

# 卒業論文

## 多段階マルチラベル分類を用いた 複合サービス推薦

指導教官 村上 陽平 教授

立命館大学情報理工学部  
先端社会デザインコース 4回生  
2600210258-6

中川 真一

2024年度(秋学期)卒業研究3(CH)  
令和7年1月31日

## 多段階マルチラベル分類を用いた複合サービス推薦

中川 真一

### 内容梗概

インターネット上の複数の Web サービスを組み合わせて作成した新たなサービスを複合サービスという。既存の Web サービスを組み合わせ、ユーザのニーズに合致する Web サービスを新たに構築する複合サービス合成の研究が行われている。現在、複合サービス合成は主には水平型サービス合成と垂直型サービス合成の 2 つのアプローチに分類される。前者は、所与のワークフローに対して、各タスクを実行するのに適した Web サービスの最適な組み合わせを選択するアプローチである。この手法では、事前にワークフローを与える必要があり、ワークフローの構築コストが高い。一方、後者は、人工知能のプランニング技術を用い、定義されたゴール状態に到達可能とする Web サービスの実行系列を生成する手法である。この手法では、ユーザがゴールを論理式で設定する必要があり、エンドユーザには困難である。

そこで、本研究では、言語モデルを用いて、自然言語で記述されたユーザの要件定義から、要件を満たす Web サービスの組み合わせを選択する手法を提案する。具体的には、本問題をマルチラベル分類問題として定式化し、BERT の事前学習済みモデルを用いてその問題を解く。本手法の実現にあたり、取り組むべき課題は以下の 2 点である。

### 膨大な予測ラベル数の解消

マルチラベル分類は予測したいラベル数が膨大になると、予測精度を大幅に下げる。これは、ラベルが多くなると学習データが疎となるからである。つまり、モデルの学習するラベルがほとんど 0 であり、関連するラベルを正確に 1 と予測するのが難しくなる。そのため予測するラベルの数を少なくする必要がある。

### カテゴリの階層化

複合サービスの構成サービスの原子サービス情報にはカテゴリ情報が含まれている。カテゴリは 152 個あり、幅広い分野のカテゴリが含まれている。カテゴリ予測を行う際に 152 カテゴリの予測をするのではなく、大まかなカテゴリから予測を行うことで予測の精度を上げることができると考えた。そのためカテゴリの階層化が必要である。

1つ目の課題に対しては、多段階でマルチラベル分類を行うことで解決する。サービスの説明文からマルチラベル分類で所属する大まかなカテゴリから予測する。そして所属するカテゴリの予測を段階的に詳細にしていき、サービスの説明文が属するカテゴリを予測し、最後に、予測したカテゴリの中に属するサービスの中でサービス推薦を行う。2つ目の課題に対しては、カテゴリをベクトル化し、クラスタリングを行うことで解決した。word2vec は GoogleNews-vectors-negative300 を用いた。GoogleNews から学習させた事前学習 word2vec モデルであり、300 万語の単語とフレーズの 300 次元ベクトルが含まれている。幅広い英単語に対応し、300 次元のベクトルで高精度に意味や文脈を表現することができる。これを用いてカテゴリをベクトル化した。Balanced K-means 法はクラスタサイズを一定にしながらクラスタリングを行うことが可能である。そのため予測するカテゴリ数の調整を行うことができる。

多段階マルチラベル分類と一段階マルチラベル分類の予測精度を比較することによって、提案手法の有効性を検証する。予測した各カテゴリとその正解カテゴリから適合率,再現率,F1 スコアを算出することで精度を評価する。また、クラスタリングで生成したクラスタの正当性は、Davies-Bouldin 指数で評価した。また本研究の貢献は以下の通りである。

### 膨大な予測ラベル数の解消

多段階マルチラベル分類を適用することで膨大な予測ラベルのマルチラベル分類の精度を保つことができることが分かった。多段階マルチラベル分類は一段階マルチラベル分類と比較して、適合率は 0.35 向上し、218%向上した。F1 スコアは 0.19 向上し、107%向上した。多段階マルチラベル分類を用いることで精度が大幅に向上している。

### カテゴリの階層化

大カテゴリは 8 個、中カテゴリは 4 個に段階的にクラスタリングを行った。各クラスタは Davies-Bouldin 指数によってクラスタリング精度が低いことが分かった。しかし精度が低いクラスタリングでカテゴリを作成し、多段階マルチラベル分類を行っても精度が向上することがわかった。

# Composite Service Recommendation with Multi-Step Multi-Label Classification

Shinichi NAKAGAWA

## Abstract

Composite service is a new service that combines multiple web services on the Internet. Service composition creates new services that users need using existing web services. Current research has two main approaches: horizontal and vertical. The horizontal approach selects the optimal combination of web services for each task in a predefined workflow; however, it requires constructing that workflow in advance, increasing development costs. The vertical approach relies on artificial intelligence planning to generate a sequence of web services that achieve a specified goal state; however, it requires representing the goal as a logical formula, hard for end users. This study proposes a method that leverages a language model to determine appropriate service combinations based on user requirements expressed in natural language. The approach formulates the task as a multi-label classification problem and uses a pre-trained BERT model to solve it. Implementing this method calls for addressing two challenges:

### **Resolving the enormous number of predicted labels**

Multi-label classification experiences a significant decline in predictive accuracy when the number of labels to be predicted becomes extremely large. This decline occurs because the training data become sparse as the label count increases. In other words, most of the labels that the model learns remain zero, making it difficult to accurately identify relevant labels as one. Consequently, reducing the number of labels to be predicted was essential.

### **Hierarchical Categorization**

Atomic service information for the constituent services of a composite service includes category information. A total of 152 categories cover many domains. Instead of directly predicting all 152 categories, starting with broader categories improved prediction accuracy. Therefore, hierarchical organization of these categories becomes essential. Addressing the first challenge involves applying multi-label classification

in multiple steps. Multi-label classification first predicted the broad categories from the service description. The predicted categories are then refined step by step, identifying the specific categories to which the service description belongs. Finally, services falling under those predicted categories are recommended.

To address the second challenge, we vectorized categories and performed clustering. The word2vec implementation relied on GoogleNews-vectors-negative300, a pretrained model learned from Google News. This model contains 300-dimensional vectors for three million words and phrases, covering a broad range of vocabulary and representing meaning with high accuracy in 300-dimensional space. These vectors were used to represent the categories. Balanced K-means enabled clustering while maintaining uniform cluster sizes, making it possible to adjust the number of predicted categories.

Verifying the predictive accuracy of multi-step multi-label classification and single-step multi-label classification evaluates the effectiveness of the proposed method. Calculating precision, recall, and F1 scores from each predicted category and correct category confirmed the accuracy. The Davies-Bouldin index confirmed the validity of the clusters generated by the clustering process. The contributions of this study are as follows.

### **Resolving the enormous number of predicted labels**

Multi-step multi-label classification preserved predictive accuracy even when the number of labels become extremely large. Compared to single-step multi-label classification, multi-step multi-label classification improved precision by 0.35, representing a 218% increase. The F1 score increased by 0.19, corresponding to a 107% improvement. Consequently, multi-step multi-label classification enhances performance.

### **Hierarchical Categorization**

A stepwise clustering approach produced eight large categories and four medium categories. The Davies-Bouldin index revealed that each cluster had low clustering accuracy. However, forming categories from these low-accuracy clusters and applying multi-step multi-label classification resulted in improved accuracy.

