テストデータを 1 企業ごとに Schulze の選挙方法で単一の順列に統合し，そのデータを予測値と比較した．抽出されたセッションに対してユーザフィードバックを受けるため，学習データとしてのセッションは短いが，実際にランキングを行うのは全文に対してであるためである．評価指標はランキング問題の指標としてよく用いられる normalized Discounted Cumulated Gain (nDCG) を採用した．

文の tf-idf を特徴量としたサポートベクトル回帰（SVR）とランダム推測をベースラインとして採用した．比較対象として，2.3 節の事前学習を行った意味表現エンコーダにより抽出した特徴量を用いた SVR と，tf-idf を特徴量とした 2.4 節のランク学習についても評価する．

提案手法とベースラインの結果を表 1 に示す．文の tf-idf を特徴量とした SVR からは 0.030，ランダム推測からは 0.191 の向上が見られた．SVR と意味表現エンコーダを組み合わせたモデルが次点であり，ドメイン知識へ依存しない意味表現の抽出を意図した意味表現エンコーダが有効であったことが示唆された．一方で，tf-idf を特徴量とした深層学習によるランク学習は，tf-idf を特徴量とした SVR よりも劣った．このことから，ランク学習に深層学習を用いたこと自体ではなく，事前学習した意味表現エンコーダを，ランク学習時に end-to-end でファインチューニングしたことが寄与したものと考えられる．

事前学習法の効果を検証するために，類似する事前学習法としてオートエンコーダを用いた場合と，事前学習を用いらなかった場合についても検証した．オートエンコーダは入力文の前後文ではなく，入力文自身を再構成する事前学習法である．

事前学習なしの条件では，tf-idf を特徴量とした SVR よりも性能が低く，ランダム推測に性能が近かった (表 2)．深層学習をそのまま経営情報キュレーションに適用することは難しく，事前学習が重要であったことが示唆された．また，[Kiros 15] の事前学習はオートエンコーダより性能が優れており，[Kiros 15] が本問題設定に対して有効であったことが確認できた．

表 3: 主観評価における順位の割合

	順位		
手法	1	2	3
提案手法	53 %	34%	13 %
SVR + tf-idf	39 %	55%	6.3%
ランダム	7.8%	11%	81 %

3.3 人手による評価実験

指標を用いた交差検証の結果は指標の選択に大きく依存するため，3.2 節の結果の妥当性を確認するために，A/B テストによる主観評価を行った．

3.2 節の提案手法，tf-idf を特徴量とした SVR，ランダム推定の 3 つの学習済みモデルを用い，それぞれ同じ企業の有価証券報告書からランキングで 4 つの文を選んだ．1 手法 1 列とし，二重盲目の設定でそれぞれ上位文を 4 人の被験者に提示し，企業ごと，モデルごとに順位 $r \in \{1, 2, 3\}$ (低い値は良い順位を示す) を付与するよう指示した．

実験結果を 表 3 に示す．提案手法は過半数の試行において，最も良いと評価された．

本研究において抽出した顧客の動向を示す文には，3.1 節に示したような単なる会計上取り決めが相当数含まれていた．これら会計上の取り決めは企業を問わず類似した表現がなされているため，提案手法と SVR はこれらを取り除くことに成功していた．会計上の取り決めを上位に出力した場合低い評価がなされることが多く，提案手法とベースラインがほとんどの場合ランダムベースラインに打ち勝っていることはこのことに起因している．

主観評価実験において相対的に提案手法が優れていたデータを抜きだしたものを表 4 に示す．SVR が最上位に出力した結果は，具体性にかけており，3.2 節のアノテーションにおいても低い評価が付与されている．“課題”，“向上” といった形態素を含む他の文は具体的な動向について触れられているものが多く，SVR はその仕組み上，形態素レベルの特徴に影響されたものと考えられる．一方で，提案手法が最上位に出力した文は企業の具体的かつ新しい取り組みについて言及されており，価値が高い．本実験では自動車業界の会社を学習にもちいているため物流事業に関する記述は稀であり，SVR では稀な内容語を捉えられなかったものと考えられる．じど提案手法の事前学習が他業界にも及んでいることが寄与したものと考えられる．

4. おわりに

効率的に重要度の高い情報を集め，経営における意思決定を支援するために，ユーザのフィードバックから，特定のユーザ群が重要と考える情報をランキングできる手法を開発した．交差検証に基づく定量的な実験において，提案手法の nDCG は 0.873 であり，非深層学習のベースライン

