1J1-4

文書分類とのマルチタスク学習による重要文抽出

Extractive Summarization using Multi-Task Learning with Document Classification

磯沼 大 *1 Masaru Isonuma　藤野 暢 *1 Toru Fujino　浮田 純平 *1 Jumpei Ukita　村上 遥 *1 Haruka Murakami　浅谷 公威 *1 Kimitaka Asatani　森 純一郎 *1 Junichiro Mori

坂田 一郎 *1 Ichiro Sakata

*1東京大学 The University of Tokyo

In the task of supervised extractive summarization, the lack of training data is crucial issue. In this paper, we propose a multi-task learning architecture with document classification to solve the issue. Generally the contents of document label such as the category correspond to the sentences that should be extracted, therefore the learning on document classification contributes to the sentence extraction. The experiment shows that our model improves the accuracy of summarization especially in the case of few training data.

1. はじめに

近年，Web の発達と電子化された情報量の飛躍的な増大により，自動文書要約のニーズはより一層高まっている．本研究では重要文抽出における新規手法を提案する．重要文抽出とは，要約対象の文書 $D = \{s_1, s_2, \ldots, s_n\}$ (s_iは文書中の i 番目の文) が与えられた時に，抽出する文の数や文長を制約として要約に適当な文 s_i を抽出するタスクである．

参照要約（人手で作成された要約）を教師データとして重要文を判別し，抽出する教師あり重要文抽出は多くの研究がなされているが，一般に現実の文書には参照要約が少ないため，汎化性能が向上せず適用困難であることが多い．

本研究では，参照要約が少量の場合においても有効な重要文手法として，文書分類の学習により重要文抽出の学習をサポートするマルチタスク学習モデルを提案する．文書分類とは，与えられた文書がどのクラスに属するかを推定するタスクであり，カテゴリ分類や極性分類などが挙げられる．マルチタスク学習とは複数の推定タスク間で共通の特徴を有する場合，それらを同時に学習することで精度が向上する学習である．単一のタスクに過学習されず，一方のタスクが擬似的にもう一方のタスクの教師データとして機能することから，汎化性能が向上する．提案モデルでは，文書分類に各文の分散表現の加重平均を利用し，その重み付けに各文の抽出確率を用いる．即ち，抽出確率の高い文ほど文書分類に大きな影響を与え，文書分類の学習時には，分類に有効な文の抽出確率が高くなるように学習が行われる．

文書分類とのマルチタスク学習を重要文抽出に適用する妥当性と，参照要約が少量の場合における有効性の根拠について，決算短信からの重要文抽出を例に説明する．決算短信には当期の売上高や利益変化率がメタ情報として付与されており，それらの業績要因が記述されている．参照要約として，決算短信の内容が要約された記事（決算記事）を利用する場合，抽出されるべき重要文は業績の主要因を記述した文である．したがって，文書から業績を判定するタスクを考えた時，判定に大きな影響を与える文は同時に抽出されるべき文であり，文書分類の学習が重要文抽出をサポートすることが期待される．また，現実の短信には参照要約である決算記事が作成されていないものが多く，これらを文書分類の学習のみに利用することで，擬似的に教師データを増やすことが可能であると考えられる．

本論文では，決算短信と New York Times 紙における重要文抽出実験を行い，提案モデルの有効性を検証する．決算短信では売上高変化率及び純利益変化率の正負判定，New York Times 紙では記事のカテゴリ分類を文書分類タスクとして設定し，文書分類とのマルチタスク学習が精度向上に寄与するか検証する．このような参照要約が少なく，かつテキスト分類が可能な文書は，実験で用いる決算短信や記事以外にも，商品レビュー文やスポーツ記事など様々なものが存在する．したがって重要文抽出を現実文書に適用するにあたり，本研究で提案するアプローチは有用であるといえる．

連絡先: 磯沼大，東京大学，isonuma@ipr-ctr.t.u-tokyo.ac.jp

2. 関連研究

2.1 学習データの充足・補完に関する研究

学習データの充足・補完に関する研究として，他ドメインや他タスクの学習を通じて精度向上を図るマルチタスク学習や，学習データとして参照要約だけでなく，非言語情報を用いた要約や情報検索に関する研究が挙げられる．自然言語処理におけるニューラルネットワークを用いたマルチタスク学習モデルとして，Luong らは複数言語間の翻訳について，出力だけでなく入力も複数にしたマルチタスク学習を行い，用いたどの言語においても翻訳精度の向上を確認した [Luong 16].

