

サブワードフレーズ抽出を併用した文書分類

木村 優介^{1,a)} 駒水 孝裕^{2,b)} 波多野 賢治^{3,c)}

概要： 深層学習を用いた文書分類は従来の手法と比べて高精度を達成してきた。文書分類は固有表現抽出とのマルチタスク学習によりその精度を向上させることが可能であると報告されている。単語に基づいた特徴量を基本とする手法において、意味のまとまりとしてのフレーズは文書分類の精度に寄与することが知られている。一方で、深層学習においてサブワードは一般的に使われているが、単語とは異なり、意味のまとまりを表すとは限らない。そのため、サブワードで構成されるフレーズを文書分類に応用する研究はされてこなかった。そこで、本研究では文書分類の精度向上を目的に、サブワードフレーズ抽出と文書分類のマルチタスク学習を行う文書分類フレームワークを提案する。従来の単語に基づくフレーズ抽出手法が出現頻度を用いてきた点に倣い、本稿では高頻度なサブワードのフレーズをサブワードフレーズとして定義し抽出する。

キーワード： 文書分類, サブワードフレーズ, マルチタスク学習

1. はじめに

膨大な量の文書に対して効果的・効率的に情報アクセスするための技術の一つとして、文書分類の技術がある [1]。一般的には、あらかじめ人間が分類基準を定め、その基準にしたがって文書を分類し、その分類ごとにラベルを付与することをいい、この文書分類を特に教師あり文書分類と呼んでいる。ある発言に対してそれをヘイトスピーチと判定 [2] したり、日々発信される多くのニュース記事を社会面や経済面といった新聞記事のカテゴリに分類したりすることも文書分類の技術で行われている。

近年は深層学習技術が発展し、従来よりも高精度な文書分類が可能になった [3]。深層学習に基づく文書分類では、英語文書をトークナイザで単語より小さな単位であるサブワードに分割し、サブワードによって分割された文書を深層学習の入力データとして使用する。サブワードを用いることにより未知語の存在を最小限に抑えることができるため、利用したサブワードで文書を表現することができるこ

とによって、文書分類に限らず、Q&A タスクや翻訳タスクといった自然言語処理に関連するタスクでも高精度な処理を実現できている。

また、深層学習ではマルチタスク学習と呼ばれる、あるメインタスクの汎化性能を向上させるために別のサブタスクを用いる方法を適用することが可能であり [4]、文書分類タスクの一つである意図分類をメインタスク、固有表現抽出をサブタスクとしたマルチタスク学習で高精度な結果を得ることができたという報告もある [5,6]。

このことから、文書中からその文書をよく表現しているフレーズ（キーフレーズ）を抽出することができれば、マルチタスク学習を行うことで文書分類に活用できる可能性があると考えられる。なぜなら、深層学習技術が発展する以前から、文書分類を行うためにキーフレーズを抽出しようとする研究が数多く行われてきたからである [7,8]。これに深層学習技術を適用することを考えれば、サブワードのフレーズ（以後、サブワードフレーズと呼称する）の抽出をサブタスクとした文書分類を行うマルチタスク学習の方法を提案できれば、深層学習時代の新たな文書分類のフレームワークとなり得る可能性がある。このような方法による文書分類の提案は、人手によるアノテーションとして与えたキーフレーズを用いたり、固有表現やキーフレーズ抽出された、人間が認識可能なフレーズを用いて行われているが、本研究の提案のような、意味のまとまりを表さない可能性のあるものをフレーズとして扱ってはいない。

そこで、本研究では文書分類の精度向上を目的に、サブ

¹ 同志社大学大学院文化情報学研究科
Graduate School of Culture and Information Science,
Doshisha University
² 名古屋大学数理・データ科学教育研究センター
Mathematical and Data Science Center, Nagoya University
³ 同志社大学文化情報学部
Faculty of Culture and Information Science, Doshisha University
a) kimura@mil.doshisha.ac.jp
b) taka-coma@acm.org
c) khatano@mail.doshisha.ac.jp

ワードフレーズ抽出と文書分類のマルチタスク学習のフレームワークを提案する。我々の文書分類フレームワークはサブワードフレーズ抽出タスクのためのラベリングとマルチタスク学習で構成されている。まず、文書分類ラベルが付与された文書に対して教師なし抽出手法で得られたサブワードフレーズでラベリングを行う。次に、サブワードフレーズでラベリングされた文書データを用いて教師ありサブワードフレーズ抽出と文書分類のマルチタスク学習を行う。教師なしサブワードフレーズの抽出には、高頻度語は重要語とならないとするジップの法則を援用し、高頻度なサブワードフレーズに着目することにし、それらが文書分類モデル中で重要視されないようなサブワードフレーズ抽出をサブタスクとするマルチタスク学習を行う。

