

# 機械学習モデルを使用した Kyutech コーパスのトピック分類

川寄慎乃介<sup>†</sup> 嶋田 和孝<sup>††</sup>

<sup>†</sup>九州工業大学 大学院情報工学府 情報創成工学専攻 〒820-8502 福岡県飯塚市川津 680-4

<sup>††</sup>九州工業大学 大学院情報工学研究院 知能情報工学研究系 〒820-8502 福岡県飯塚市川津 680-4

E-mail: <sup>†</sup>kawasaki.shinnosuke933@mail.kyutech.jp, <sup>††</sup>shimada@ai.kyutech.ac.jp

**あらまし** 複数人議論を行う際、議事録は決定事項の記録や非参加者への情報共有のために必要不可欠である。議論の自動議事録生成のためには、議論の文脈に沿った話題（トピック）を捉えることが第一に必要である。そこで、本研究では発話単位のトピック分類に取り組む。データセットには複数人議論コーパスである Kyutech コーパスを使用する。Kyutech コーパスでは1発話に複数のトピックタグが付与されている場合がある。したがって、トピック分類では、そのうち一つを正しく推定する多値分類と全てのタグを正しく推定するマルチラベル分類の2種類の問題を取り扱う。複数の機械学習技術を適用し、その有効性を比較、検証する。

**キーワード** トピック分類, マルチラベル分類, 機械学習, 議論マイニング

## Topic Classification of Kyutech Corpus by Machine Learning

Shinnosuke KAWASAKI<sup>†</sup> and Kazutaka SHIMADA<sup>††</sup>

<sup>†</sup> Department of Creative Informatics, Kyushu Institute of Technology  
680-4 Kawazu, Iizuka, Fukuoka 820-8502, JAPAN

<sup>††</sup> Department of Artificial Intelligence, Kyushu Institute of Technology  
680-4 Kawazu, Iizuka, Fukuoka 820-8502, JAPAN

E-mail: <sup>†</sup>kawasaki.shinnosuke933@mail.kyutech.jp, <sup>††</sup>shimada@ai.kyutech.ac.jp

**Abstract** Discussion summarization is one of the most important tasks for discussion analysis. Utterances in a discussion contains several topics, and the topics have an important role for the summarization. In this paper, we report a topic classification task of utterances in a multi-party discussion corpus: Kyutech corpus. In the corpus, each utterance contains one to three topic tags. We compare several machine learning methods for the topic tag classification task.

**Key words** Topic classification, Multi-label classification, Machine learning, Argument mining

### 1. はじめに

社会生活を送るうえで、組織内の意思決定や意見交換を目的とした複数人による議論を行う機会が多い。また、議論に直接参加していない人が議論の内容を理解するためには議事録が必要である。議事録を見直すことで議論の内容や決定事項などをいつでも把握することができる。しかし、人手による議事録作成では、理解のしやすさが製作者の能力によって大きく左右されてしまうという問題に加え、理解しやすい議事録の作成には時間がかかってしまうという問題も存在する。そのため、システムによる議事録の自動作成が望ましい。

議論では文脈に沿って様々なトピックが話されるため、議論中の発話を分析し、各トピックごとにまとめ上げることで、議事録を読む時間を短縮することができる。実際に、Nakayamaら [1] は、Kyutech コーパス [2] を対象に各トピックごとの自動

要約生成を行っている。このような自動要約生成を行うには、各発話がどのトピックに該当するかというタグ付けが必要である。Nakayama らの場合、人手によって付与されたトピックタグを使用した自動要約生成を行っている。しかし、議事録生成の自動化のためには、要約の前処理となるトピックのタグ付けも自動的に行われるべきである。

