機械学習を用いたデマ検知アルゴリズムの評価

QU　LIJING[†1]　成　凱[†1]

概要：インターネット上で自由に発信された情報の中で人々に役立つ情報がありながら、政治的・利益的な意図を持って拡散されるデマ情報も含まれており、社会問題となっている。デマ対策の一環として、デマ検知を素早く行う必要がある。本研究では、デマの特徴や発生原因等を考察し、機械学習の分類アルゴリズムによるデマ検知の有効性を評価する。実験で中国 SNS となる Weibo に関するデータセットを使い、SVM 分類器と Naive Bayes 分類器を評価した。SVM 分類器では線形分類モデル Linear-SVC と非線形分類モデル Rbf-SVC があるが、予測精度が予想より低かったため、Grid Search という手法で Hyperparameter を調整することで Rbf-SVC の予測精度を改善した。実験結果により、SVM 分類器が Naive Bayes 分類器より優れ、特に Rbf-SVC モデルの適合率が 9 割となり、デマ検知に有効であることが分かった。

キーワード：デマ検知、機械学習、テキスト分類、SVM、Naive Bayes

Evaluation of Machine Learning Based Rumor Detection Algorithms

Lijing QU[†1]　Kai CHENG[†1]

Abstract: Rumor information transmitted on the Internet becomes the social issues that required related measures. As part of measures against rumors, it is necessary to detect rumors quickly. In this paper, we consider the characteristics and causes of rumor and evaluate the effectiveness of rumor detection by machine learning classification algorithms. In the experiments, we evaluated the SVM classifier and the Naive Bayes classifier using a dataset related to Weibo, one of the Chinese SNS. The SVM classifier includes a linear classification model Linear-SVC and a non-linear classification model Rbf-SVC, but the prediction accuracy was lower than expected, so the prediction accuracy of Rbf-SVC was improved by adjusting the Hyperparameter with Grid Search approach. The results showed that the performance of the SVM classifier is better than the Naive Bayes classifier. Additionally, the 90% precision of Rbf-SVC model identified the rumor detection effectiveness.

Keywords: Rumor detection, Machine learning, Text classifier, SVM, Naive Bayes

1. はじめに

インターネットの発達により、SNS（Social Networking Service）で誰でもどこでも自由に発信ができるようになっている。昔から情報を入手する手段としては主にテレビと新聞などであるが、携帯の普及に従って、年齢、性別、収入と関わらず子供から高齢者まで SNS で情報を受け入れることができていた。近年から、人々はますます情報を受信する方から情報を発信する、共有する方に転変した。テキスト情報の編集や画像とビデオの加工は簡単になっていたので、日常を記録することが時代のファションになる。

大きな事故や地震などの災害が発生した場合は様々な情報が出てきた。しかし、その情報の中にデマ情報も含まれている。情報はデマかデマでないかの判別は難しい。また、デマを拡散することは起こしやすい。特に、高齢者は認識力の低下なのでソーシャルメディア上知らずにデマを拡散しまったことも多い。

2019 年年末から新型コロナウイルスが世界中範囲で流行されているが、SNS で真偽不明の情報がインターネット上で拡散していた。デマの拡散による大きな影響を与えられた。例えば、「日本で販売されているトイレペーパーの原材料は中国に依存しているため、近いうちに入手できなくなる」というデマが SNS で拡散した。この情報が拡散されていたことで日本のトイレペーパーが買い占める動きが広まっていた。香川県高松市にあるスーパーは「店舗の関係者が新型コロナウイルスを感染した」というデマがインターネット上に書き込まれた結果、系列の 7 店のうち 6 店は売り上げが上がったが、デマが流行された店だけ 10～15%ほど売り上げが減少した。これで、デマの伝播により、民衆の恐慌と利益の損失を引き起こすことができる。

政治的・利益的な意図を持って拡散されるデマ情報も含まれており、社会問題となって対策が求められている。この社会問題を解決するために、専門家や各関連部門から真相を明らかにすることが一番重要である。しかし、現在の