2. 関連研究

本節では、マルチタスク学習を用いた文書分類について紹介する。Zhang ら [9] は、文書分類におけるマルチタスク学習を以下の三種類に分類している。

- **Multi-Cardinality** : 異なるデータに対する同種のタスクをサブタスクにする。例えば、Liu らは、映画レビューの分類を複数のデータセットを使ってマルチタスク学習を行い、分類性能の向上を示した [10]。
- **Multi-Domain** : ドメインの異なる類似タスクをサブタスクにする。例えば、Zhang らは、本やDVDなどの商品に対するレビューの分類をサブタスクとして、映画レビュー分類を行うマルチタスク学習を行った [9]。
- **Multi-Objective** : 異なる目的を持つタスクをサブタスクにする。例えば、Zhang らは、ニュース記事の分類と質問の種類分類をサブタスクとして、映画レビュー分類を行うマルチタスク学習を行った [9]。

これらの研究は、複数の教師あり学習を組み合わせることで、その有効性を示している。しかし、一般に、教師データを作成することは人的・金銭的コストが高いため、より低コストな手法が望まれる。

加えて、分類問題とはやや異なるが、ニュース推薦においてマルチタスク学習を利用している研究も存在する [11]。Bi ら [11] は、ニュース推薦タスクに対して、ニュース記事の分類と固有表現抽出をサブタスクとしたマルチタスク学習により、ニュース推薦の性能を向上させた。このような研究により、同一データセット内での異なるタスクをサブタスクとしたマルチタスク学習が有効であることが示された。上述の問題同様に、Bi らの研究も教師データに依存したサブタスクを利用しており、依然としてコストの高い方法である。

また、1つの文書が複数のラベルに分類されるマルチラベル分類問題において、同一データセット内のラベル間の関係を利用したマルチタスク学習も提案されている。Zhang ら [12] は、ラベル間の関係を学習するサブタスクを設計す

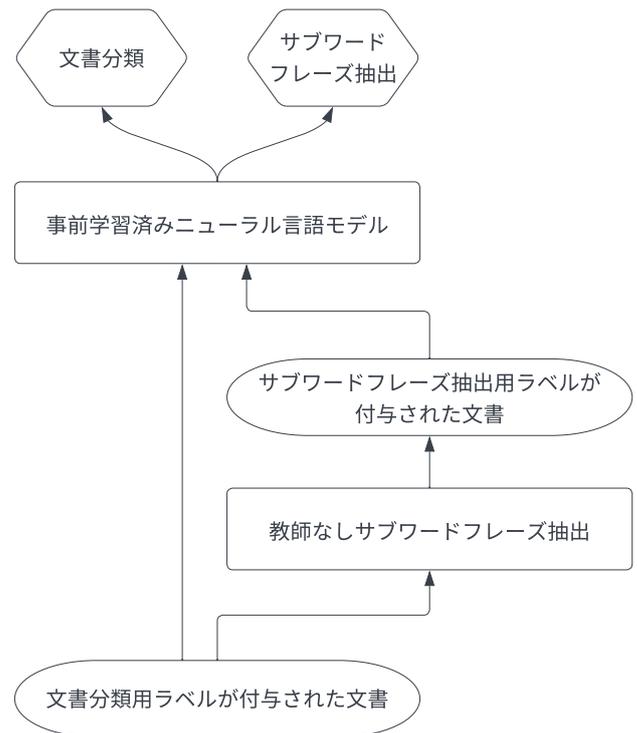


図1 サブワードフレーズ抽出を併用した文書分類フレームワーク

ることで、分類性能の向上を示した。

本研究は、機械的にサブタスクの教師データを作成することで、上述のコストを大幅に削減したマルチタスク学習に基づく文書分類を提案する。また、サブワードに基づいたサブタスクの設計を試みた研究は、著者らの知るところでは存在しない。したがって、本研究は、(1) 低コストなサブタスク設計と (2) サブワードフレーズの導入、という2点において、マルチタスク学習を用いた文書分類の新しい方法に取り組むものである。