# 多段階マルチラベル分類を用いた複合サービス推薦

## 目次

<b>第1章</b>	<b>はじめに</b>	<b>1</b>
<b>第2章</b>	<b>サービス合成</b>	<b>3</b>
2.1	水平型サービス合成	3
2.2	垂直型サービス合成	4
<b>第3章</b>	<b>多段階マルチラベル分類</b>	<b>5</b>
3.1	BERT	5
3.1.1	事前学習済みモデル	6
3.2	複合サービス推薦の定式化	6
3.2.1	データセット	6
3.2.2	マルチラベル分類問題の適用	8
3.3	多段階化	9
3.3.1	多段階マルチラベル分類モデルの構築	9
<b>第4章</b>	<b>カテゴリの階層化</b>	<b>13</b>
4.1	カテゴリのベクトル化	13
4.1.1	word2vec	13
4.1.2	GoogleNews-vectors-negative300	13
4.2	クラスタリング	14
4.2.1	BalancedK-means	14
4.2.2	カテゴリの階層化	14
<b>第5章</b>	<b>評価</b>	<b>16</b>
5.1	多段階マルチラベル分類モデルの性能評価	16
5.1.1	評価手法	16
5.1.2	結果	16
5.1.3	多段階マルチラベル分類モデルの実行例	17
5.2	クラスタリング精度の結果	18
5.2.1	クラスタリング精度の評価手法	18
5.2.2	クラスタリング精度の結果	19

<b>第6章</b>	<b>考察</b>	<b>21</b>
6.1	多段階マルチラベル分類と一段階マルチラベル分類の比較 . . . .	21
6.2	カテゴリの階層化 . . . . .	23
<b>第7章</b>	<b>おわりに</b>	<b>25</b>
	<b>謝辞</b>	<b>26</b>
	<b>参考文献</b>	<b>27</b>

## 第1章 はじめに

近年、サービスコンピューティングの研究が盛んに行われている。サービスコンピューティングとは、ユーザと情報通信技術とのギャップを埋める技術であり、様々なサービスを組み合わせることでユーザのニーズに適した複合サービスを構築することができる。例えば、翻訳サービスと音声認識サービスを組み合わせることでリアルタイムに音声を翻訳する複合サービスを構築することが可能である。このように既存のサービスを組み合わせることでユーザが必要とするサービスを構成することができる。通常、複合サービスを構成する際にはユーザが構成するために必要なサービスを選択する必要がある。膨大な数のサービスの中から必要なサービスはどれなのかを判断して選択する必要があるため非常に手間のかかる作業である。その作業をより簡単に行うことができるようにするために複合サービス合成の研究が行われている。複合サービス合成は主には水平型サービス合成と垂直型サービス合成の2つのアプローチに分類される。前者は、所与のワークフローに対して、各タスクを実行するのに適したWebサービスの最適な組み合わせを選択するアプローチである。この手法では、事前にワークフローを与える必要があり、ワークフローの構築コストが高い。一方、後者は、人工知能のプランニング技術を用い、定義されたゴール状態に到達可能とするWebサービスの実行系列を生成する手法である。この手法では、ユーザがゴールを論理式で設定する必要があり、エンドユーザには困難である。したがって多くのwebサービスからユーザが低コストで必要なサービスを選択することができるようにする必要がある。

そこで本研究では、言語モデルを用いて、自然言語で記述されたユーザの要件定義から、要件を満たすサービスの組み合わせを選択する手法を提案する。具体的には、本問題をマルチラベル分類問題として定式化し、BERTの事前学習済みモデルを用いてその問題を解く。本手法の実現にあたり、取り組むべき課題は以下の2点である。

### 膨大な予測ラベル数の解消

マルチラベル分類は予測したいラベル数が膨大になると、予測精度を大幅に下げる。これは、ラベルが多くなると学習データが疎となるからである。つまり、モデルの学習するラベルがほとんど0であり、関連するラベルを正確に1と予測するのが難しくなる。そのため予測す

るラベルの数を少なくする必要がある。

### カテゴリの階層化

複合サービスの構成サービスの原子サービス情報にはカテゴリ情報が含まれている。カテゴリは152個あり、幅広い分野のカテゴリが含まれている。カテゴリ予測を行う際に152カテゴリの予測をするのではなく、大まかなカテゴリから予測を行うことで予測の精度を上げることができると考えた。そのためカテゴリの階層化が必要である。

以下、本論文では2章では従来の研究の複合サービス合成のアプローチとして、水平型サービス合成、垂直型サービス合成について説明する。次に3章では本研究で構築した多段階マルチラベル分類について説明する。続いて4章では、カテゴリの階層化に必要なクラスタリングについて説明する。5章では3章と4章で説明を行った多段階マルチラベル分類とクラスタリング精度に対する評価を行う。6章にて考察を行う。最後に、7章にて今後の展望や課題について述べ、結論とする。

## 第2章 サービス合成

本章では、既存の複合サービス合成の手法について説明する。複合サービスとは、インターネット上の複数の Web サービスを組み合わせて作成した新たなサービスのことである。複合サービスの構築事例として図 1 の manaba の事例を用いる。manaba はクラウド型の教育支援サービスであり、教員と学生間のコミュニケーションを支援している。主な機能として学習管理サービスを提供している。このサービスはユーザのログイン機能を Microsoft が提供している Authentication を用いて作られている。また、課題の提出の機能では Google が提供している Google カレンダーを用いてユーザが使用している Google カレンダーに予定を追加する機能を提供している。さらに教材を配布する機能では、Google の提供している Youtube を用いて動画の埋め込みを可能にしている。このように既存のサービスを組み合わせることでユーザが必要としているサービスを提供できる。このような複合サービスを構築する際には必要な機能を提供できるサービスを探す必要がある。しかしインターネット上には膨大な数のサービスが存在している。その中からユーザが必要な機能を持つサービスを探すのは非常に困難である。このような問題に対して、複合サービス合成を用いることでユーザが必要な機能を持つ Web サービスを簡単に知ることができる。本章では、複合サービス合成に関する研究として、水平型サービス合成と垂直型サービス合成の2つのアプローチを説明する

### 2.1 水平型サービス合成

水平型サービス合成は、所与のワークフローに対して、各タスクを実行するのに適した Web サービスの最適な組み合わせを選択するアプローチである。Zeng



図 1: 複合サービスの例

らは、サービス品質に着目し、同一の機能を持ったサービスの中からサービス品質を重視したサービスを選択する複合サービス合成を提案した [1]。具体的には以下の方法である。1つ目はサービス選択に基準となるサービス品質の算出である。サービスの非機能的特性から各サービスのサービス品質ベクトルを算出する。2つ目は算出したサービス品質ベクトルからサービス選択を行うことである。サービス選択ではローカル最適化アプローチとグローバル最適化アプローチの2つのアプローチを行う。ローカル最適化アプローチは複合サービスの所定タスクを実行するサービスの選択は関係する他のタスクのサービスを考慮せず、タスクのユーザ制約を満たす最大スコアのサービスを選択する。グローバル最適化アプローチは、サービス選択は各タスクごとに個別に行われる。しかし他のタスクのサービスとの関係は考慮していないため全体の品質は最適ではない可能性がある。全体の品質を考慮してサービスを選択する。