1 九州産業大学理工学部
Faculty of Science and Engineering
Kyushu Sangyo University

デマの検知は主に人力的に行うので、デマを拡散される時点から真相を明らかするまでは時間がかかる。

そこで、既存な実証されたデマデータを機械学習の教師データとして、大量な学習をした後、SNS 上に乗っている疑わしい情報の真偽を自動的に判別できる開発が期待される。

本研究では、デマの特徴や発生原因等を考察し、機械学習の分類アルゴリズムによるデマ検知の有効性を評価する。また、分類器の実験結果を比較し、分類器の違いによる分類精度への影響を確認する。

2. 関連研究

2.1 情報拡散の分析についての研究

大きな事故や地震などの災害が発生した時点で情報拡散に関る研究の一つに、安田氏[1]は震災時におけるソーシャルメディア上の情報拡散行動および情報拡散の状況の分析を行った。ソーシャルメディアを利用する個々のユーザの影響力と、そのユーザの持つ情報の正確さや知識量は独立であるにも関わらず、一部の人々が著しい情報伝播力を持つことが、不正確な情報が膨大に拡散する重要な要因と指摘した。

情報拡散の経由に関する研究も行われている。榎ら[2]はソーシャルメディアの友人関係のネットワークを用いることによって、実際にどのような経由で情報が流れたのかを推定する手法を提案した。また、情報が他のユーザへ段々と広けることができているユーザほど、高い値が付与されるようにユーザ段層ランクつけをおこなった。

デマ情報拡散の様子をモデル化した研究も行われている。白井ら[3]は SNS でのデマ情報の拡散の様子をモデル化した。SIR モデルの拡張を行い、デマ情報と訂正情報両方が伝播するモデルを構築し、早急にデマの拡散を収束される方法についての検討を行った。

2.2 機械学習におけるテキスト分類についての研究

フェイクニュースの分類に関する研究が行われている。岡山ら[4]はフェイクニュース分類器を用いたロコミサイトのレビューを分析した。Fake News Challenge のデータを教師あり学習によって学習することで、テキスト表現の特徴を学習し、分類器を作成する。さらに、ニュース以外のテキストとしてロコミサイト Yelp のレビューはフェイクかフェイクでないかを分類し、結果を分析する。廣瀬ら[5]は SNS 上で拡散するニュース説明文を自動的に生成するために、新聞社の記事中から拡散に寄与する重要文を RNN（Recurrent neural network）を用いて自動的に選択する仕組みの開発を行った。

大学の連絡通知の分類に関する研究も行われている。黒木ら[6]は機械学習を用いて連絡通知を自動分類し、重要な連絡通知を見逃すことを防ぐことを目的とする。前処理としてテキストデータの形態素解析を行い、Bag-of-Words モデルを適用し、単語の出現頻度による重み TF-IDF を計算する。最後に機械学習のアルゴリズムを訓練し、分類器を作成し、テストを行う。

デマ情報の分類に関する研究では、劉ら[7]は中国語メディア Weibo を例として、収集したデマについて意味分析とタイミング分析など色々な角度から分析した。宋ら[8]は新しいデマ検知モデル CED（Credible Early Detection）を開発した。彼らは三つのデータセットで実験を行った結果、予測する時間は大幅に短縮した。

中国 SNS の一つとする Weibo の情感傾向性、伝播過程およびユーザ情報に関する研究も行われている。Mao ら[9]から、分類特徴を抽出し、J48 分類器を用いた DTEC（Decision tree ensemble classifier）と SVM を用いた SVMEC（support vector machine ensemble classifier）を採用した。