そこで、本研究では Kyutech コーパスを対象とした発話単位のトピック分類に取り組む。Kyutech コーパスには、人手によるアノテーションによって1発話につき最低1つ、最大3つのトピックタグが付与されている。しかし、それらの中には前後の発話を参考に付与されたトピックタグが存在する。そのため、発話単位のトピックタグの推測はたとえ人間であっても困難なタスクである。一方で、複数個のトピックタグのうち少なくとも1つのトピックタグを正しく推定することができれば、各発話をいずれかのトピックの要約生成に使用することができ

る。そこで、多値分類手法によるトピック分類実験を行い、その精度を検証する。しかし、本来は複数のトピックタグすべてを正しく推定して適切なトピックの要約生成に使用することが望ましい。そこで、各発話に含まれるすべてのトピックタグを推定するマルチラベル分類手法によるトピック分類実験についても同様にを行い、その精度も検証する。また、これらの実験の際には複数の機械学習モデルを使用し、各モデルの有効性を検証する。

## 2. 関連研究

トピック分類タスクにおいては、これまでに数多くの研究が行われている。Bharath ら [3] は、Twitter 上のツイートデータを対象とするトピック分類を SVM を用いて行った。彼らは研究の中で企業アカウントや個人アカウントといったアカウントの属性によって使われる語彙に差異がみられることに着目し、特徴量を新たに定義することで高精度なトピック分類を可能にしている。Xu ら [4] は、ニュース記事データセットをはじめ4つのデータセットを対象にトピック分類を行い、SVM や k-NN、ナイーブベイズ分類器、およびランダムフォレスト法などの分類モデルの比較検証を行った。このように、機械学習モデルを用いてトピック分類を行っている研究が多数存在している。しかし、これらの研究の大半は書き言葉がベースのデータセットを対象としている。Kyutech コーパスは議論中の発話を転記して作成されたコーパスであるため、これらの分類モデルが同様に Kyutech コーパスにおいても有効であるか定かではない。そこで、本研究では複数の分類モデルを使用して、精度の比較検証を行う。

議論コーパスを対象にトピック分類を行うことで、議論の事後分析に役立てるための研究は複数存在する。森崎ら [5] は単語の TF-IDF 値や共起頻度に基づいてトピックの定義、分類を行うことで発話単位での討議の分析を行った。Maria ら [6] はワクチン接種に関する議論を対象に教師なし機械学習モデル LDA [7] によるトピック分類を行うことで医療分野での議論分析に役立てた。しかし、これらの研究はルールベースに基づいたトピック分類もしくは教師なし機械学習モデルによるトピック分類を行っている。一方、本研究では教師あり機械学習モデルを使用することでより精度の高い議論データのトピック分類を行うことが期待できる。

## 3. データセット

本節では、本研究で使用したデータセットについて説明する。

3.1 節では Kyutech コーパスの概要について説明する。3.2 節ではデータセットとして用いるために Kyutech コーパスに施した処理について説明する。

### 3.1 Kyutech コーパス

本研究では、複数人議論コーパスとして Kyutech コーパス<sup>1</sup>を使用する。Kyutech コーパスは4人1組で行われる議論の発話を書き起こして作成されている。参加者は架空のショッピング

表1 Kyutech コーパスのデータ例

ID	必須	サブ1	サブ2	発話文
:	:	:	:	:
D	CandX			ラーメン ラーメンのかいぶつだけは 今のところ
A	CandX			そうやねー
A	CandX			まあ他の他の比較点とかはないかな
D	CandX	Price		高校生に 学割がある
D	CandX	Price		(笑)
:	:	:	:	:

モールの経営陣と仮定して議論を行う。議論のテーマはフードコートに参入させるのにふさわしい飲食店を候補店3つの中から1つ選ぶ、というものである。候補店や既存店にはそれぞれメニュー例、予算、座席数、営業時間、口コミなどの情報が設定されており、議論の参加者はそれらをまとめた資料を事前に手渡される。参加者はそれらの資料を10分間黙読した後、20分の議論を行う。対話データは参加者や飲食店の組み合わせなどを変えて全9種類存在する。