3. 提案手法

本研究では、サブワードフレーズ抽出と文書分類を行う文書分類フレームワークを提案する。図1で示す通り、本研究で提案する文書分類フレームワークはラベリングとマルチタスク学習の2段階に分けられている。まず、教師なし抽出手法で得られたサブワードフレーズのラベルを文書に付与する。次に、文書分類ラベルとサブワードフレーズのラベルが付与された文書を用いて、文書分類とサブワードフレーズ抽出のマルチタスク学習を行う。本節では、それぞれの段階を3.1項、3.2項で詳しく説明する。

3.1 サブワードフレーズのラベリング

サブワードフレーズのラベリングの段階では、まず教師なし抽出手法によってサブワードフレーズを抽出する。一般のフレーズ抽出に倣えば、キーフレーズ抽出など様々な

関連技術が存在するが、本稿では、単純な手法である頻度に基づくサブワードフレーズを用いて、本フレームワークの有効性を確認する。一般的に、単語の出現頻度はジップの法則に従っており、また、サブワードフレーズもジップの法則に従っていることが報告されている [13]。そのため、高頻度なサブワードフレーズも高頻度な語と同様に文書の意味を表さない語である可能性がある。本研究では、高頻度なサブワードフレーズを抽出対象とすることで、それらが文書分類モデルに重要視されないようにする。

高頻度なサブワードフレーズを抽出する手順は、以下の2ステップから構成される。

- (1) トークナイザを用いたサブワード分割
- (2) Byte Pair Encoding (BPE) を用いた高頻度なサブワードフレーズの抽出

本研究では、トークナイザに事前学習済みニューラル言語モデルで利用されているものを用いる。また、元来の BPE は、文書を1文字単位に分割後、単語内の隣接頻度に基づいて高頻度な文字同士から結合を行う。この手順を指定した語彙サイズになるまで繰り返すことで、文書を分割する手法である [14, 15]。本研究では、高頻度なサブワードフレーズを取得するために BPE のアルゴリズムをサブワード単位で実行する。アルゴリズムを変更した BPE を用いて得られたサブワードフレーズを高頻度なサブワードフレーズとする。BPE は、指定された語彙数になるまで、共起しやすい文字を結合するものである。したがって、高頻度なサブワードフレーズの抽出数がコントロールできる点において、抽出数が文書分類に与える影響を調査する際に有用な性質である。

また、上述の方法では、高頻度なサブワードフレーズが異なる文書分類クラスに出現する。したがって、特定のクラスに有効なサブワードフレーズを区別することができない。これに対し、本研究では、クラスごとに高頻度なサブワードフレーズを n 語取得し、どのクラスに属するフレーズかが区別できる方法を用いてラベリングを行う。また、各クラスで抽出対象とする n 語の決め方は、3.2 項で説明するマルチタスク学習を行ったモデルに対し、検証データでの分類性能が最も高いサブワードフレーズ数を採用する。

高頻度なサブワードフレーズを文書にラベリングする作業では、固有表現抽出タスクで広く使われている Inside-outside-beginning (IOB2) タグに倣って、ラベリングする [16]。IOB2 タグを用いたラベリングは、高頻度なサブワードフレーズに含まれるサブワードについて、最初のサブワードに B タグを付与し、それ以外のサブワードには I タグを付与し、高頻度なサブワードフレーズに含まれないサブワードには O タグを付与する。あるクラスにのみ出現する高頻度なサブワードフレーズを区別するために、クラスラベルを IOB2 タグ付けの方法に従って付与する。具体的には、B- i , I- i ($1 \leq i \leq$ データセットのクラス数) のよ

うにラベリングを行う。また、2クラス以上に出現する高頻度なサブワードフレーズには、複数クラスでの出現を表すラベル 0 を特別に使用する。具体的には、B-0, I-0 タグをラベリングする。

3.2 文書分類とサブワードフレーズ抽出のマルチタスク学習

まず、文書分類は事前学習済みニューラル言語モデルに文 X を入力して得られる [CLS] トークンの埋め込み表現を使用して式 (1) に従って分類を行う。

$$Pr(c | X_{cls}) = \text{softmax}(W_{dc}^T \cdot \mathbf{x}_{cls}) \quad (1)$$

ただし、 W_{dc} は文書分類タスク固有のパラメータ行列、 \mathbf{x}_{cls} は [CLS] トークンの文脈依存ベクトル、 c はクラスを表す。