3. デマ検知アルゴリズム

3.1 デマ情報について

辞書によるデマの定義は以下の表のように示す。

表 1　デマの定義

出典	定義
デジタル大辞泉	政治的な目的で、意図的に流す扇動的かつ虚偽の情報。 事実に反するうわさ。流言飛語。
百科事典マイペディア	語源的には扇動政治家が民衆を扇動するために用いた宣伝の意。現在では，権力機構が行う虚偽宣伝を本義とし、嘘、噂、流言などを意味する場合もある。
日本国語大辞典	政治的な目的で相手を誹謗し、相手に不利な世論を作り出すように流す虚偽の情報。また、社会情勢が不安な時などに発生して、人心を惑わすような憶測や事実誤認による情報。 単なる悪口や根拠のないうわさ話。

表 1 のように、デマは「虚偽の情報」であると強調した。事実に反する、根拠がないことを指す。頻繁に出現したデマは政治ニュース、社会ホットスポット問題、生活常識と有名人への誹謗などの四つの分類をまとめた。

これからなぜデマを信じている人が多いかとなぜデマを広く拡散するかについて述べる。人間は初めて情報を見る時本能で情報を質疑せずに受ける。その情報は嘘だろうかと思った時に専門的な分野における知識が足りないので、真偽を判断するのは難しい。また、目を引くものを他人に分かち合うことが自分の消息通のイメージを打ち立てる。情報社会では、人間は情報を交換することで認められた感を得る。デマを捏造する人はその点を利用して、デマ情報の中で、例えば、「早く家族を教えて！」、「怖い！友達に転送しろ」というような扇動的な言論と錯誤的な誘導も含まれている。情報共有の本心は悪くないのに、デマの拡散者

になってしまう。

本研究はデマ情報の特徴とデマ情報の発生原因について考察した。

(1) デマ情報の特徴

デマ情報の特徴の一つとしてはある特定の情報を狙って発信する。また、デマ情報のもう一つの特徴は期限が切れる情報を現在の状況を再度組み合わせる。例えば、「2018 年 XX 市の ATM 機は三竹台風で吹き飛ばされたお金がいっぱいだった」というデマは実際に発生していなかった。ATM 機が吹き飛ばされたのは 2017 年の天鴿台風だった。デマ情報の最も重要な特徴は似ているデマが繰り返し出てくる。広くて伝播されたデマ情報を分析すると、それらはいくつかの固定されたトピック領域に集中し、安定した物語の枠組みを形成された。例えば、「中国政府は納税者に数千億元を費やす」というようなデマは 2008 年北京オリンピック大会の時現れたが、2010 年上海世界博覧会のときと 2016 年 G20 杭州サミットのときも現れた。

(2) デマ情報の発生原因

まず、最も影響力があるのは政治的な原因でデマを拡散する。このようなデマ情報は悪意で国の顔に泥を塗っているかもしれないため、民衆は国への信任が失っていて動乱が起こしやすい。

次に、商業競争の原因でデマを拡散する。それは商家から芸能事務所まで様々なデマが出てきた。競争相手を中傷するため、ステマ工作員を雇用しネット上で大量的にデマを拡散する。

また、ネット有名人になりたいからデマを拡散する。近年からネット有名人または人気ブロガーは高収入という印象が残った。フォロワーを増やす、注目度を高めるために、真偽を判断せずに目を引くものであれば SNS に投稿する。

3.2 テキストデータの前処理

データを入手した後、データ前処理が必要である。コンピュータは二進数の 0 と 1 しか読めないから自然言語のテキストデータを変更しなければいけない。データという概念は画像、音声、テキストなどいろいろな形があるが、本研究の研究対象としてはテキストデータのみとなる。

本研究で行なったテキストデータの前処理には主に三つのプロセスがある。

(1) 形態素解析

形態素解析（Morphological Analysis）は自然言語のテキストデータから、言語で意味を持つ最小単位と呼ばれる形態素に分割する作業である。本研究で扱うテキストデータは全て中国語で記述された。中国語における自然言語処理の形態素解析ツールは色々あるが、本節で説明する jieba 以外に NLPIR（北京理工大学開発）、LTP（ハルビン工業大学開発）、THULAC（清華大学開発）など大学で開発したツールがあるが、BaiduNLP、AliyunNLP、BosonNLP などの商用ツールもある。jieba は単語分割と品詞付け機能以外に TF-IDF または Text Rank に基づいてキーワード抽出機能も持つが、中国語形態素解析ツールの中で最も使いやすいものかもしれない。