## 2.2 垂直型サービス合成

垂直型サービス合成は、人工知能のプランニング技術を用い、定義されたゴール状態に到達可能とする Web サービスの実行系列を生成する手法である。Carman らは、サービス合成を計画問題と捉え、サービス合成アルゴリズムを構築した [2]。具体的には以下の方法である。1つ目はセマンティック型マッチングアルゴリズムである。異なるサービス間でのデータ型の非互換性を解消するために設計されている。あるサービスが出力するデータ型と、別のサービスが受け入れる入力データ型、最終的に達成すべき目標型との間に意味的な関連性があるかを確認する。データ型の名称を単純に比較するのではなく、そのラベルが示す意味や構造全体を考慮する。2つ目は、探索と実行を組み合わせたアルゴリズムである。動的で不確実な環境で計画を進めるために設計されている。計画の全体像をあらかじめ完全に把握することが難しい状況において、試行錯誤を繰り返しながら目標に近づく。最初にサービスを選択し、その操作を実行して得られた結果をもとに次のステップを決定するという手順を繰り返す。探索と実行を交互に行い、不完全な情報に基づいて計画を修正しながら進む柔軟なアプローチを可能にする。セマンティック型マッチングアルゴリズムが適切なサービス候補を特定し、探索と実行を組み合わせたアルゴリズムが動的な環境においてこれらのサービスを統合的に利用してサービス合成が行われる。

## 第3章 多段階マルチラベル分類

本章では、複合サービス推薦をマルチラベル分類で行うモデルの構築について説明する。入力テキストを処理する自然言語処理モデル BERT の概要、サービス推薦モデルの構築方法について説明する。

### 3.1 BERT

BERT[3]とは、Bidirectional Encoder Representations from Transformers の略で、2018年10月にGoogleが発表した自然言語処理モデルである。Transformer のエンコーダ部分のアーキテクチャを採用している。従来のニューラルネットワークと異なり、Transformer は Self-Attention を用いている。そのため文全体の長距離依存関係を捉えることができ、文脈を理解することが可能である。そのため Transformer のエンコーダを多層に積み重ねている BERT は入力テキストの双方向の文脈を考慮した自然言語処理が可能となる。BERT は事前学習と転移学習の2つの学習を行うことができる。

事前学習は単語穴埋めと次文予測の2種類のタスクから行われる。単語穴埋めは、文中で隠された単語の予測を行うタスクである。予測する単語はテキスト中の単語から15%をランダムに選択される。そのうち80%は [MASK] という特殊トークンに置き換えられる。10%は選択した単語をランダムな別単語に置き換える。10%は元の単語のままにする。隠された単語に対して、左右の文脈を用いながら元の単語を予測する。これにより、前後の文脈を考慮した言語表現を学習することができる。次文予測は2つのテキストが元の文書で連続して続いていた文か無関係で連続していない文かを判別させる。コーパスのある文に対して50%で実際に続く文、残りの50%を実際に続かない文からランダムで選択する。テキストの先頭に [CLS] という特殊トークンを配置し、それに対応する最終層の埋め込み表現を用いて、文が連続するものかしないものかを二値分類する。これにより、文同士の関係を学習し、文脈を理解する能力を向上させる。これらのタスクにより、BERT は文全体を参照し、文中の単語の意味と文脈をより深く理解したモデルとなった。

転移学習は、既存のモデルに新たなタスクに対応するように設計したニューラルネットワーク層を重ね、そのタスクに対応できるようにすることである。BERT は Transformer のエンコーダに基づいているため転移学習のタスクは文

や単語の分類問題が中心となる。事前学習にて得た既存の知識を活用するため転移学習に必要なデータ量は少ない量で済む。

### 3.1.1 事前学習済みモデル

事前学習済みモデルとは大規模なデータセットであらかじめ学習したモデルである。大規模なデータセットであらかじめ学習されているため一からモデルを学習させる必要がないのでコストを削減できる。言語の広範囲の知識を学習した事前学習済みモデルを活用して転移学習を行い、タスクに特化したモデルを作成することが一般的であり、一般的な特徴抽出能力が高いため高い精度が期待できる。本研究では、Hugging Face Hub で公開されている BERT 事前学習済みモデルを用いる。本研究で扱うデータセットは全て英語で書かれているため英語の事前学習済みモデルを用いる。今回用いる事前学習済みモデルは Google が公開している英語学習済みモデルの bert-base-uncased を用いる。

## 3.2 複合サービス推薦の定式化

サービス推薦モデルとは、サービスを説明しているテキストからそのサービスを構築するのに必要なサービスを予測するモデルである。データセット、マルチラベル分類問題の適用について述べる。

### 3.2.1 データセット

サービス推薦モデルを構築するために複合サービスデータと原子サービスデータを使用した。データは programmableweb に記載されているデータを用いた。複合サービスデータは 4284 件あり、図 2 のデータ構造となっている。複合サービス名をキーとし、複合サービスカテゴリ、複合サービス説明文、構成サービス、id の情報がバリューとなっている。複合サービスカテゴリと構成サービスは複数ある可能性があるためリスト型にしている。原子サービスデータは 21526 件あり、図 3 のデータ構造となっている。原子サービス名をキーとし、原子サービスメインカテゴリとサブカテゴリ、原子サービス説明文、サービス提供者、サービス利用者がバリューとなる。原子サービスのメインカテゴリはカテゴリのリストの始めものをメインカテゴリとする。それ以外は全てサブカテゴリとして扱う。これらのデータは全てユーザによって付与されたデータである。

複合サービスデータには欠損データがある。欠損データに対してフィルタリングを行った。フィルタリング条件は「複合サービスの構成サービスが原子サービスデータに含まれない」である。フィルタリングを行った結果複合サービス

```

“複合サービス名”:[
  [
    複合サービスカテゴリ
  ],
  “複合サービス説明文”,
  [
    構成サービス
  ],
  id
],

```

図 2: 複合サービスデータのデータ構造

```

“原子サービス名”:[
  [
    原子サービスメインカテゴリ
    原子サービスサブカテゴリ
  ],
  “原子サービス説明文”,
  [
    原子サービス提供者
  ],
  [
    原子サービス利用者
  ],
  id
],

```

図 3: 原子サービスデータのデータ構造

データは 4284 件となった。

複合サービスデータのデータ数が少ないためデータの拡充を行った。カテゴリ予測タスクにて、複合サービスデータと原子サービスデータを合わせたサービスデータを用いた。サービスの構成カテゴリを予測するタスクのためカテゴリデータも含まれている原子サービスデータも使用した。複合サービスのカテゴリは構成サービスのメインカテゴリを用い、データ拡充に用いた原子サービスデータのカテゴリはメインカテゴリのみを用いた。原子サービスは複合サービスデータに含まれているカテゴリがメインカテゴリであるもののみデータ拡充に用いた。データ拡充を行った後のデータ数は 22180 件である。データ拡充後のサービスデータは図 4 のデータ構造となっている。サービス名、サービス

```
“サービス名”:[
  [
    サービス構成カテゴリ
  ],
  “ サービス説明文 ”,
  id
],
```

図 4: データ拡充後のサービスデータのデータ構造

の説明文, 構成カテゴリ, id が含まれている. 構成カテゴリは複数ある場合があるためリスト型にする. サービスデータの学習データは 19746 件, 検証データは 2434 件にした. 本研究ではデータ拡張後のサービスデータを用いて BERT の事後学習を行う. サービスの説明文に適する構成カテゴリを予測する事ができるように学習を行う. よってサービスの説明文を入力とし, 構成カテゴリを出力として学習を行う.