非言語情報を用いた要約・情報検索として，Titov らはレビュー文について単語と評価値及びその属性との共起関係を抽出し，要約に用いた [Titov 08]．Liu らは情報検索において，クエリ分類とクエリに対する検索結果を同時に学習することで，出力される検索結果の精度向上を報告している [Liu 15].

2.2 分散表現を用いた文抽出

文を分散表現により表現し，重要文を抽出する手法として，Cao らは Recursive Neural Network を用いた手法を提案している [Cao 15]．Cao らは句構造解析によって単語やフレーズをノードにした木構造に文を変換し，参照要約との ROUGE 値を学習し，整数計画法による抽出判定を行っている．Cheng らは LSTM-RNN（Long Short-Term Memory Recurrent Neural

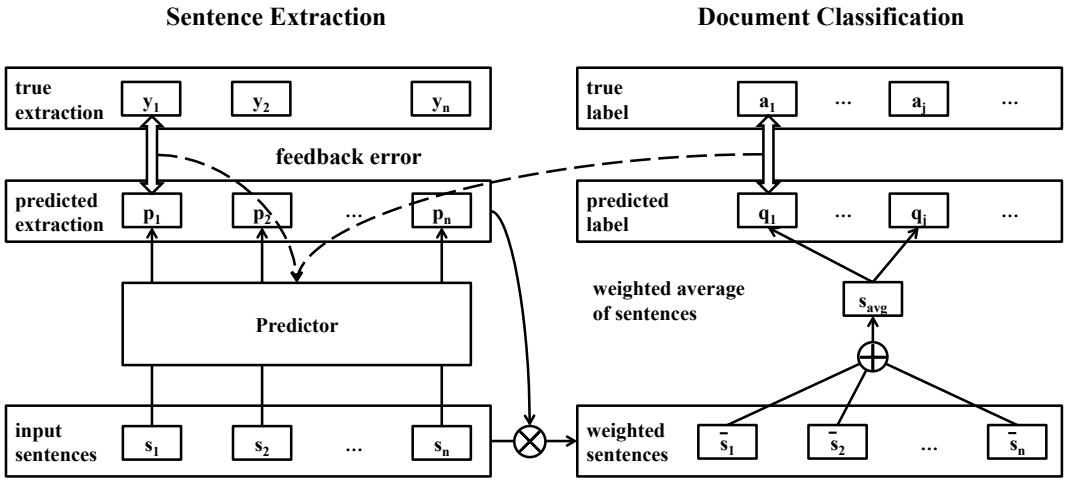

図 1: 提案手法の概要図

Network）Encoder-Decoder モデルを用いて文書全体の情報を元に各文の抽出判定を行っている [Cheng 16].

2.3 分散表現を用いたテキスト分類

Yang らは文書を文と単語の 2 階層の RNN によって表現し，各階層間にアテンション機構を適用し，文書分類を行っている [Yang 16]．Wang らは Aspect Sentiment Classification において，Aspect を分散表現によって表現し，それに基づいたアテンション機構により極性分類を行った [Wang 16].