Kyutech コーパスには転記する発話単位が異なる2種類のデータ(転記単位と長い発話単位)が存在するが、本研究では長い発話単位のデータのみを対象に分類実験を行う。発話データの一部を表1に示す。各発話文には人手によるアノテーションによって、話者ID、発話の開始時間、発話の終了時間、トピックタグ、談話行為タグが付与されている。なお、表1では話者ID、トピックタグ、発話内容以外の情報は省いている。トピックタグの全てとその説明を表2に示す。

表1のように、各発話には必須タグとサブタグという2種類のトピックタグが付与されている。必須タグはアノテーションの際に前後の文脈を見た後、その文脈に即して付与されている。そのため、発話単位でトピック分類を行う本研究において、必須タグのみを正解とした場合、たとえ人間であっても非常に困難なタスクになってしまう。例として、表1の「高校生に学割がある」という発話は、前文脈から同じ候補店についての話が継続しているため必須タグはCandXとなり、サブタグはPriceとなる。しかし、この発話のみを見た場合、単語からPriceの推定は可能であるが、CandXの推定は非常に困難である。このことを考慮し、分類実験において、分類モデルが推定したタグは必須タグとサブタグを合わせた最大3種類のタグのうち1つでも一致していれば正しい結果であると評価する。

### 3.2 Kyutech コーパスの前処理

Kyutech コーパスはトピック分類のデータセットとして使用する際に、ノイズとなる発話文やデータのばらつきなどの問題が存在する。本節ではそれらの問題点を解消するために行った処理について説明する。

#### 3.2.1 発話文の整形

表1中にもみられるように、Kyutech コーパスの発話文の中には笑い声を表す「(笑)」や書き起こしの際に聞き取ることが

(注1) : <http://www.pluto.ai.kyutech.ac.jp/~shimada/resources.html>

表2 トピックタグの種類

トピックタグ	タグの説明
CandX	候補店 X についてのトピック
CandY	候補店 Y についてのトピック
CnadZ	候補店 Z についてのトピック
CandS	複数の候補店に関連するトピック
Exist1	既存店 1 についてのトピック
Exist2	既存店 2 についてのトピック
Exist3	既存店 3 についてのトピック
Exist4	既存店 4 についてのトピック
Exist5	既存店 5 についてのトピック
Exist6	既存店 6 についてのトピック
Exists	複数の既存店に関連するトピック
Closed	閉店したレストランについてのトピック
CIEx	既存店及び閉店した店の両方に関連したトピック
Mall	ショッピングモール全体についてのトピック
OtherMall	他のショッピングモールについてのトピック
Area	地域や都市についてのトピック
Access	ショッピングモールへのアクセスについてのトピック
Price	価格に対するトピック
Menu	メニューに関するトピック
Atomos	雰囲気に関するトピック
Time	営業時間などに関するトピック
Seat	座席数や回転率などに関するトピック
Sell	売上に関するトピック
Location	お店の立地に関するトピック
People	ターゲットにする顧客についてのトピック
Meeting	議事提案や最終決定部に関するトピック
Chat	雑談 (議題には直接関係のない発話)
Vague	何を言っているか判別不可能な発話

できなかった発話を表す「(?)」のような発話としての情報を持たないものが存在する。本研究では Kyutech コーパスからそのような発話を除く処理を行う。その結果、全発話から計 172 発話のデータが削除された。それに伴い、データ数は全 3120 発話となった。

Kyutech コーパスはその議題の設定から、候補店や既存店の名前には架空の店名が付けられている (表 1 中の「かいぶつ」など)。この架空店名は対話データごとに異なるものが定義されており、架空店名から得られる情報はトピックの推定には貢献しない。

そこで、本研究では架空の店名は全て記号に置換して一般化する。例として、表 1 中の「ラーメンのかいぶつだけは今のところ」という発話文中の架空店名「かいぶつ」を一般化した場合は「ラーメンの候補店 X だけは今のところ」となる。