### 3.2.2 マルチラベル分類問題の適用

マルチラベル分類問題は, あるテキストに対して付与されている複数のラベルを選択肢の中から予測する問題である. 例えばあるニュース記事に対して予測を行う. ニュース記事は複数の要素を扱うことがほとんどであるため予測されるラベルは複数である. なので「衆議院解散で日経平均株価が急落」というニュース記事に対して予測を行った場合, ニュースのカテゴリの中から「経済」「政治」と複数のラベルが予測される. このようにマルチラベル分類では, 各テキストが保持している複数の側面を考慮し, それに対応したラベルを適切に割り当てることが求められる.

3.2.1 で説明したデータを用いて, 複合サービス推薦をマルチラベル分類として定式化する. まず, 入力するテキストはサービスの説明文とする. サービスの説明文はサービスについての情報が充実していて, 細かくサービスについて表しているためマルチラベル分類の入力に適していると考えた. 出力するラベルはサービスを構成することに関係しているものとする. これはマルチホットベクトルで表す. これにより自然言語の説明文からサービスの構成に必要なものをマルチホットベクトルで予測することができるマルチラベル分類をすることができる.

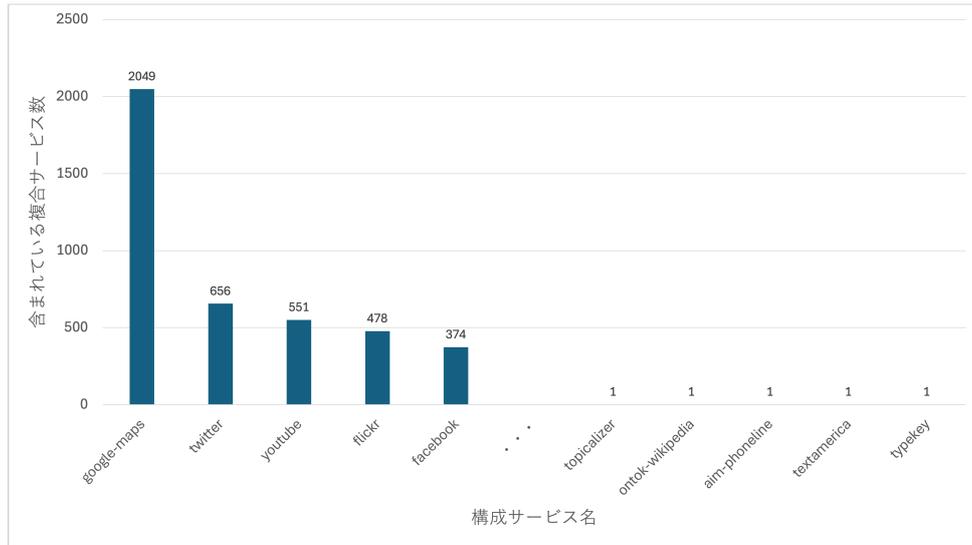


図 5: 構成サービスごとのデータ数

### 3.3 多段階化

マルチラベル分類は予測したいラベル数が膨大になると、予測精度を大幅に下げる。これは、ラベルが多くなると学習データが疎となるからである。つまり、モデルの学習するラベルがほとんど0であり、関連するラベルを正確に1と予測するのが難しくなる。複合サービスの構成サービスを学習する場合、1623件のラベルを付与することになる。それは非常に疎なラベルになるため精度が高くなる。またラベルの不均衡も発生する。ラベルの不均衡は特定のラベルのデータ数が他のラベル数に比べて極端に多い際に発生する問題である。この問題により、全体的な精度は高く見えるが少数ラベルに対する精度が低下する。そのためモデルの精度が高いとすることができない。複合サービスの構成サービスを学習する場合、図 5 の通り、各サービスの最小データ数は1件で最大データ数は2049件であるため非常に不均衡なデータである。サービスデータの特性上、有名なサービスが構成サービスとして用いられることが多い。データ数の上位は”google-map”, ”twitter”などの誰でも知っているサービスである。そのため不均衡なデータになると考える。

#### 3.3.1 多段階マルチラベル分類モデルの構築

この課題を解決することができるモデルを提案する。多段階でマルチラベル分類を行うことで解決する。サービスの説明文からマルチラベル分類で所属する大まかなカテゴリから予測する。そして所属するカテゴリの予測を段階的に

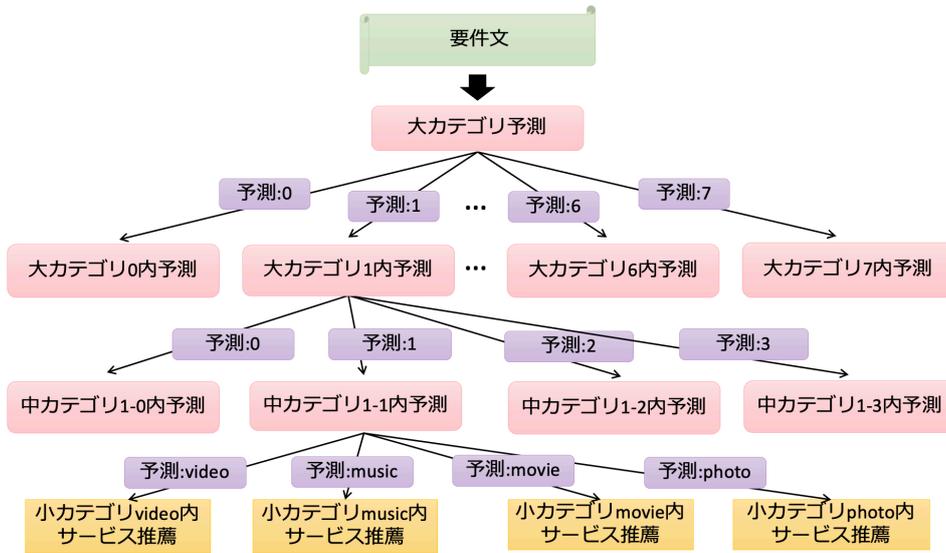


図 6: 多段階マルチラベル分類の全体像

詳細にしていき，サービスの説明文が属するカテゴリを予測し，最後に，予測したカテゴリの中に属するサービスの中でサービス推薦を行う．具体的にはまず複合サービスの説明文に対し，大カテゴリ，中カテゴリ，小カテゴリの順で多段階のマルチラベル分類を行う．最初に，大カテゴリ予測モデルを用いて該当する大カテゴリを予測し，その中で中カテゴリを予測する．同様に，中カテゴリと小カテゴリの予測を行う．最後に，予測された小カテゴリに属する原子サービスの中から最適なものを推薦し，これを推薦サービスとする (図 6)．これにより 1 度に予測するラベル数を減少させることができる．