高頻度なサブワードフレーズ抽出は、事前学習済みニューラル言語モデルに文 X を入力した後に得られる各トークンの埋め込みをもとに式 (2) に従って抽出を行う。

$$Pr(c | X_{token}) = \text{softmax}(W_{sb}^T \cdot \mathbf{x}_{token}) \quad (2)$$

ただし、 W_{sb} は高頻度なサブワードフレーズ抽出タスク固有のパラメータ行列を表し、 \mathbf{x}_{token} はあるトークンの文脈依存ベクトルを表す。高頻度なサブワードフレーズ抽出のクラス c は、数値に変換した IOB2 タグを表している。

2つのタスクの損失関数はいずれも各クラスへの2値分類と考え、式 (3) の Binary Cross Entropy*1を採用する。

$$L = \frac{1}{N} \sum_{i=1}^N -w_i [y_i \cdot \log x_i + (1 - y_i) \cdot \log(1 - x_i)] \quad (3)$$

ただし、 N はバッチサイズ、 $y \in \{0, 1\}$ は真のラベル、 x はモデルの出力確率を表す。

文書分類と高頻度なサブワードフレーズ抽出のマルチタスク学習は両タスクの損失値 L_{dc} , L_{sb} の合計値をモデル全体の損失値 L_{joint} として学習を行う。

$$L_{joint} = L_{dc} + L_{sb} \quad (4)$$

また、optimizer には Python ライブラリ Transformers*2の AdamW を採用する [17]。

4. 評価実験

本節では、比較手法、評価実験の方法と評価結果および考察について説明する。

*1 <https://pytorch.org/docs/stable/generated/torch.nn.BCELoss.html#torch.nn.BCELoss>

*2 https://huggingface.co/docs/transformers/main_classes/trainer

4.1 比較手法：BertGCN

BertGCN は事前学習済みニューラル言語モデルと GNN のインダクティブ学習を組み合わせた文書分類の state-of-the-art である [18]. BertGCN は, TextGCN [19] に倣い, 文書と単語, 単語と単語の共起関係からなるグラフを構築し, 事前学習済みの BERT から得られたベクトルをノードの初期ベクトルとして扱い, GCN (Graph Convolutional Network) を用いてノードのベクトルを更新することで文書ノードの最終的な埋め込み表現を獲得する. GCN から得られた各文書の埋め込み表現をソフトマックス関数に入力することで分類を行う. また, [18] では, BERT モデルの出力を全結合層とソフトマックス関数を用いて分類した結果と統合することで, より分類性能が向上することを報告している. 具体的には, 前述の GCN から得られた予測確率 Z_{GCN} と BERT を用いた分類器から得られた予測確率 Z_{BERT} の線形和 (式 (5)) を新たな予測確率としている.

$$Z = \lambda Z_{GCN} + (1 - \lambda) Z_{RoBERTa} \quad (5)$$

ただし, λ は GCN の分類に対する重みである.

BertGCN は, 事前学習済みニューラル言語モデルとして, BERT 以外の手法を用いることが可能で, [18] では, RoBERTa が最も良い性能を示した. そのため, 本稿の評価実験においては, RoBERTaGCN を比較手法とする.

4.2 評価方法

本研究では下記のデータセットを用いる.

- **20NG (20 Newsgroups)**: ニュース文書を 20 種類のカテゴリに分類したデータセット
- **R8**: ニュース記事の文書分類データセット Reuters-21578*³の一部であり, カテゴリ数を 8 個に限定したデータセット
- **R52**: Reuters-21578 の一部であり, カテゴリ数を 52 個に限定したデータセット
- **OHS (Ohsumed)**: 1991 年の医療抄録を MESH カテゴリとよばれる 23 種類の医学分野の概念に分類するデータセット
- **MR (Movie Review)**: 映画レビューがポジティブかネガティブかの 2 値に分類するデータセット

データセットの統計情報を表 1 に示す. 同表が示すように, クラス数やクラスごとのデータ数のばらつき (クラスデータ数 STD; クラス内データ数の標準偏差) が異なるデータセットを実験に用いており, 提案手法の得手不得手を分析可能なデータセットを用いている.

評価指標は, 式 (8) の F 値を基本に, マクロ平均 F_{macro} , マイクロ平均 F_{micro} , 重み付き平均 $F_{weighted}$ を用いて比較する.