### 3.2.2 トピックタグの集約

表 3 に Kyutech コーパスにおける必須タグのトピックタグ毎のデータ数を示す。表 3 からわかるように、Kyutech コーパスに付与されている 28 種類のトピックタグにはタグごとのデータ数の差異が大きいという特徴がある。このようなデータ数の偏りは分類モデルの学習や分類に悪影響をおよぼす。具体的には、データ数の多いトピックタグばかりを選択するモデルと

表3 トピックタグごとのデータ数

トピックタグ	データ数	トピックタグ	データ数
CandX	325	OtherMall	13
CandY	364	Area	46
CnadZ	393	Access	21
CandS	107	Price	82
Exist1	8	Menu	24
Exist2	14	Atomos	0
Exist3	28	Time	8
Exist4	81	Seat	49
Exist5	0	Sell	29
Exist6	45	Location	15
Exists	22	People	787
Closed	98	Meeting	428
CIEx	8	Chat	97
Mall	20	Vague	8

なってしまう場合がある。そのため、トピックタグごとのデータ数は可能な限り均一であることが望ましい。

表 2 や表 3 からわかるように、Kyutech コーパスに付与されている 28 種類のトピックタグは非常に細分化されている。しかしながら、これらのタグには候補店や既存店など、ある程度意味の似通っているタグが存在しており、それらは階層的に集約することが可能である。1 節で述べた Nakayama らの自動要約生成では、28 種類のトピックタグを利用した要約を生成しているが、本研究においてトピックタグの集約を行うことで、Nakayama らが生成した要約とは異なる粒度の要約生成を選択することも可能になる。

そこで、本研究では 28 種類のトピックタグを 8 種類に集約する処理を行う。表 4 に集約前後のトピックタグの対応関係を示す。表 5 に集約後における必須タグのトピックタグごとのデータ数を示す。集約によって、たとえば、高々数個から数十個のタグしか持ち得なかった Exist1~Exist6 が Exist として 304 事例で構成されるようになり、集約前と比較して、トピックタグごとのデータ数のばらつきが小さくなっている。このトピックタグのアンバランスさの解消は、分類モデルの分類精度の向上にも貢献すると考えられる。

## 4. 分類モデル

本節では分類モデルおよび使用する素性について説明する。本研究では各発話につき最低 1 つのトピックタグを推定するために多値分類実験と発話に付与された全てのトピックタグを推定するマルチラベル分類実験の 2 つを扱う。4.1 節では、多値分類を行う際に用いる分類モデルと素性について説明する。4.2 節では、マルチラベル分類を行う際に用いる分類モデルと素性について説明する。

### 4.1 多値分類モデル

本節では、多値分類実験で用いる分類モデルについて説明する。4.1.1 節では、自然言語処理分野では古くから一般的な素性である Bag-of-Words を使用する分類モデルについて説明する。4.1.2 節では、近年の自然言語処理分野で頻りに用いられ

表4 トピックタグの集約前後の対応関係

集約後のトピックタグ	集約前のトピックタグ
Cand	CandX
	CandY
	CandZ
	CandS
Exist	Exist1
	Exist2
	Exist3
	Exist4
	Exist5
	Exist6
	Exists
	Closed
	CIEx
Region	Mall
	OtherMall
	Area
	Access
Info	Price
	Menu
	Atmos
	Time
	Seat
	Sell
	Location
People	People
Meeting	Meeting
Chat	Chat
Vague	Vague

表5 集約後におけるトピックタグごとのデータ数

トピックタグ	長い発話単位
Cand	1189
Exist	304
Region	100
Info	207
People	787
Meeting	428
Chat	97
Vague	8