BERT モデルからマルチラベル分類を行うためのモデルを作成する．BertMultiLabelClassification を Pytorch lightning を用いて実装する．まず事前学習済み BERT モデルを読み込み，モデルの最終層にて線形変換を行う．線形変換をすることでマルチラベル分類に適したモデルを実装することができる．次にトレーニング，モデルの保存を定義する．トレーニングでは学習データの 20% の検証データを用いたアリーストッピングをすることで過学習にならないようにした．最後に定義した BertMultiLabelClassification と学習データを用いてファインチューニングを行う．バッチサイズは 8 とし，trainer メソッドを用いてトレーニングをした．

大カテゴリ予測モデルの実装である．サービスの説明文からそのサービスを構成するのに必要なサービスの大カテゴリを予測するモデルである．複合サー



図 7: 大カテゴリ作成方法

ビスの構成サービスのメインカテゴリは 152 個ある。152 個のカテゴリでクラスタリングを行い、クラスタサイズが一定の 8 クラスタを作成した。そのクラスタを大カテゴリとして大カテゴリ 0 から大カテゴリ 7 とした (図 7)。学習データ 19746 件のカテゴリを対応する大カテゴリに置き換えて学習を行った。

中カテゴリ予測モデルの実装である。サービスの説明文からそのサービスを構成するのに必要なサービスの中カテゴリを予測するモデルである。各大カテゴリ内でクラスタリングを行い、クラスタサイズが一定の 4 クラスタを作成した。そのクラスタを大カテゴリに属する中カテゴリとして中カテゴリ (大カテゴリ-0) から中カテゴリ (大カテゴリ-3) とした。中カテゴリ 0-0 から中カテゴリ 7-3 まで 32 個の中カテゴリが作成した (図 8)。学習データ 19746 件から所属する大カテゴリごとにデータを分割したものを中カテゴリ予測モデルの学習データとした。学習データの構成カテゴリは各モデルで対象となる中カテゴリのみとした。学習データに対象ではない中カテゴリを入れると予測ラベルの減少に繋がらない。対象ではない中カテゴリのデータ数が少なくなってしまうためラベルの不均衡の問題が発生する (図 9)。

小カテゴリ予測モデルの実装である。サービスの説明文からそのサービスを構成するのに必要なサービスのカテゴリを予測するモデルである。各中カテゴリに属する小カテゴリを予測する。学習データ 19746 件から所属する中カテゴリごとにデータを分割したものを小カテゴリ予測モデルの学習データとした。中カテゴリ予測モデルと同様に学習データに含むカテゴリは予測対象のカテゴリに限定した。

多段階マルチラベル分類モデルを構築する。作成した大カテゴリ分類モデル、中カテゴリ分類モデル、小カテゴリ分類モデルを組み合わせる。まず入力テキ

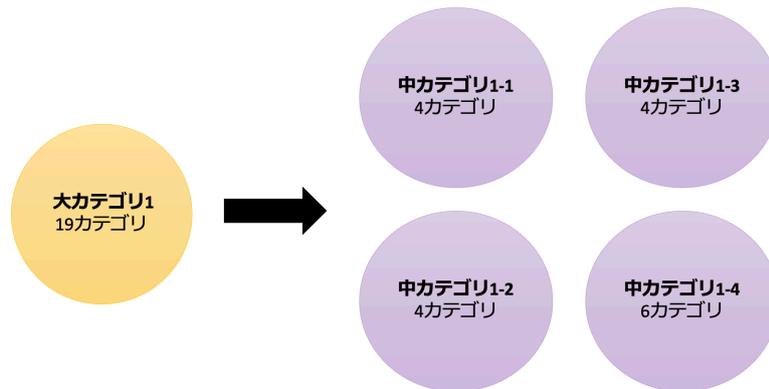


図 8: 中カテゴリ作成方法

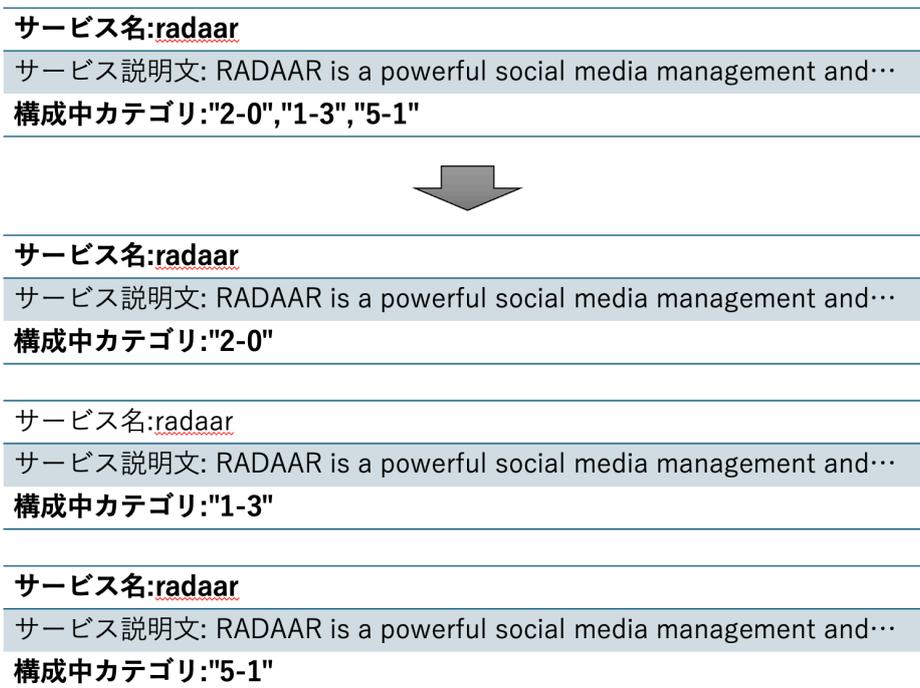


図 9: 中カテゴリ学習データの構成中カテゴリ

ストに対して大カテゴリ分類モデルで予測を行う。各大カテゴリに対する予測値が 0.5 を超えた大カテゴリの中カテゴリ予測モデルに移行する。次に入力テキストに対して中カテゴリ予測モデルで予測を行う。各中カテゴリに対する予測値が 0.5 を超えた中カテゴリの小カテゴリ予測モデルに移行する。最後に入力テキストに対して小カテゴリ予測モデルで予測を行う。各小カテゴリに対する予測値が 0.5 を超えた小カテゴリを入力テキストの推薦カテゴリとする。

## 第4章 カテゴリの階層化

本章では、カテゴリの階層化について説明する。第3章で構築する多段階マルチレベル分類を実現するためにカテゴリを階層化する必要がある。本研究ではカテゴリをベクトル化し、クラスタリングを用いて階層化を行った。

### 4.1 カテゴリのベクトル化

サービスデータにはカテゴリ情報が含まれている。本研究で扱うカテゴリは152個であり、複合サービスの構成サービスのメインカテゴリである。カテゴリは幅広い分野のカテゴリが含まれている。

#### 4.1.1 word2vec

カテゴリのベクトル化に word2vec を用いた。word2vec は、Mikolov らによって提案された自然言語処理システムである [4]。大規模なデータセットを用いてモデルの学習を行うことにより単語ベクトルの品質を向上させることを目的にしている。具体的な手法としては CBOW と Skip-gram というモデルを組み合わせる。CBOW は文中の周囲の単語から中央の単語を予測するタスクを行うモデルである。Bag-of-Words を拡張して実装されている。前に出てくる単語と後に出てくる単語を入力として対数線形分類モデルを構築することで中央の単語を正確に予測する。各単語の順序を無視し、周囲の単語の埋め込み表現を平均化して中央の単語を予測するためのベクトルを生成する。Skip-gram は中央の単語から周囲の単語を予測するタスクを行うモデルである。文脈の中央の単語を入力として対数線形分類モデルを構築することでその文脈で前後の単語を正確に予測する。前後の単語の予測はそれぞれの単語で行われる。