(2) ベクトル化

ベクトル化はテキストデータを数値のベクトルに変換することを指す。一般的に、形態素解析したデータを対象とする。重みつけによく採用されるのは TF-IDF 重みである。TF-IDF は tf と idf の組み合わせたものであるが、単語の重要度を評価する手法の一つである。

(3) 次元削減

次元削減は文字の通り、データの次元数を減らすことを指す。機械学習における次元削減は高次元データをそのまま低次元へ情報を落とし込むことである。

次元削減の目的はデータの圧縮とデータの可視化二つがある。よく利用されている次元削減手法は主成分分析（PCA）、ランダム・プロジェクション（Random Projections）と特徴集約（Feature Agglomeration）があるが、本研究で使った次元削減の手法としては主成分分析（PCA）である。

主成分分析（Principal Component Analysis；PCA）は相関のある多数の変数から相関のない少数で全体のばらつきを最もよく表す主成分と呼ばれる変数を合成する多変量解析の手法の一つである。

3.3 機械学習の手法

(1) サポートベクターマシン（SVM）

サポートベクターマシン（Support Vector Machine）は教師あり学習を用いるパターン認識モデルである。主な特徴はマージン最大化法とカーネル関数法の二つである。

マージン最大化法は訓練サンプルから、各データ点との距離が最大となるマージン最大超平面（Maximum-margin Hyperplane）を求めるという基準で線形入力素子のパラメータを学習する。ニューラルネットワークで用いられる最適点の探索法であるバックプロパゲーション法と比較して、マージン最大化法は理論的には過学習を起こしにくい。

n 次元のデータ集合 x＝$\{x_i\}$から超平面$w^T x_i$+b=0 までの距離は

$$d=\frac{w^T x_i+b}{\|w\|} \tag{1}$$

ここで、‖w‖は重みベクトルのノルムを表す。

SVM では n 次元実数ベクトル w とバイアスと呼ばれるスカラー変数 b をマージン最大化するように最適化問題を解くことによって決定する。二つのクラスを扱いやすくなるために、ラベル変数t_iを用意する。

$$t_i=\begin{cases}1 & (x_i \in K_1)\\ -1 & (x_i \in K_2)\end{cases} \tag{2}$$

それで、マージン最大化を最適化問題へ遷移する。

$$\max_{w,b}\frac{1}{\|w\|}, t_i(w^T x_i+b) \geq 1_{(i=1,2,3,\ldots,n)} \tag{3}$$

という関数を与えられる。最大化は逆数をとって最小化するのと変わらないので、式（3）は以下のように書き換えることができる。

$$\min_{w,b}\frac{1}{2}\|w\|^2, t_i(w^T x_i + b) \geq 1_{(i=1,2,3,\ldots,n)} \quad (4)$$

そこまで、線形分離可能を仮定したハードマージン[2]の標準的な定式化を行った。

マージン最大化法では、データは完全に分離できるような超高次元空間に投射する。この高次元空間への拡張は写像[3]という。しかし、投射する関数はわからなくて計算量も多い。ここでカーネル関数を使った計算を代わりに行う。

線形分離できない場合はスラック変数ξを導入して制約を弱める。この拡張はソフトマージン[4]と呼ぶ。式（4）は次のように変換する。

$$\min_{w,b,\xi}\left\{\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i\right\},\ t_i(w^T x_i + b) \geq 1 - \xi_i,\ \xi_i \geq 0\ _{(i=1,2,3,\ldots,n)} \quad (5)$$