る汎用言語モデル BERT について説明する。4.1.3 節では、各分類モデルの出力結果を利用した多層的な分類モデルについて説明する。

#### 4.1.1 Bag-of-Words を利用するモデル

図1に、本節で説明する分類モデルが行う処理の流れを示す。Bag-of-Wordsとは、文中に出現する単語の頻度を数え上げることでベクトル化する手法である。具体的には、文書内の全単語をリスト化し、各発話文中の各単語の頻度をカウントすることで文書の語彙数と同じ次元数の特徴ベクトルを求める。

本研究では、単語の出現頻度など表層的特徴から発話文のトピックをとらえる目的で使用する。なお、ベクトル化する対象の単語は、名詞、動詞、形容詞の単語のみとし、それ以外の品

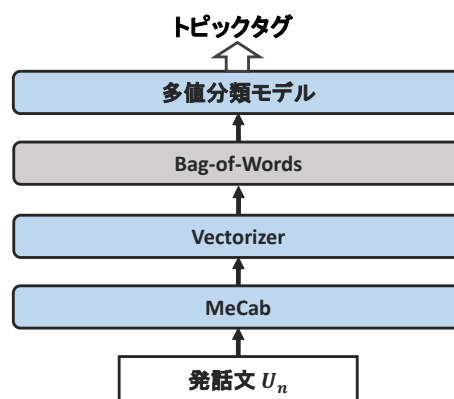


図1 Bag-of-Words 利用モデルによるトピック分類

詞の単語はすべて除外する。形態素解析には MeCab<sup>2</sup>を使用する。素性に Bag-of-Words に使用するモデルを以下に示す。

- SVM<sub>BoW</sub> : Support Vector Machines は深層学習以前によく利用された機械学習モデルである。マージン最大化戦略に基づき、正例と負例を分離する分離超平面を構築する。
- NB : Naive Bayes 分類器はベイズの定理に基づく確率分類器である。
- k-NN : k 近傍法 (Nearest Neighbor) は与えられた特徴ベクトルを特徴空間にマッピングし、もっとも距離の近いデータ k 個のラベルから多数決を行いラベルを推定する分類モデルである。
- RF : Random Forest 分類器は、複数の決定木で個別にラベル推定を行った後、その推定ラベルで多数決を行い、1つのラベルを決定する分類モデルである。

#### 4.1.2 BERT

BERT [8] は、事前学習モデルの一つである。事前学習モデルとは、大規模言語コーパスで事前学習を行うことで単語の汎用的な意味表現を学習する。その後、目的とするタスクのデータセットでファインチューニングすることで様々なタスクに適応可能な汎用言語モデルである。事前学習で大規模コーパスから単語の意味表現を学習しているため、タスクを問わず精度の高い分類を行うことが可能である。またその特徴から、小規模データセットにおいても従来の機械学習モデルと同等もしくはそれ以上の分類精度を期待できるということが知られている。Kyutech コーパスのデータ数はテキスト分類タスクにおいては小規模の部類であるため、BERT の有効性を期待して採用する。

図2に本タスクにおける BERT による推論の流れを示す。各発話はサブワード単位に分割され、トピックタグの推定に使用される。本研究では東北大学の乾研究室が公開している日本語用事前学習済み BERT モデル<sup>3</sup>を Kyutech コーパスでトピック分類タスクへファインチューニングしたものを使用する。