#### 4.1.2 GoogleNews-vectors-negative300

本研究では事前学習済み word2vec モデルで GoogleNews-vectors-negative300 を用いた。このモデルは約 1000 億語の Google ニュースデータセットの一部でトレーニングされた事前学習済みのベクトルで 300 万語の単語の 300 次元ベクトルが含まれている。そのため幅広いカテゴリの単語に対応することができる。サービスデータのカテゴリはユーザが設定したものであるため様々な単語に対応することが求められる。そのためこのモデルが適していると考えた。

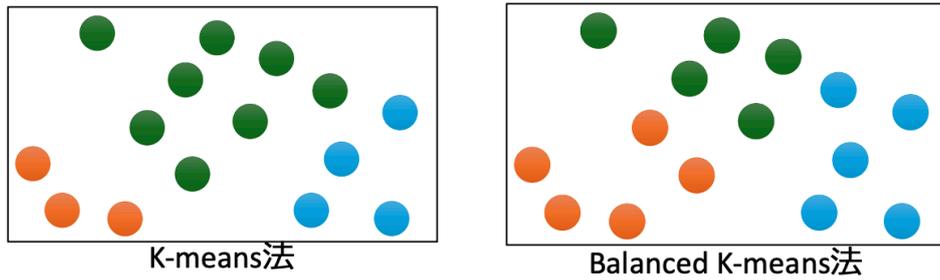


図 10: K-means 法と BalancedK-means 法の違い

## 4.2 クラスタリング

カテゴリのクラスタリングを行う。word2vec でベクトル化されたカテゴリをクラスタリングし、作られたクラスタをそのカテゴリが属するカテゴリとする。

### 4.2.1 BalancedK-means

クラスタリングは大カテゴリと中カテゴリを構築するのに用いられる。本研究では、大カテゴリと中カテゴリを作成するクラスタリング手法を BalancedK-mean とした。BalancedK-means は Malinen らが提案したクラスタサイズを一定にするように制約を加えたクラスタリング手法である [5]。BalancedK-means は K-means と異なり、クラスタサイズを一定にするようにクラスタリングを行う (図 10)。クラスタリングの流れは K-means と変わらないが割り当て部分が異なる。K-means ではユークリッド距離を基準に各データを最も近いセントロイドに割り当てる。割り当てるデータ数は考慮しないためクラスタサイズは自由に変動する。しかし BalancedK-means はクラスタサイズを一定にすることを制約にしているため、データ間の距離を基準にデータを事前に決めたクラスタのロットに割り当てる。ロットはデータ数とクラスタ数から決められる。ロットに割り当てる問題を Hungarian アルゴリズムで解く。本研究では、予測するラベル数の減少を目的としているためクラスタサイズを均等にすることを重視したクラスタリング手法を選択した。クラスタサイズを均等にすることで各カテゴリの学習データを確保することもできる。

### 4.2.2 カテゴリの階層化

サービスカテゴリから大カテゴリと中カテゴリを作成する。3.3.1 で作成した多段階マルチラベル分類では、大まかなカテゴリから予測するようにしている。そのためカテゴリを階層化し、徐々に詳細なカテゴリとなるようなカテゴリの階層化が必要である。カテゴリは大カテゴリ、中カテゴリ、小カテゴリと段階

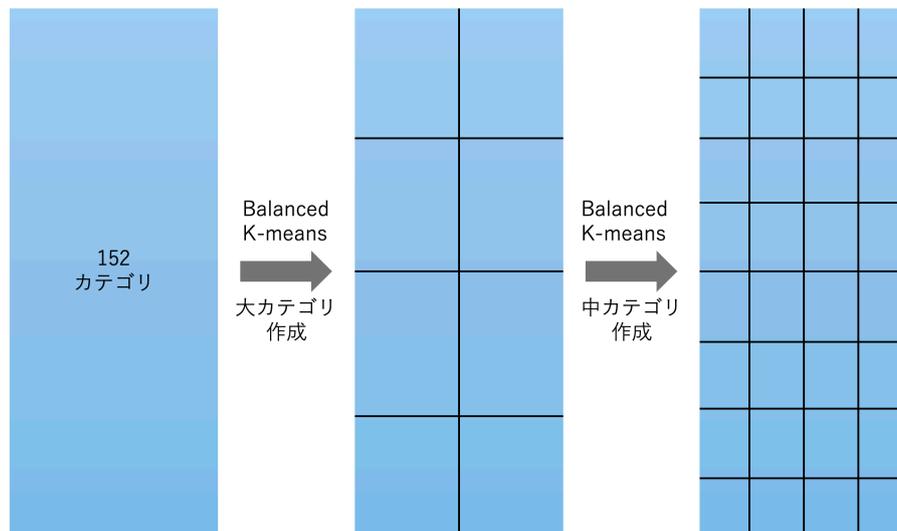


図 11: カテゴリの階層化

的に詳細なカテゴリとなる。図 11 の通り、段階的にクラスタリングをすることでカテゴリを階層化する。まず大カテゴリを作成する。大カテゴリは 8 カテゴリにする。152 カテゴリを 8 クラスにクラスタリングする。クラスサイズは均一であるため 1 つの大カテゴリに所属するカテゴリは 19 個のカテゴリとなる。次に中カテゴリを作成する。中カテゴリは大カテゴリ内で作成する。そのため各大カテゴリに所属する 19 個のカテゴリを対象にクラスタリングを行う。中カテゴリは大カテゴリ内で 4 つ作成する。そのため全体では 32 個の中カテゴリが作成される。それぞれのクラスサイズは均一であるため、1 つの中カテゴリに所属するカテゴリは 4 個から 6 個のカテゴリとなる。中カテゴリに所属するカテゴリを小カテゴリとする。

## 第5章 評価

本章では、第3章で作成した多段階マルチラベル分類の有効性と第4章で作成したカテゴリの精度を検証する手法について説明する。

### 5.1 多段階マルチラベル分類モデルの性能評価

多段階マルチラベル分類と一段階マルチラベル分類で比較を行い、多段階マルチラベル分類の有効性を確認する。一段階マルチラベル分類は、入力テキストから直接対象カテゴリを予測することである。つまり多段階マルチラベル分類を用いていない場合の通常マルチラベル分類と言える。

#### 5.1.1 評価手法

モデルの評価には適合率、再現率、F1スコアを用いる。scikit-learnのclassification\_reportを用いて評価関数を構築し、適合率、再現率、F1スコアの算出を行った。適合率、再現率、F1スコアは式(1)、式(2)、式(3)から算出される。各予測モデルの予測値は少数値であり二値ではないため、予測値が0.5以上の場合を1とし、それ以外は0とした。

$$\text{適合率} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{再現率} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{F1スコア} = 2 \cdot \frac{\text{適合率} \cdot \text{再現率}}{\text{適合率} + \text{再現率}} \quad (3)$$