ここで、C はどこまで制約条件を緩めかを指定するパラメータであり、正則化係数と呼ぶ。制約条件を最小化する問題に関してラグランジェ乗数$\alpha = \{\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n\}$と$\beta = \{\beta_1, \beta_2, \beta_3, \ldots, \beta_n\}$を使ったラグランジェ関数

$$L(w,b,\xi;\alpha,\beta) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\{t_i(w^T x_i + b) - 1 + \xi_i\} - \sum_{i=1}^{n}\beta_i\xi_i \quad (6)$$

を用いて最適化は

$$Q(\alpha) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{i=j}^{n}\alpha_i\alpha_j t_i t_j x_i^T x_j \quad (7)$$

を制約条件

$$\sum_{i=1}^{n}\alpha_i t_i = 0, 0 \leq \alpha_i \leq C\ (i = 1,2,3,\ldots,n) \quad (8)$$

のもとで解く問題に帰着する。

これを最適化するαを求めれば、w とβを求めることもできる。b については Karush-Kuhn-Tucker[5]条件から、式（7）が

$$\begin{cases}\alpha_i = 0 & : t_i\{\alpha_i(w^T x_i + b)\} > 1 \\ 0 < \alpha_i < C & : t_i\{\alpha_i(w^T x_i + b)\} = 1 \\ \alpha_i = C & : t_i\{\alpha_i(w^T x_i + b)\} < 1\end{cases}$$

の条件を満たす。

以上 w 、α、b を用いてデータx_jについての線形判別関数は

$$f(x_j) = sign\left(\sum_{i=1}^{n}\alpha_i t_i x_i^T x_j + b\right) \quad (9)$$

となる。ここで、sign（・）は符号関数を示す[11]。

本研究では、カーネル関数ガウスカーネルの中で、$\sigma > 0$の実数の場合である RBF カーネル（Radial basis function kernel）を用いた。RBF カーネルは内積のみを扱う線形アルゴリズムが非線形化できる。

ソフトマージンに RBF カーネル関数を適用するのは

$$\min_{w,b,\xi}\left\{\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i\right\}, K(x_i, x_j) = exp\left\{-\gamma\|x_i - x_j\|^2\right\}$$

（10）となる。

(2) ナイーブベイズ

ナイーブベイズ（Naive Bayes）はベイズの定理に基づいて、元に確率モデルであることがわかる。一般的に、条件付き確率に関して、以下が成立する。

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (11)$$

この中で、$P(A|B)$はデータ B が与えた時、その結果に対して原因の推定 A が正しい確率を指す。$P(B|A)$は原因の推定 A が正しい時その結果（データ B）の確率を指す。機械学習における分類問題は与えられたデータの全ての推定の確率を計算し、確率が一番高いものを推定結果をとして考えれば簡単になる。

このアルゴリズムの名前の前に、ナイーブという理想的な、単純な意味を持つ言葉をつく理由がある。データの属性はお互いに関係なし、つまり属性の独立性と同じ重要度であれば、分類問題に役に立つということがナイーブベイズ分類器の特徴である。現実世界でデータの属性は他のデータの属性に依存したり、複数の属性が持ったり、独立な属性はあまり成立しない[11]。