#### 4.1.3 SVM<sub>Meta</sub>

機械学習モデルでは、メタ学習という学習方法が用いられることがある。メタ学習とは、複数の機械学習モデルの学習結果

(注2) : <http://taku910.github.io/mecab/>

(注3) : <https://github.com/cl-tohoku/bert-japanese>

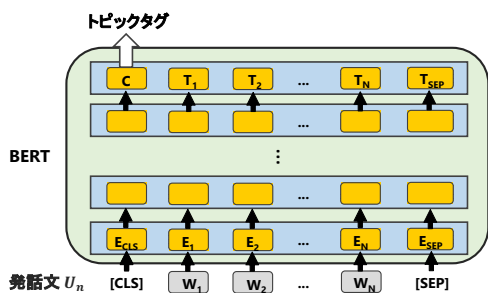


図2 BERTによるトピック分類

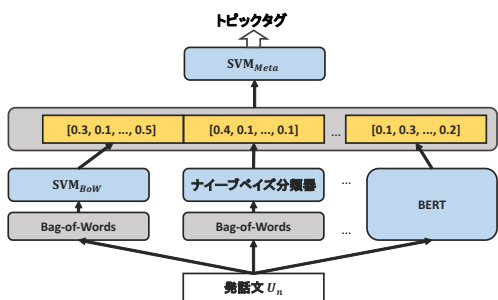


図3 SVM<sub>Meta</sub>によるトピック分類

を利用し、さらに良いモデルを作る学習法である。その中の一つにスタッキングという手法がある。スタッキングでは、複数のベースモデルが学習用データを基に出力した推定値を入力として学習を行う。各ベースモデルの推定値を入力として学習することで、それぞれのベースモデルの良い点を組み合わせた統合モデルが構築されることを期待する。

本研究では、4.1.1節および4.1.2節で説明した各モデル(SVM<sub>BoW</sub>, NB, k-NN, RF, BERT)をベースモデルとし、その結果を学習するメタ学習器をSVMとしたSVM<sub>Meta</sub>を用いる。図3に、SVM<sub>Meta</sub>による推論の流れを示す。図3のように、各ベースモデルは8トピックそれぞれの尤度を数値に変換し、8つの値で構成されるベクトルを生成する。これを1つの特徴ベクトルとみなし、分類モデルの数だけ結合させることでSVM<sub>Meta</sub>の入力として取り扱う。すなわち、入力ベクトルの次元数は(8次元のベクトル) × (ベースモデルの数5) = 40次元となる。

#### 4.2 マルチラベル分類モデル

表1に示したように、本研究で用いるKyutechコーパスにはトピックタグが同時に複数個付与されている発話が存在している。しかし、4.1節で示した多値分類モデルでは、1つの発話につき1つのトピックしか推定できないため、完全な発話のトピック分類ができていないとは言い難い。そのため、本来はマルチラベル分類によって複数のトピックタグを同時に推定することが望ましい。

本節では、マルチラベル分類に使用する分類モデルと使用する素性について説明する。本研究では、素性をBag-of-Wordsのみに統一し、マルチラベル分類を行う。本研究のマルチラベル分類で使用する分類モデルを以下に示す。

- ML-KNN: マルチラベルk近傍法[9]は、k-NNから派

表6 多値分類結果

分類モデル	Accracy
SVM <sub>BoW</sub>	0.551
NB	<b>0.612</b>
k-NN	0.494
RF	0.578
BERT	0.556
SVM <sub>Meta</sub>	0.522

生じたマルチラベル分類器であり、基本的な処理は最も距離の近いk個のデータを使用するというk-NNと同様である。

- ML-RF: Random Forest分類器に基づき、各ラベルごとに複数の決定木を使用して多数決を行うことでマルチラベル分類を行う。
- BR<sub>model</sub>: Binary Relevance手法は、各ラベルごとに2値分類をするモデルを割り当てることでマルチラベル分類を行う手法である。本研究ではSVMとNBをモデルとしたBR<sub>SVM</sub>とBR<sub>NB</sub>の2つを用いる。

## 5. 分類実験

本節では各分類実験の結果から精度の比較や考察を行う。なお、評価の際には全9対話のうちテストデータに用いる1対話と訓練データに用いる8対話の組み合わせを変えて9回分類実験を行い、その平均値を使用する。