*³ Reuters-21578 <http://www.daviddlewis.com/resources/testcollections/reuters21578/>, 閲覧日: 2022 年 8 月 4 日

表 1 データセットの統計情報: クラスデータ数 AVG およびクラスデータ数 STD はクラス内データ数の平均と標準偏差を表す.

	20NG	R8	R52	OHS	MR
訓練データ数	10,183	4,937	5,879	3,022	6,398
検証データ数	1,131	548	653	335	710
テストデータ数	7,532	2,189	2,568	4,043	3,554
クラス数	20	8	52	23	2
クラスデータ数 AVG	942	959	175	321	5,331
クラスデータ数 STD	193	583	268	139	1,161

$$P = \frac{TP}{TP + FP} \quad (6)$$

$$R = \frac{TP}{TP + FN} \quad (7)$$

$$F = 2 \times \frac{P \times R}{R + P} \quad (8)$$

ただし, TP は分類モデルが推測し正解したラベル数を表し, FP は不正解したラベル数を表す. また, FN は分類モデルの推測結果に出力されていない真のラベル数を表す. マクロ平均およびマイクロ平均は, クラスレベルおよびインスタンスレベルの F 値の平均値である. 重み付き F 値 $F_{weighted}$ は, クラスごとのインスタンス数のばらつきを考慮した F 値であり, 以下の式のように計算する.

$$F_{weighted} = \sum_{c \in C} r_c \cdot F_c \quad (9)$$

ただし, C はクラス集合を, r_c はクラス $c \in C$ に属するインスタンスの割合を, F_c はクラス c における F 値を表す.

この評価実験の環境は, CPU として AMD Ryzen Threadripper PRO 3975WX 3.5GHz, 主記憶として 256GB RAM, GPU として NVIDIA RTX A6000 $\times 3$, OS に Ubuntu 20.04 LTS を搭載したワークステーション上で行った. 本研究の評価実験では比較対象として, 上述の RoBERTaGCN を用いる. また, マルチタスク学習の効果を確認するために Huggingface 社の RoBERTa-base*⁴ に全結合層を加えた分類モデルを用いた. エポック数は 100, バッチサイズは 128, 最大トークン長は 256 で実験を行った. 本研究の提案手法の高頻度サブワードフレーズの数を表すハイパーパラメータ n は, $\{10, 100, 1000, 10000\}$ から検証データに対して最も高い分類性能を出した値を用いる.

4.3 結果, 考察

RoBERTa-base, RoBERTaGCN との比較評価の結果, マクロ F 値, マイクロ F 値, 重み付き F 値はそれぞれ表 2, 表 3, 表 4 のようになった. 全体的に見ると, RoBERTaGCN の分類性能が良く, 提案手法は, R52 と Ohsumed において, F_{micro} と $F_{weighted}$ が高性能を示している. 一方で,

*⁴ <https://huggingface.co/roberta-base>

表 2 F_{macro}

モデル	20NG	R8	R52	OHS	MR
RoBERTa-base	0.812	0.946	0.616	0.541	0.878
RoBERTaGCN	0.892	0.981	0.931	0.716	0.881
提案手法	0.848	0.939	0.844	0.600	0.876

RoBERTa-base と提案手法を比べるとほぼすべてのデータセットと評価指標で、提案手法の方が同等かそれ以上の性能を示している。特に、クラス数が多い R52 と OHS において、その差は大きい。このことから、提案手法がクラス数の比較的多い多クラス分類問題において、有効である可能性が示唆される。また、R52 はクラスごとのデータ数のばらつきが非常に大きく（表 1 のクラスデータ数 AVG とクラスデータ数 STD より）、不均衡なデータである。この R52 において、その差が顕著となることから、提案手法が不均衡な多クラス分類問題において、有効である可能性もまた示唆される。

表 3 の結果から、本研究の提案手法はマイクロ F 値において、クラスが不均衡なデータである R52, OHS で最も高いことを分かった。また、表 4 の重み付き平均 F 値においても本研究の提案手法は R52 で最も高い分類性能を達成した。以上から、本研究の提案手法は不均衡なデータでインスタンス数が多いクラスに対して予測が正しくできていると考えられる。また、分類するクラス数が増えるほど本研究の提案手法は他手法より分類性能が高い傾向にある。