しかしながら、シンプルなナイーブベイズは単純という名前をついているにも関わらず、実際のデータセット上ではうまく機能する。

文章の i 番目の単語w_iが、クラス C から取り出された文章に出現する確率は、次のように書き表せる。

$$P(w_i|C) \quad (12)$$

問題を簡単にするため、単語はランダムに分布すると仮定している。あるクラス C が与えられた時、文章 D が取り出される確率は次のようになる。

$$P(D|C) = \prod_i P(w_i|C) \quad (13)$$

解きたい問題は「ある文章 D が、あるクラス C に属する

2 ハードマージン：線形分離可能の場合、訓練データの点全てを正しく分類できるようなパラメータ w と b の組が存在している仮定のもと SVM を用いる。このような SVM による分類をハードマージンと呼ぶ。
3 写像：元のデータの持つ特徴量をあらゆる形で組み合わせて、新しい特徴量で表現されたデータのこと
4 ソフトマージン：訓練データが線形分離可能な場合は理論的にはあるが、実問題ではかなり珍しい。多少の識別誤りは許すように制約を緩める方法はソフトマージンと呼ぶ。
5 Karush-Kuhn-Tucker 条件：変量 x∈ R^nについて、目的変数$f_p = e^T x$の最小化を考える。
線形計画問題の標準形は次の形で表見できる。
最小化：$f_p = e^T x$　制約条件：Ax＝b（$x \geq 0$）
これに対して、次の線形計画問題を双対問題（dual problem）と呼ぶ。
最大化：$f_d = b^T y (y \in R^m)$ 制約条件：$A^T y \leq c$
あるいは、スラック変数$\xi \in R^n$を導入すれば、
最大化：$f_d = b^T y (y \in R^m)$ 制約条件：$A^T y + \xi = c\ (\xi \geq 0)$
と書ける。

確率」、つまり、$P(C|D)$の値になる。ベイズの定理より、

$$P(C|D) = \frac{P(D|C)P(C)}{P(D)} \quad (14)$$

となる。

ここで、2 値分類の問題を考え、クラスは K と$\neg K$しかないと仮定する。

$$P(D|K) = \prod_i P(w_i|K) \quad (15)$$

かつ

$$P(D|\neg K) = \prod_i P(w_i|\neg K) \quad (16)$$

となる。式（14）に適用し、式（15）と式（16）を接合して次のようになる。

$$P(K|D) = \frac{P(D|K)P(K)}{P(D)} = \frac{\prod_i P(w_i|K)P(K)}{P(D)} \quad (17)$$

$$P(\neg K|D) = \frac{P(D|\neg K)P(\neg K)}{P(D)} = \frac{\prod_i P(w_i|\neg K)P(\neg K)}{P(D)} \quad (18)$$

したがって、確率比率

$$\frac{P(K|D)}{P(\neg K|D)} = \frac{\frac{\prod_i P(w_i|K)P(K)}{P(D)}}{\frac{\prod_i P(w_i|\neg K)P(\neg K)}{P(D)}} \quad (19)$$

を簡単化すると、

$$\frac{P(K|D)}{P(\neg K|D)} = \frac{P(K)}{P(\neg K)} \prod_i \frac{P(w_i|K)}{P(w_i|\neg K)} \quad (20)$$

となる。

これらを全て対数にすると、次の式が得られる。

$$\ln \frac{P(K|D)}{P(\neg K|D)} = \ln \frac{P(K)}{P(\neg K)} + \sum_i \ln \frac{P(w_i|K)}{P(w_i|\neg K)} \quad (21)$$

式（2.29）のように、$\ln \frac{P(K|D)}{P(\neg K|D)} > 0$なら、その文章はクラス K に属し、そうではない場合はクラス$\neg K$に属する。

4. 実験評価

4.1 データセット

Song ら[8]はデマ検知モデル CED（Credible Early Detection）の開発に関する研究で収集されたデータは Weibo 偽情報報告プラットフォームから提供されたデータである。そのうちに、オリジナル Weibo データ、デマデータとデマでないデータの転送情報をそれぞれ集めている。

本研究で適用するデータはデマデータとデマでないデータを含むオリジナル Weibo データのみとなる。また、デマデータは全部 Weibo 偽情報報告プラットフォームが実証された不確実な情報とデマデータであるが、デマでないデータは一般的なユーザの投稿である。

全データは 5135 件のうち、内容が意味不明なテキストや長さが短いテキスト合計 1748 件を除いて、デマデータは 1538 件とデマでないデータは 1849 件を用いた。

4.2 評価方法

本研究で利用する評価方法は混同行列（Confusion matrix）と Scikit-learn クラス分類レポート（Classification report）の適合率（Precision）、再現率（Recall）と F 値（F1-score）を使い、テストデータの分類性能を評価する。