2.4 本研究の位置づけ

文書分類及び重要文抽出は様々な既存研究が存在するが，文書分類を重要文抽出のサポートとして利用する手法は提案されていない．本研究ではこれをマルチタスク学習モデルによって実現すると共に，マルチタスク学習を効率的に進めるための新規のカリキュラム学習を提案する．

3. 提案手法

本節では，重要文抽出の学習に加え，文書分類の学習を同時に行うことで重要文抽出をサポートするマルチタスク学習モデルを提案する．図 1 に本手法の概要図を示す．図の左側は通常の文抽出学習部分であり，s_i は文書の i 番目の文の分散表現ベクトルであり，y_i は文 i の抽出ラベル，p_i は推定された文 i の抽出確率である．右側は本節で提案する文書分類による文抽出学習部分である．a_j は文書のラベル情報の一つで，$\bar{s}_i$ は推定抽出確率 p_i によって重み付けされた s_i である．$\bar{s}_i$ の要素和 s_{avg} は抽出確率による文の重み付け平均ベクトルであり，抽出確率の高い文の情報が多く含まれる．このベクトルから文書のラベルを推定し，真のラベルとの誤差を得る．抽出される確率の高い文がラベル推定に有効でない場合は誤差は大きくなる．この誤差を最小化するように文抽出の学習機を更新することで，ラベルの内容を反映する重要文を抽出する．

本研究では Cheng ら [Cheng 16] の重要文抽出モデルにマルチタスク学習を適用するが．この枠組みは分散表現を用いた一般的な重要文抽出モデルに汎用可能である．Cheng らは重要文抽出に LSTM-RNN を用いており，以下ではそれの簡易な説明と，文書分類とのマルチタスク学習及びそれを効率的に進めるためのカリキュラム学習の導入について説明する．

3.1 文の分散表現獲得

文中の単語の分散表現から文の分散表現を獲得する．単語 w_i の分散表現ベクトル x_i は，単語のカテゴリベクトル t_i，分散表現射影行列を $W_{emb} \in \mathcal{R}^{k \times |V|}$ とした時，以下の式 (1) で定義される．本研究では skip-gram により事前学習を行う．

$$x_i = W_{emb} \cdot t_i \tag{1}$$

CNN により複数の単語の分散表現を畳み込み，文の分散表現を獲得する．畳み込む単語の数を N とした時，畳み込みは以下の式 (2) で表される．ただし，$x_{i:i+N}$ は文の i 番目から $i+N-1$ 番目までの N 個の単語分散表現ベクトルを結合したベクトル，f は活性化関数である．全ての i に対する最大値を，文の分散表現の要素の 1 つとして用いる．

$$c_i = f(w_{conv} \cdot x_{i:i+N} + b_{conv}) \tag{2}$$

3.2 文書の隠れ層への埋め込みと重要文推定

LSTM-RNN Encoder-Decoder モデルにより文書を隠れ層へ埋め込み，各文の抽出確率を推定する方法について述べる．LSTM-RNN では t 番目の文の分散表現 s_t と直前の隠れ層 h_{t-1} から隠れ層 h_t が逐次更新される．(式 (3))

$$h_t = \mathrm{LSTM}(s_t, h_{t-1}) \tag{3}$$

Encoder 部分では s_1 から s_n まで LSTM-RNN に入力され，隠れ層が更新される．h_n には文書全体の情報が埋め込まれ，これをもとに各文の抽出確率を推定する．Decoder 部分の隠れ層 $\bar{h}_t$ は Encoder 部分と同様に LSTM によって以下の式により更新される．ただし Decoder 部分では直前の文 s_{t-1} が後述する推定抽出確率 p_{t-1} を乗じて入力される．即ち抽出確率の高い文は多くの情報が入力される．

$$\bar{h}_t = \mathrm{LSTM}(p_{t-1} \cdot s_{t-1}, \bar{h}_{t-1}) \tag{4}$$

$\bar{h}_t$ から，文 t の抽出確率 p_t は以下の式で得られる．: はベクトルの結合を表す演算子であり，Encoder 部分の h_t を結合することで，入力文 s_t の情報をより直接的に参照し，抽出確率が推定される．抽出確率上位 3 文を重要文として抽出する．

$$p_t = \sigma(W_y \cdot [h_t : \bar{h}_t] + b_y) \tag{5}$$

3.3 文書のラベル推定

推定された文抽出確率と文の分散表現をもとに文書のラベル推定を行う方法について説明する．ラベル j が a_j である確率 q_j は以下の式から推定される．

$$s_{avg} = \frac{\sum_t p_t \cdot s_t}{\sum_i p_i} \tag{6}$$

$$q_j = \sigma(w_a \cdot s_{avg} + b_a) \tag{7}$$

$s_{avg} \in \mathcal{R}^{lm}$ は p_t による各文 s_t の加重平均ベクトルである．$w_a \in \mathcal{R}^{lm}$ は重みベクトルで，$b_a \in \mathcal{R}$ はバイアス項である．