本研究では、式(1)の分母にモデルが1と予測したカテゴリ数、分子にモデルが1と予測したカテゴリで正解していたカテゴリ数を当てはめる。式(2)の分母に検証データで1のカテゴリ数、分子にモデルが1と予測したカテゴリで正解していたカテゴリ数を当てはめる。式(1)と式(2)で算出したものを式(3)に当てはめる。

#### 5.1.2 結果

表1は中カテゴリ分類の多段階マルチラベル分類と一段階マルチラベル分類の結果である。中カテゴリ分類は入力テキストから32個の中カテゴリを予測することである。一段階マルチラベル分類の適合率は0.39、再現率は0.74、F1スコアは0.51となった。多段階マルチラベル分類の適合率は0.56、再現率は0.64、

表 1: 多段階と一段階の中カテゴリ分類の結果

分類方法	適合率	再現率	F1 スコア
一段階	0.39	0.74	0.51
多段階	0.56	0.64	0.60

表 2: 多段階と一段階の小カテゴリ分類の結果

分類方法	適合率	再現率	F1 スコア
一段階	0.16	0.77	0.27
多段階	0.51	0.62	0.56

F1, スコアは0.60 となった。適合率は多段階マルチラベル分類が高くなった。再現率は一段階マルチラベル分類が高くなった。F1 スコアは多段階マルチラベル分類が高くなった。このことから多段階マルチラベル分類の方が精度が高くなったと言える。

表 2 は小カテゴリ分類の多段階マルチラベル分類と一段階マルチラベル分類の結果である。小カテゴリ分類は入力テキストから 152 個の小カテゴリを予測することである。一段階マルチラベル分類の適合率は0.16, 再現率は0.77, F1 スコアは0.27 となった。多段階マルチラベル分類の適合率は0.51, 再現率は0.62, F1, スコアは0.56 となった。適合率は多段階マルチラベル分類が大幅に高くなった。再現率は一段階マルチラベル分類が高くなった。F1 スコアは多段階マルチラベル分類が大幅に高くなった。このことから多段階マルチラベル分類の方が精度が大幅に高くなったと言える。

表 1 では予測するカテゴリ数が 32 個であったが、表 2 では予測するカテゴリが 152 個であった。予測するカテゴリ数が約 4 倍になり予測難易が上がったにもかかわらず適合率は 0.05 減少, 再現率は 0.02 減少, F1 スコアは 0.04 減少となった。つまり、予測カテゴリ数が増加し、予測難易度が上がった場合でも予測精度は保つことができたと言える。

### 5.1.3 多段階マルチラベル分類モデルの実行例

3.3.1 で構築した多段階予測モデルの実行例を示す。図 12 の通り、入力したテキストから各段階のカテゴリで予測したカテゴリとその予測スコアを出力する。まず入力テキストから大カテゴリを予測する。例の場合では大カテゴリ 1,

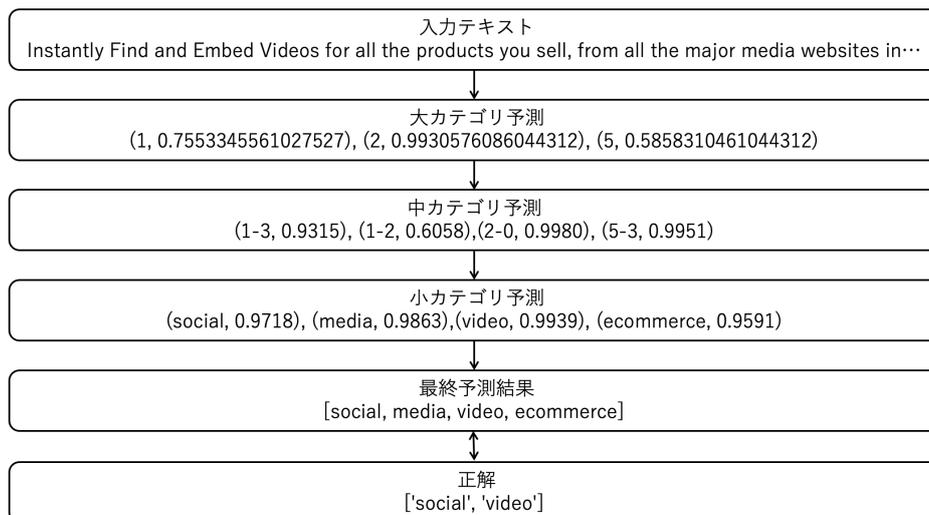


図 12: カテゴリ推薦の例

大カテゴリ 2, 大カテゴリ 5 を予測した。次に中カテゴリ予測である。予測した大カテゴリ内で中カテゴリ予測を行う。例の場合は, 大カテゴリ 1 内の中カテゴリ予測で中カテゴリ 1-2, 中カテゴリ 1-3 を予測した。大カテゴリ 2 内の中カテゴリ予測では中カテゴリ 2-0 を予測した。大カテゴリ 5 内の中カテゴリ予測では中カテゴリ 5-3 を予測した。最後に中カテゴリ内で小カテゴリを予測する。例の場合では, 中カテゴリ 1-2 では media を予測した。中カテゴリ 1-3 では social を予測した。中カテゴリ 2-0 では video を予測した。中カテゴリ 5-3 では ecommerce を予測した。最終予測結果と正解データを照らし合わせて精度を出す。

## 5.2 クラスタリング精度の結果

クラスタリングの精度を検証する。カテゴリを階層化する際にクラスタリングを用いた。その精度を検証し, 階層化が適切にできているかを確認する。

### 5.2.1 クラスタリング精度の評価手法

クラスタリングの精度を検証するのに, Davies-Bouldin 指数を用いる。Davies-Bouldin 指数 [6] は, クラスタ内の分散とクラスタ間の距離の比率で計算される指標である。クラスタの分散が小さく, 互いに十分に離れているほどスコアが小さくなり, 良いクラスタリングと評価される。式 4 でクラスタ内の分散を算出する。この値は, クラスタ内のデータ点がクラスタの重心にどれだけ密集し

表 3: 各カテゴリの Davies-Bouldin 指数

カテゴリ	Davies-Bouldin 指数
大カテゴリ	6.08
中カテゴリ 0	2.53
中カテゴリ 1	3.03
中カテゴリ 2	2.89
中カテゴリ 3	2.87
中カテゴリ 4	2.73
中カテゴリ 5	2.95
中カテゴリ 6	2.92
中カテゴリ 7	3.06

ているかを示す.

$$S_i = \left( \frac{1}{T_i} \sum_{j=1}^{T_i} |X_j - A_i|^q \right)^{\frac{1}{q}} \quad (4)$$

式 5 でクラスタ間の距離を算出する. この値は, クラスタ  $i$  とクラスタ  $j$  の重心間の距離を表す.

$$M_{ij} = \left( \sum_{k=1}^p |a_{ki} - a_{kj}|^p \right)^{\frac{1}{p}} \quad (5)$$

クラスタ内の分散とクラスタ間の距離を用いて, 式 6 でクラスタ間の類似度を算出する. この値が小さいほど, クラスタ間の分離が良好であることを意味する.

$$R_{ij} = \frac{S_i + S_j}{M_{ij}} \quad (6)$$

クラスタ間の類似度を用いて, 式 7 で Davies-Bouldin 指数を算出する. この値は, クラスタリング結果全体の適切性を示し, 小さいほど良好なクラスタリングであると評価される.