### 5.1 多値分類実験

SVM<sub>BoW</sub>はカーネルにLinearカーネルを使用し、正則化手法はL2正則化、正則化パラメータは1.0とした。ナイーブベイズ分類器は多項分布モデルを使用した。k-NNの推論に使用する近傍サンプル数は5とした。ランダムフォレスト分類器の推論に使用する決定木の数は100とする。BERTは学習率0.0001、エポック数10、バッチサイズ64でファインチューニングしたものを使用する。SVM<sub>Meta</sub>はカーネルにRBFカーネルを使用し、正則化手法はL2正則化、正則化パラメータは1.0とする。

表6に各モデルのAccuracyを示す。この実験では、最大3つ付与されたトピックタグのうち1つが当たれば正解としているため、PrecisionとRecallの算出はできない。したがって、実験ではAccuracyのみを示している。

最高精度はNBであった。最新の大規模事前学習モデルであるBERTは、NB以外にもRFに勝てておらず、SVM<sub>BoW</sub>とも大きな差はない結果となった。BERTとそれ以外のモデルの違いは、BERTが単語のembeddingを利用しているのに対し、それ以外のモデルがBag-of-Words(つまり、単語の表層)を利用している点である。今回実験の対象としているKyutechコーパスのような特殊な語彙を用いる小規模な対話データにおいては、事前学習により得られた汎用的な知識を用いるモデルよりも単語の表層の特徴を利用するモデルの方が効果的である可能性が示唆される。追加実験として、各単語のword2vecのベクトルの総和を入力としたSVM<sub>w2v</sub>でも同様の実験を行った。結果として、SVM<sub>BoW</sub>と比較して、精度は向上したが(0.591)、本タスクでのBag-of-WordsによるNBのほうが高精度であった。

表7 マルチラベル分類結果

分類モデル	Precision	Recall	F1	Accuracy
ML-KNN	0.359	0.187	0.198	0.130
ML-RF	0.410	0.230	0.262	0.162
BR <sub>SVM</sub>	0.429	0.262	0.298	0.173
BR <sub>NB</sub>	<b>0.446</b>	<b>0.272</b>	<b>0.307</b>	<b>0.193</b>

複数のモデルの統合によって精度向上を期待した SVM<sub>Meta</sub> についても、統合による効果は得られなかった。これは、それぞれのベースモデルのエラーの傾向<sup>4</sup>が大きく変わらないことや学習のためのデータサイズが十分ではないなどの原因が考えられる。

## 5.2 マルチラベル分類実験

ML-KNN の推定に使用する近傍サンプル数は 10 とした。ランダムフォレスト分類器の推論に使用する決定木の数は 100 とした。Binary Relevance 手法で用いる SVM のカーネルには Linear カーネルを使用し、正則化手法は L2 正則化、正則化パラメータは 1.0 とした。Binary Relevance 手法で用いるナイーブベイズ分類器は多項分布モデルを使用した。

表 7 に各モデルの Precision/Recall/F1-Score のマクロ平均および Accuracy の平均値を示す。ここで示している Accuracy は付与されているタグ全てと予測結果が完全一致した場合のみを正解として算出している。表 7 から、マルチラベル分類においてもナイーブベイズ分類器を使用したモデルが一番高いスコアを出すことが分かった。しかしながら、いずれのモデルも十分な精度とはいえ、Recall が低いことから複数のタグを正確に予測することが困難であることがわかった。

## 5.3 文脈情報を追加した分類実験

3.1 節で説明したように、Kyutech コーパスには前後の文脈情報無しには推定が難しいトピックタグが存在している。しかし、発話単位でのトピック分類では、それらを正しく推定するのは困難である。そのため、何らかの文脈情報を素性に追加することによって、分類精度の向上が期待できる。

5.1 節および 5.2 節から、Kyutech コーパスにおけるトピック分類には単語の表層的特徴を使用したナイーブベイズ分類器が最も有効であることが分かった。そこで、本節ではナイーブベイズ分類器の入力に対し、1 つ前の発話文を利用した文脈情報の追加を行い、さらなる分類精度の向上を図る。ナイーブベイズ分類器への文脈情報の追加には、以下の 2 つの方法を提案する。