表 2　混同行列の指標

		予測結果	
		「regular data」と識別	「rumor data」と識別
実際結果	regular data	TP (True-Positive)	FN (False-Negative)
	rumor data	FP (False-Positive)	TN (True-Negative)

$$Precison = \frac{TP}{TP + FP} \quad (22)$$

$$Recall = \frac{TP}{TP + FN} \quad (23)$$

$$F1 = \frac{2Recall \cdot Precision}{Recall + Precision} \quad (24)$$

そのうち、Precision は「regular data」と識別した件数の中で実際に分類された regular data の割合である。Recall は全ての regular data の中で正しく識別した件数の割合である。F 値は Precision と Recall をバランスよく合わせているかを示す指標である。

4.3 実験結果

今回のトレーニングデータはデマデータ 1391 件、デマでないデータ 1656 件で評価実験を行なった。モデルの分類性能を評価するため、F1 値に関して 5 回交差検証を行った。予測精度が予想より低かったため、グリッドサーチという手法でパラメータを最適化した。Rbf-SVC の最適なパラメータ C は 100 で、γは 0.001 となって予測精度を 0.71 から 0.86 に改善した。

テストデータはデマデータ（Class 1）147 件、デマでないデータ（Class 0）192 件で評価実験を行なった。混同行列を分析すると、192 件デマでない情報のうちに、Linear-SVC モデルは 27 件、Rbf-SVC モデルは 11 件、Naive Bayes モデルは 180 件が誤ってデマ情報と認識したが、147 件デマ情報のうちに、Linear-SVC モデルは 36 件、Rbf-SVC モデルは 48 件、Naive Bayes モデルは 18 件が誤ってデマでない情報と認識した。

表 3　分類器の分類結果

	Class	Precision	Recall	F1-score	Accuracy
Linear -SVC	0	0.82	0.86	0.84	0.80
	1	0.80	0.76	0.78	
Rbf-SVC	0	0.79	0.94	0.86	0.86
	1	0.90	0.67	0.77	
Naive Bayes	0	0.40	0.06	0.11	0.72
	1	0.42	0.88	0.57	

5. 考察

モデルを作成する目的はデマ検知に有効で、デマデータ

を分類できることである。今回はモデルが有効かどうかとデマ検知ができるかどうかの二つの方面から考察する。

一つ目は前提としたモデルが有効かどうかを評価する必要がある。SVM 分類器（Linear-SVC と Rbf-SVC）はクラス分類レポートで評価した結果が全部 1.00、つまりモデルのトレーニングデータに関しては 100%で識別した。非常に良い結果と示したが、しかし、それはトレーニングに対して一回のみの表現なので、実際のデータへの分類性能にはまだわからない。その問題を解決するために、F 1 値に対して交差検証を行った。実験が 5 回検証を行うので、データが少なくてもモデルの性能を最大化に評価できる。クラス分類レポートの F 1 値は Recall と Precision の調和平均 $\frac{2Recall \cdot Precision}{Recall+Precision}$なので、F 1 値を使って交差検証でモデルが有効かどうかを評価するには役に立つ。

二つ目はデマ検知ができるかどうかについて評価する。新しいデマが広く拡散するとき、モデルは専門家が真相を明らかにする前に「デマ情報」と判断すれば、デマ検知ができると設定する。分類器が「デマ情報」と判断したものは実際のデマ情報の割合が高ければ、デマ情報が検知できると考えられる。本研究の実験結果を見ると、テストデータのデマ情報（1 クラス）に対して、クラス分類レポートの precision 値を注目する。