次に、クラスごとに得られた高頻度なサブワードフレーズから考察を行う。表 5 はデータセット R52 で得られたクラス別高頻度なサブワードフレーズの上位 10 件である。サブワードフレーズに記載されている記号 “G” は RoBERTa-base のトークナイザが付与する記号であり、空白区切りで分割した際の単語の開始位置を示す。表 5 を見ると、クラス “bop” (BOP; 国際収支) と “orange” (オレンジ) には money に関係するサブワードフレーズが多い。例えば、bop クラスには、money supply や business loan などが高頻度なサブワードフレーズとして出ており、国際収支に関する特徴を捉えられている。同様に、orange クラスには、consumer prices や consumer price index (消費者物価指数) のような用語が出現している。一方で、複数クラスで共通するサブワードフレーズには別のラベルを付与したため、汎用なサブワードフレーズが選出されている。これらのことから、それぞれのクラスの細かな差異を表すことが可能なフレーズを抽出できたと考えられる。

しかし、本研究の問題点としてクラス別に高頻度なサブワードフレーズを n 語抽出した後、2 クラス以上で共通するサブワードフレーズに分類したため、データセット R52 の場合、全クラスの内 25% のクラスに属する高頻度なサブワードフレーズは 0 個になった。これは、サブワードフ

表 3 F_{micro}

モデル	20NG	R8	R52	OHS	MR
RoBERTa-base	0.818	0.978	0.765	0.674	0.878
RoBERTaGCN	0.875	0.923	0.654	0.555	0.881
提案手法	0.844	0.977	0.963	0.698	0.876

表 4 $F_{weighted}$

モデル	20NG	R8	R52	OHS	MR
RoBERTa-base	0.819	0.978	0.763	0.668	0.878
RoBERTaGCN	0.891	0.980	0.924	0.708	0.882
提案手法	0.855	0.977	0.961	0.696	0.876

レーズがクラス間で共通する度合いを考慮せずにした結果である。したがって、今後はクラス間で共通して出現する度合いを考慮したラベリングを実現することで、一定のクラス間に共通して出現するサブワードフレーズを捉えることが可能となる。

5. おわりに

従来の文書分類のためのマルチタスク学習がサブタスクにおいて、人的・金銭的成本のかかる教師データの作成が必要であるのに対し、本研究では教師なしのサブワードフレーズ抽出を用いて教師データが作成したサブタスクを用いるマルチタスク学習フレームワークを提案した。本稿では、教師なしで抽出可能なサブワードフレーズの中でも最も単純な頻度に基づく方法を採用した。また、高頻度なサブワードフレーズをクラス別に抽出することで、提案マルチタスク学習フレームワークが文書分類に与える影響を実験した。評価実験の結果、不均衡なデータでインスタンス数が多いクラスに対して高精度での分類が可能になった。

今後の方針は、より高度なサブワードフレーズ抽出手法を試みることで、複数種類のサブワードフレーズ抽出タスクを本研究のフレームワークに加え、複合タスクとすることによる性能向上の可能性を模索し、複数タスクに異なる重みを与えることによって文書分類精度が向上するように調整する手法を提案する。

謝辞

本研究は日本学術振興会科学研究費助成事業基盤研究 (A) 19H01138 および基盤研究 (B) 19H04218 と 21H03555, JST 科学技術イノベーション創出に向けた大学フェローシップ創設事業 JPMJFS2145 の助成を受けて遂行された。ここに記して謝意を表す。

参考文献

- [1] Qian Li, Hao Peng, Jianxin Li, Congying Xia, Renyu Yang, Lichao Sun, Philip S. Yu, and Lifang He. A survey on text classification: From traditional to deep learning.

表 5 R52 のクラス別高頻度なサブワードフレーズの上位 10 語の一例 (n=100)

	2 クラス以上で共通するサブワードフレーズ	クラス名		
		'bop'	...	'orange'
1	Ġre uter	Ġmoney Ġsupply		Ġconsumer Ġprices
2	Ġfe b	Ġmln Ġdlrs Ġbillion Ġdlrs		Ġconsumer Ġprice Ġindex
3	Ġl me	Ġweek Ġended		Ġstatistics Ġinstitute
4	Ġjan uary	Ġborrow ings		Ġrose Ġpct Ġfebruary
5	Ġfeb ruary	Ġbusiness Ġloans		Ġpct Ġmarch
6	Ġspokesman Ġsaid	Ġmoney Ġsupply Ġgrises		Ġcost Ġliving
7	Ġp ct	Ġfed Ġsays		Ġpct Ġcompared
8	Ġm ln	Ġbillion Ġdlrs Ġbillion Ġdlrs	-	
9	Ġseason ally	Ġweek Ġfed Ġsays	-	
10	Ġseasonally Ġadjusted	Ġpct Ġbillion	-	