- Addition: 1 つ前の発話文の Bag-of-Words ベクトルと推定を行う発話文の Bag-of-Words ベクトルを要素ごとに加算する。
- Concatenation: 1 つ前の発話文の Bag-of-Words ベクトルと推定を行う発話文の Bag-of-Words ベクトルを連結する。

また、BERT には文節記号 [SEP] を利用して 2 文を同時に入力することで、文脈情報を追加する入力方法がある。そこで、BERT に対しても同様に文脈情報の追加を行い、精度の向上を

(注4): どのモデルも多数派のタグを優先して推定する傾向があり、少数派のトピックタグの推定精度は十分ではない場合が多い。この傾向は NB 以外ではさらに顕著であった。

表8 文脈情報を追加した分類結果

実験対象	分類モデル	操作なし	+Addition	+Concatenation
シングル	NB	0.612	<b>0.639</b>	<b>0.632</b>
	BERT	0.556	<b>0.578(+[SEP])</b>	
マルチ	BR <sub>NB</sub>	0.193	<b>0.226</b>	<b>0.226</b>

図る。

表 8 に各モデルの文脈情報追加前と追加後における Accuracy を示す。表でシングルとは 5.1 節の実験設定を意味し、マルチは 5.2 節の実験設定を意味する。したがって、表での「操作なし」の数値は表 6 と表 7 での対応する数値と同じである。結果として、文脈情報の追加はどのモデルに対しても有効に機能した。結果を分析したところ、NB では最もデータ数の多いタグである Cand への誤分類が減り、People や Region の正解が増えるという傾向があった。これは文脈を加味した結果であると考えられる。

## 6. おわりに

本研究では、Kyutech コーパスのトピックタグを使用して発話単位のトピック分類実験を行った。今後は分類精度の向上とトピックタグを利用した要約生成に取り組む。

謝辞 本研究は科研費 20K12110 の助成を受けたものです。

## 文 献

- [1] Yuri Nakayama, Tsukasa Shiota, and Kazutaka Shimada. Corpus construction for topic-based summarization of multi-party conversation. In *Proceedings of the 2021 International Conference on Asian Language Processing (IALP)*, pp. 229–234, 2021.
- [2] Takashi Yamamura, Kazutaka Shimada, and Shintaro Kawahara. The kyutech corpus and topic segmentation using a combined method. In *Proceedings of the 12th Workshop on Asian Language Resources (ALR12)*, pp. 95–104, 2016.
- [3] Bharath Sriram, Dave Fuhry, Engin Demir, Hakan Ferhatosmanoglu, and Murat Demirbas. Short text classification in twitter to improve information filtering. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pp. 841–842, 2010.
- [4] Baoxun Xu, Xiufeng Guo, Yunming Ye, and Jiefeng Cheng. An improved random forest classifier for text categorization. *J. Comput.*, Vol. 7, No. 12, pp. 2913–2920, 2012.
- [5] 森崎孔太, 塚井誠人, 難波雄二, 桑野将司. 司会者の関与が討議参加者の納得に及ぼす影響. 土木学会論文集 D3 (土木計画学), Vol. 70, No. 1, pp. 28–43, 2014.
- [6] Maria Skeppstedt, Andreas Kerren, and Manfred Stede. Vaccine hesitancy in discussion forums: computer-assisted argument mining with topic models. In *Building Continents of Knowledge in Oceans of Data: The Future of Co-Created eHealth*, pp. 366–370, 2018.
- [7] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, Vol. 3, pp. 993–1022, 2003.
- [8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, 2019.
- [9] Min-Ling Zhang and Zhi-Hua Zhou. Ml-knn: A lazy learning approach to multi-label learning. *Pattern recognition*, Vol. 40, No. 7, pp. 2038–2048, 2007.