SVM 分類器と Naive Bayes 分類器の比較は以下のように示す。

表 4　デマ情報に対して precision 値の比較

	Linear-SVC	Rbf-SVC	Naive Bayes
Precision(1 クラス)	0.80	0.90	0.42

SVM 分類器の Linear-SVC モデルと Rbf-SVC モデルはかなり良い結果を出していると示す。特に、Rbf-SVC モデルは「デマ」と判断した件のうちに、90%実際のデマ情報を識別したので、デマ検知に有効だと考えられる。Naive Bayes 分類器の結果は良くないと分かるが、本研究のデマ検知アルゴリズムに関しては SVM 分類器の表現は Naive Bayes 分類器より良い。

6. おわりに

本研究では、デマの特徴や発生原因を考察し、テキスト分類では、形態素解析、ベクトル化、次元削減というデータの前処理を行い、トレーニングデータとテストデータを分け、機械学習により、関数原理で支えられている SVM 分類器と確率原理で支えられている Naive Bayes 分類器を作成し、デマ検知アルゴリズムを開発した。

SVM 分類器は線形分類モデル Linear-SVC と非線形分類モデル Rbf-SVC があるので、Grid Search という手法で Hyperparameter を調整することで Rbf-SVC の精度を大幅に上げた。

実験評価では、Scikit-learn のクラス分類レポート機能を使い、F 1 値の交差検証を加えて作成したモデルを評価し、混同行列を加えてテキスト分類性能を評価した。結果により、本研究はデマ検知に有効であることがわかった。

今後の課題として、データ以外のテキストデータに対し、デマ検知アルゴリズムは適用できるかどうかことが挙げられる。デマは繰り返して出る事が分かったから、いくつかの固定されたトピック領域に集中し、安定した物語の枠組みを形成されたので、理論的には学習したデータの量が大きければ、分類精度が高くなる。他のテキスト分類に適用するのは検討すべきだと考えられる。

また、本研究で扱うデータセットは投稿されるユーザの情報とテキスト本文のコメントなども載せるので、デマを拡散されている地域の比較およびコメントに基づいて感情分析する事が可能であると考えられる。

最後、中国語の体系は複雑で膨大なので、意味解析におけるアルゴリズムの開発はまだ成熟ではない。データの意味を残したい場合は、データ処理の次元削減方式を変更しなければいけない。

参考文献

[1]安田雪（2013）．ソーシャルメディア上の情報拡散の特性――東日本大震災時のデマの事例とハブの役割．関西大学『社会学部紀要』第 45 巻第 1 号．pp.33-46
[2]榎美紀，村上明子，レイモンドルディー，小口正人.ソーシャルメディア上の情報拡散分析．第 6 回データ工学と情報マネジメントに関するフォーラム（DEIM Forum 2014），B4-6
[3]白井嵩士，榊剛史，不二夫鳥海等．Twitter におけるデマツイートの拡散モデルの構築とデマ拡散防止モデルの推定．人工知能学会全国大会（第 26 回），1C3-OS-12-1
[4]岡山光平，石川博，廣田雅春（2019）．分類器を用いた口コミサイトのレビューの分析．第 11 回データ工学と情報マネジメントに関するフォーラム（DEIM Forum 2019），pp.3-5
[5]廣瀬友香，木村昭悟，藤代裕之（2017）．SNS 上で拡散するニュース説明文の信頼性向上に向けた楽手データ整理．第 9 回データ工学と情報マネジメントに関するフォーラム（DEIM Forum 2017），I6-1
[6]黒木亮人，高山拓斗，成凱，安部恵介(2020)．機械学習を用いた連絡通知の自動分類に関する研究．情報処理学会研究報告
[7]刘知远, 张乐, 涂存超, 等．中文社交媒体谣言统计语义分析．中国科学: 信息科学, 2015, 45: 1536–1546, doi: 10.1360/N112015-0024
[8]Song, Changhe and Tu, Cunchao and Yang, Cheng and Liu, Zhiyuan and Sun, Maosong（2018）.CED: Credible Early Detection of Social Media Rumors．arXiv preprint arXiv:1811.04175
[9]柴田淳（2018）.『みんなの Python 』第 4 版．SB クリエイティブ