3.4 マルチタスク学習とカリキュラム学習の導入

マルチタスク学習では，文抽出の損失関数 E_y と文書ラベル j の推定に関する損失関数 E_{a_j} が共に最小になるようにモデルの各パラメータ θ を更新する．共に損失関数として負の対数尤度を用いる．$\lambda_\theta \|\theta\|$ は各パラメータに関する正則化項である．

$$E_y(\theta) = -\sum_{t=1}^{n} y_t \log p_t + (1-y_t)\log(1-p_t) + \lambda_\theta \|\theta\| \tag{8}$$

$$E_{a_j}(\theta) = -\log q_j \ + \lambda_\theta \|\theta\| \tag{9}$$

マルチタスク学習は文抽出と文書分類の学習を同時に行うため，学習が複雑になる．この対応策として，カリキュラム学習 [Bengio 09] を導入する．カリキュラム学習とは単純なモデルやデータに対する学習を初期に行い，徐々に複雑な学習を行うことで，複雑なモデルやデータにおける精度を高める学習方法である．

本手法では 2 種類のカリキュラム学習を導入する．1 つ目はデータセットを 3 分割し，難易度が低い順にデータを追加する方法である．文書分類の学習による文抽出は難易度が高いため，最初は参照要約が有る文書について文抽出学習のみを行い，文抽出判定を安定化させる．次に同一の文書について文書分類を行うデータを追加し，文抽出とラベル推定の同時学習を安定化させる．最後に参照要約が無い文書について，ラベル推定の学習のみを行うデータを追加し，文書分類が重要文抽出のサポートとして機能するよう学習する．

2 つ目はラベル推定の際に，推定文抽出確率 p_t と，真の文抽出ラベル y_t の双方を用いて文の加重平均を導出する方法である．式 (6) における p_t を，式 (10) に示す $\bar{p}_t$ で代替する．κ は文抽出確率と真の文抽出ラベルの混合比であり，学習が収束していない段階では κ を 0 近辺にすることで，文抽出とメタ情報推定の学習が互いに干渉せずに，出力層のパラメータのみが学習される．学習が進むにつれ κ を徐々に増加させることで，ラベル推定の学習が文抽出判定をサポートする．

$$\bar{p}_t = \kappa p_t + (1-\kappa) y_t \tag{10}$$

表 1: 手法毎の重要文抽出精度比較（%）

モデル	F-measure	Precision	Recall
LEAD	33.3	36.2	38.6
LREG	60.5	67.6	66.5
NN-SE	64.2	71.9	69.1
NN-MP	54.1	61.0	58.5
NN-MP-CL	65.7	74.3	70.2

4. 実験

4.1 決算短信と記事を用いた実験

本節では，決算短信を要約対象に，それをもとに作成された日本経済新聞社の決算記事を参照要約とした実験内容とその結果について説明する．訓練，検証，評価で用いた決算短信の数はそれぞれ 12,262 件，191 件，183 件である．訓練データ中の短信は記事が作成されていない短信 8,725 件を含み，それらは文書分類の学習のみに利用した．文書分類で用いるラベルとして，前年同期比売上高変化率と前年同期比純利益変化率を用いた．どちらも，値が正（増加）の場合 $a = 1$ を，値が負（減少）の場合 $a = 0$ とした 2 値ラベルを用いた．

実験では，短信の各文について抽出フラグが必要である．本実験では要約の自動評価指標 ROUGE-1 に閾値を設け，抽出フラグを付与した．評価データ 183 件の短信における人手で付与した抽出フラグに対する自動付与した抽出フラグの精度は，AUC において 88% だった．