- ACM Trans. Intell. Syst. Technol.*, Vol. 13, No. 2, pp. 31:1–31:41, 2022.
- [2] Ona de Gibert, Naiara Pérez, Aitor García Pablos, and Montse Cuadros. Hate speech dataset from a white supremacy forum. In *Proceedings of the 2nd Workshop on Abusive Language Online, ALW@EMNLP 2018, Brussels, Belgium, October 31, 2018*, pp. 11–20, 2018.
- [3] Shervin Minaee, Nal Kalchbrenner, Erik Cambria, Narjes Nikzad, Meysam Chenaghlu, and Jianfeng Gao. Deep learning-based text classification: A comprehensive review. *ACM Comput. Surv.*, Vol. 54, No. 3, 2021.
- [4] Rich Caruana. Multitask learning. *Machine Learning*, Vol. 28, No. 1, pp. 41–75, 1997.
- [5] Chaochen Wu, Guan Luo, Chao Guo, Yin Ren, Anni Zheng, and Cheng Yang. An attention-based multi-task model for named entity recognition and intent analysis of chinese online medical questions. *Journal of Biomedical Informatics*, Vol. 108, p. 103511, 2020.
- [6] Alberto Benayas, Reyhaneh Hashempour, Damian Rumble, Shoaib Jameel, and Renato Cordeiro De Amorim. Unified transformer multi-task learning for intent classification with entity recognition. *IEEE Access*, Vol. 9, pp. 147306–147314, 2021.
- [7] Yongzheng Zhang, Nur Zincir-Heywood, and Evangelos Milios. Narrative text classification for automatic key phrase extraction in web document corpora. In *Proceedings of the 7th Annual ACM International Workshop on Web Information and Data Management, WIDM '05*, p. 51–58, New York, NY, USA, 2005. Association for Computing Machinery.
- [8] Fabrizio Sebastiani. Machine learning in automated text categorization. *ACM Comput. Surv.*, Vol. 34, No. 1, p. 1–47, mar 2002.
- [9] Honglun Zhang, Liqiang Xiao, Yongkun Wang, and Yao-hui Jin. A generalized recurrent neural architecture for text classification with multi-task learning. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, pp. 3385–3391, 2017.
- [10] Pengfei Liu, Xipeng Qiu, and Xuanjing Huang. Recurrent neural network for text classification with multi-task learning. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016*, pp. 2873–2879, 2016.
- [11] Qiwei Bi, Jian Li, Lifeng Shang, Xin Jiang, Qun Liu, and Hanfang Yang. MTRec: Multi-task learning over BERT for news recommendation. In *Findings of the Association for Computational Linguistics: ACL 2022*, pp. 2663–2669, Dublin, Ireland, May 2022. Association for Computational Linguistics.
- [12] Ximing Zhang, Qian-Wen Zhang, Zhao Yan, Ruifang Liu, and Yunbo Cao. Enhancing label correlation feedback in multi-label text classification via multi-task learning. In *Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, Online Event, August 1-6, 2021*, pp. 1190–1200, 2021.
- [13] 木村優介, 駒水孝裕, 波多野賢治. ストップフレーズ抽出を併用した文書分類. 第 14 回データ工学と情報マネジメントに関するフォーラム予稿集, 2022.
- [14] Philip Gage. A new algorithm for data compression. *C Users J.*, Vol. 12, No. 2, p. 23–38, feb 1994.
- [15] Rico Sennrich, Barry Haddow, and Alexandra Birch. Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1715–1725, Berlin, Germany, August 2016. Association for Computational Linguistics.
- [16] Lance Ramshaw and Mitch Marcus. Text chunking using transformation-based learning. In *Third Workshop on Very Large Corpora*, 1995.
- [17] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019.
- [18] Yuxiao Lin, Yuxian Meng, Xiaofei Sun, Qinghong Han, Kun Kuang, Jiwei Li, and Fei Wu. BertGCN: Transductive text classification by combining GNN and BERT. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pp. 1456–1462, Online, August 2021. Association for Computational Linguistics.
- [19] Liang Yao, Chengsheng Mao, and Yuan Luo. Graph convolutional networks for text classification. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pp. 7370–7377, 2019.