表 4: 主観評価における出力例と主観評価．文の隣に 3.2 節のアノテーションを Schulze 法で並べ替えた時の順位を示す．

	提案手法		SVR + tf-idf		ランダム	
出力	今後は宅配便の受け取りなど「日々の生活に時間とゆとりを生みだす」サービスの拡充を図るとともに，企業・保育園・病院・駅といった人々が集まる場所への設置による物流ロッカー事業のさらなる拡大も視野に入れ，本サービスを開始します．♠	*	より強固な経営基盤を築き企業価値の一層の向上に向け，グループの総力をあげて次の課題に取り組んでまいります．♡	4	（注）1「自動車」につきましてはトヨタ自動車株式会社および株式会社デンソーから生産計画の提示を受け，生産能力を勘案し，見込生産を行っているため，記載を省略しております．◇	3
	連結財務諸表提出会社を中心として，「魅力ある新商品の開発」という考えのもとに，年々高度化・多様化する市場のニーズを先取りし，お客様の満足が得られるよう，先進技術を導入した積極的な新商品開発を進めております．◇	2	品質第一に徹してお客様の信頼におこたえいたしますとともに，各市場の動きに的確に対応して，販売の拡大に努めてまいりました．♣	1	より強固な経営基盤を築き企業価値の一層の向上に向け，グループの総力をあげて次の課題に取り組んでまいります．♡	4
	品質第一に徹してお客様の信頼におこたえいたしますとともに，各市場の動きに的確に対応して，販売の拡大に努めてまいりました．♣	1	連結財務諸表提出会社を中心として，「魅力ある新商品の開発」という考えのもとに，年々高度化・多様化する市場のニーズを先取りし，お客様の満足が得られるよう，先進技術を導入した積極的な新商品開発を進めております．◇	2	連結財務諸表提出会社を中心として，「魅力ある新商品の開発」という考えのもとに，年々高度化・多様化する市場のニーズを先取りし，お客様の満足が得られるよう，先進技術を導入した積極的な新商品開発を進めております．◇	2
	（注）1「自動車」につきましてはトヨタ自動車株式会社および株式会社デンソーから生産計画の提示を受け，生産能力を勘案し，見込生産を行っているため，記載を省略しております．◇	3	（注）1「自動車」につきましてはトヨタ自動車株式会社および株式会社デンソーから生産計画の提示を受け，生産能力を勘案し，見込生産を行っているため，記載を省略しております．◇	3	加えて，当期純利益等の表示の変更および少数株主持分から非支配株主持分への表示の変更を行っております．♣	6
人手評価	3.0		1.75		1.25	

♠ 日本経済新聞，2016 より抜粋　♣ 株式会社豊田自動織機，2015 年度有価証券報告書より抜粋　◇ 株式会社豊田自動織機，2014 年度有価証券報告書より抜粋
♡ 株式会社豊田自動織機，2013 年度有価証券報告書より抜粋　* 順位情報なし．3.2 節ではセッションを無作為抽出しているため，アノテーションがなされていない文がある．

を上回った．事前学習なし，あるいは異なる方法で事前学習を行った場合と比較し性能が向上したことから，事前学習を行うことが重要であったこと，前後のコンテキストを予測する事前学習法が定型性の高い経営文書において有効であったことが示唆された．実験の妥当性を評価するために A/B テストを行った．提案手法は過半数の試行において，最も良いと評価され，経営活動のための重要文キュレーションにおいて提案手法が有効であることが示唆された．

本報告における実験は規模が小さかったため，今後規模を大きくし実験を行う．また，異なる手法間の比較実験だけではなく，実際の経営活動を更に詳しく分析し，より多面的に手法の有用性を評価する．

参考文献

[Beigman Klebanov 14] Beigman Klebanov, B., Madnani, N., Burstein, J., and Somasundaran, S.: Content Importance Models for Scoring Writing From Sources, in *the 52nd Annual Meeting of the Association for Computational Linguistics*, pp. 247–252 (2014)

[Cao 07] Cao, Z., Qin, T., Liu, T.-Y., Tsai, M.-F., and Li, H.: Learning to Rank: From Pairwise Approach to Listwise Approach, in *the 24th International Conference on Machine Learning*, pp. 129–136 (2007)

[Duan 10] Duan, Y., Jiang, L., Qin, T., Zhou, M., and Shum, H.-Y.: An Empirical Study on Learning to Rank of Tweets, in *the 23rd International Conference on Computational Linguistics (Coling 2010)*, pp. 295–303, Beijing, China (2010)

[Hochreiter 97] Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, *Neural Computation*, Vol. 9, No. 8, pp. 1735–1780 (1997)

[Kingma 14] Kingma, D. and Ba, J.: Adam: A Method for Stochastic Optimization, in *3rd International Conference on Learning Representations* (2014)

[Kiros 15] Kiros, R., Zhu, Y., Salakhutdinov, R. R., Zemel, R., Urtasun, R., Torralba, A., and Fidler, S.: Skip-Thought Vectors, in *Advances in Neural Information Processing Systems 28*, pp. 3294–3302 (2015)

[Lee 02] Lee, Y. W., Strong, D. M., Kahn, B. K., and Wang, R. Y.: AIMQ: A Methodology for Information Quality Assessment, *Information & Management*, Vol. 40, No. 2, pp. 133–146 (2002)

[Liu 09] Liu, T.: *Learning to Rank for Information Retrieval*, Foundations and Trends(r) in Information Retrieval, Now Publishers (2009)

[Page 99] Page, L., Brin, S., Motwani, R., and Winograd, T.: The PageRank Citation Ranking: Bringing Order to the Web., Technical Report 1999-66, Stanford InfoLab (1999)

[Srivastava 14] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R.: Dropout: A Simple Way to Prevent Neural Networks from Overfitting, *Journal of Machine Learning Research*, Vol. 15, pp. 1929–1958 (2014)

[Yanai 17] Yanai, K., Sato, M., Yanase, T., Kurotsuchi, K., Koreeda, Y., and Niwa, Y.: StruAP: A Tool for Bundling Linguistic Trees through Structure-based Abstract Pattern, in *the 2017 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 31–36 (2017)

[Zhu 00] Zhu, X. and Gauch, S.: Incorporating Quality Metrics in Centralized/Distributed Information Retrieval on the World Wide Web, in *the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 288–295 (2000)