実装の詳細について，単語の分散表現ベクトルの次元数は 200，LSTM の隠れ層の次元数は 400 を設定した．CNN では畳み込む単語の数は $\{1, 2, 3, 4, 5, 6\}$ の 6 種類を設定し，各単語数について 50 枚のフィルタを用いた．したがって文の分散表現ベクトルの次元数は $6 \times 50 = 300$ である．パラメータの学習は Adadelta を利用し，学習率の初期値は 10^{-6} である．単語の分散表現ベクトルの事前学習では，日本語版 Wikipedia の全記事を用いた．

4.2 決算短信を用いた実験結果

4.2.1 手法毎の重要文抽出精度比較

本実験では 5 つの手法について，精度比較を行った．LEAD は文書の初めの 3 文を抽出する方法である．LREG は文の文書中における位置や長さを特徴量としたロジスティック回帰による重要文抽出である．NN-SE はニューラルネットワークによる重要文抽出手法 [Cheng 16] であり，提案モデルとの差分はマルチタスク学習の適用有無である．これら 3 つは比較のためのベンチマークとして採用した．NN-MP は本研究で提案する文書分類とのマルチタスク学習を追加した手法であり，NN-MP-CL は NN-MP にカリキュラム学習を適用した手法である．手法毎の重要文抽出精度比較結果を表 1 に示す．

NN-SE に，マルチタスク学習を追加した NN-MP は，NN-SE より精度が大幅に下回り，LREG よりも精度が下回った．しかしカリキュラム学習を適用した NN-MP-CL では，NN-SE に対して精度の向上が確認された．通常の学習では文抽出と文書分類の学習が競合し精度が向上しないが，提案するカリキュラム学習を導入した場合，文抽出が安定化した後に文書分類を行うため双方の学習が競合せず，文書分類が精度向上に寄与したと考えられる．

4.2.2 参照要約の数を変化させた場合の比較

学習データ中の参照要約の数が 125 件，250 件，500 件，1000 件，2000 件の場合において，既存手法である NN-SE と提案手法である NN-MP-CL の重要文抽出精度を確認した．F 値による抽出精度比較を行った結果を表 2 に示す．参照要約の数が少ない場合，提案手法の既存手法に対する増加幅が特に大きいことが確認される．

表 2: 参照要約の数を変化させた場合の F 値の比較（%）

モデル	125	250	500	1000	2000
NN-SE	60.8	61.4	62.1	65.2	63.5
NN-MP-CL	62.7	62.5	63.5	65.8	64.3

表 3: 人手による手法の評価順位比較（%）

モデル	1 位	2 位	3 位	4 位	平均順位
LEAD	21.7	20.0	28.3	30.0	2.67
SE-ALL	20.0	28.3	26.7	25.0	2.45
NN-SE	31.7	21.7	16.7	30.0	2.57
NN-MP-CL	51.7	20.0	21.7	6.7	**1.83**

4.2.3 人手による手法の順位付け評価

各手法により短信から抽出した文について，人手による順位付け評価を行った．LEAD，EF，NN-SE，NN-MP-CL の 4 手法により短信から抽出された 3 文について，記事との合致度の観点から順位付けを依頼した．評価は東京大学大学院工学系研究科の研究員及び学生 6 名に依頼した．用いられた短信は 4.1 節で述べた評価に用いられた 183 件から，異なる手法で同一の文が抽出された短信を除いた上で，無作為に 20 件を抽出した．

各手法の順位分布とその平均値を表 3 に示す．提案手法である NN-MP-CL は最も順位が高くなり，特に 4 位と評価された件数は他のベンチマーク手法に比して非常に少ないことが確認される．提案手法の有用性が，人手による評価において検証された．

4.3 NYTAC を用いた実験

本節では New York Times Annotated Corpus（NYTAC）を用いた実験内容とその結果について説明する．NYTAC とは New York Times 紙の記事と人手で作成された要約が格納されたコーパスであり，各記事にはカテゴリなどのラベルが付与されている．本実験においても，参照要約の数を変化させた場合における精度検証を行った．検証，評価で用いた記事の数は共に 200 件であり，訓練では参照要約の数が 125 件，250 件，500 件それぞれの場合を用意した．また，参照要約が付与されていない 3000 件の記事を文書分類の学習のみに利用した．文書分類で用いるラベルとして，記事のカテゴリを用いた．カテゴリ数は 26 で，”Business”や”Arts”といったカテゴリが付与されている．

ハイパーパラメータは，単語の分散表現ベクトルの次元数以外決算短信を用いた実験と同一である．単語の分散表現ベクトルについては，Google News をコーパスとした事前学習済み分散表現ベクトルを利用した．次元数は 300 である．

評価では，推定抽出確率の上位 3 文を抽出判定し，抽出文と要約文との ROUGE-1 及び ROUGE-2 を算出し，比較した．比較結果を表 4 に示す．ROUGE-1, ROUGE-2 のどちらにおいても，参照要約の数が 250 の場合，提案手法の既存手法に対する精度の増加幅が最も大きかった．次いで 125，500 の順にどちらの指標においても増加幅が大きかった．NYTAC を用いた実験においても，参照要約の数が少ない場合，特に精度向上が確認された．

表 4: 参照要約の数を変化させた場合の ROUGE 値比較（%）

モデル	参照要約数	ROUGE-1	ROUGE-2
NN-SE	125	17.2	12.1
	250	16.8	11.3
	500	18.0	12.5
NN-MP-CL	125	18.1	12.7
	250	18.3	12.7
	500	18.5	12.9

表 5: 各手法で抽出成功した文のラベルとの合致率（%）

	合致	非合致	その他
NN-MP-CL	**85.7**	14.3	0.0
NN-SE	**40.0**	40.0	20.0

5. 考察

本節では精度向上が文書分類の学習によって実現したことを示す．決算短信を用いた実験において，提案手法である NN-MP-CL と既存手法である NN-SE について，抽出成功した文内容と，ラベルとして用いた純利益変化率や売上高変化率との合致率をまとめた結果を表 5 に示す．提案手法で抽出成功した文の 85.7%はラベルと合致していた一方，逆の場合だと 40.0%のみであった．文書分類の学習により業績を反映する文がより多く抽出判定され，精度向上に寄与したと考えられる．

6. 結論

本研究では，参照要約が少量の場合においても有効な重要文手法として，文書分類に各文の分散表現の加重平均を利用し，その重み付けに各文の抽出確率を用いることで，文書分類の学習が重要文抽出の学習をサポートするマルチタスク学習モデルを提案した．また，マルチタスク学習を効率的に進めるためのカリキュラム学習手法を新規に提案した．

決算短信と New York Times 紙を対象とした重要文抽出の精度評価実験では，参照要約の数が少ない場合に特に精度が向上することが確認された．

謝辞

本研究は，東京大学大学院工学系研究科技術経営戦略学専攻松尾豊特任准教授と，同学術支援専門職員である椎橋徹夫氏から多くのご助言を頂きました．厚く御礼を申し上げます．

参考文献

[Bengio 09] Bengio, Y., et al.: Curriculum learning, in *ICML*, pp. 41–48 (2009)

[Cao 15] Cao, Z., et al.: Ranking with Recursive Neural Networks and Its Application to Multi-Document Summarization., in *AAAI*, pp. 2153–2159 (2015)

[Cheng 16] Cheng, J., et al.: Neural Summarization by Extracting Sentences and Words, in *ACL*, pp. 484–494 (2016)

[Liu 15] Liu, X., et al.: Representation learning using multi-task deep neural networks for semantic classification and information retrieval, in *NAACL-HLT*, pp. 912–921 (2015)

[Luong 16] Luong, M., et al.: Multi-task sequence to sequence learning, in *ICLR* (2016)

[Titov 08] Titov, I., et al.: A Joint Model of Text and Aspect Ratings for Sentiment Summarization, in *ACL*, pp. 308–316 (2008)

[Wang 16] Wang, Y., et al.: Attention-based LSTM for Aspect-level Sentiment Classification, in *EMNLP*, pp. 606–615 (2016)

[Yang 16] Yang, Z., et al.: hierarchical attention networks for document classification, in *NAACL-HLT* (2016)