$$R = \frac{1}{N} \sum_{i=1}^N \max_{j \neq i} R_{ij} \quad (7)$$

### 5.2.2 クラスタリング精度の結果

表 3 は各カテゴリの Davies-Bouldin 指数である. 大カテゴリは 6.08 となった. これはクラスタリングの精度が低いことを表している. 中カテゴリ 0 は 2.53 となった. これもクラスタリングの精度が低いことを示す. 他の中カテゴリも同様

の結果となっている。そのため中カテゴリでもクラスタリングの精度が低いことを示している。各カテゴリでクラスタリングの精度が低いことを示した。つまりカテゴリの階層化は適切にできていないことがわかる。

## 第6章 考察

本章では、5.1と5.2で提示した結果を受けて、カテゴリ分類の精度について考察を行う。

### 6.1 多段階マルチラベル分類と一段階マルチラベル分類の比較

従来手法である一段階マルチラベル分類と本手法である多段階マルチラベル分類を利用した場合の分類結果を比較し、考察する。カテゴリごとの学習データ数と多段階マルチラベル分類と一段階マルチラベル分類の適合率、再現率、F1スコアを用いて、学習データ数によって多段階マルチラベル分類と一段階マルチラベル分類のスコアがどのように推移したかを比較する。まず、カテゴリごとの学習データ数に基づき、カテゴリを5つの区分(Q1~Q5)に分割した。具体的には、学習データ数が3~18のカテゴリをQ1, 19~38をQ2, 40~71をQ3, 72~167をQ4, 173~1929をQ5と定義する。各区分に含まれるカテゴリについて、一段階マルチラベル分類から多段階マルチラベル分類の上昇率を適合率、再現率、F1スコアを算出し、区分ごとの平均値を比較した。つまり多段階マルチラベル分類で値がどのくらい改善したかを検証した。図13は学習データ数別の適合率の上昇率の比較のグラフである。全体的に学習データ数が適合率が上がっていて、Q3, Q4, Q2, Q5, Q1の順で適合率が上がっている。これは、全体的に予測が正確になっていると言える。特に学習データ数が40から70のQ3での予測が3倍以上正確になったことが分かる。学習データが少ないカテゴリは予測の難易度が高いカテゴリである。Q2とQ1の学習データ数が少ないカテゴリでも予測が正確になった。多段階でマルチラベル分類を行うことで予測の正確性を上げることができることが分かる。

図14は学習データ数区分別の再現率の上昇率の比較のグラフである。全体的に再現率が下がっていて、Q2, Q4, Q3, Q5, Q1の順で再現率が下がっている。これは全体的に見逃しが増えたことになる。特にQ2のデータ数が若干少ないカテゴリに対しての見逃しが増えている。これは多段階マルチラベル分類はカテゴリの絞り込みを段階的に行うことにより、カテゴリに対しての予測を大カテゴリと中カテゴリの予測の部分で間違えてしまった場合にそのカテゴリを無視することになるため、見逃しが増えたと考える。そのため難易度の高い学習データ数が少ないカテゴリにて見逃しが増えてしまったと考える。しかし

最も高いQ2で-39%であり適合率と比べると低い値であるためモデルへの影響は少ないと考える。

図 15は学習データ数区分別の F1 スコアの上昇率の比較のグラフである。全体的に F1 が上がっていて、Q3, Q2, Q4, Q5, Q1 の順で F1 スコアが上がっている。これは精度が全体的に上がっていることになる。特に Q3 の学習データ数が 40 から 70 のカテゴリで精度が上がっている。Q3 は適合率の上昇率が非

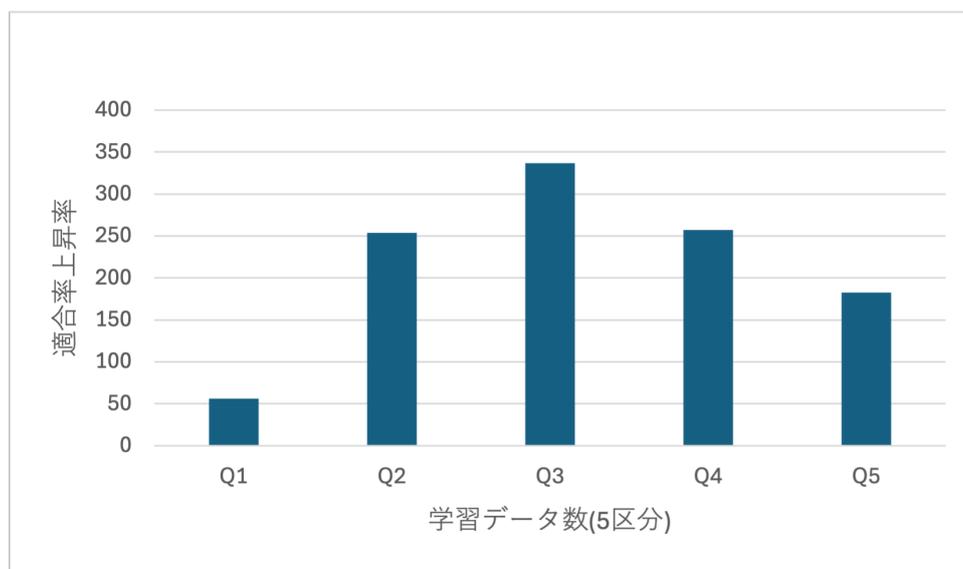


図 13: 多段階マルチラベル分類における学習データ数区分別の適合率の比較

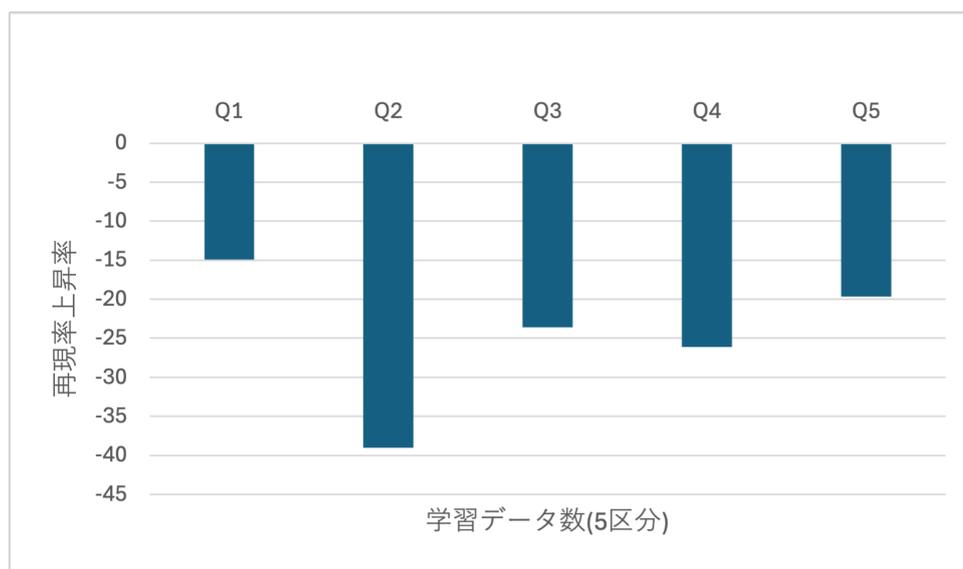


図 14: 多段階マルチラベル分類における学習データ数区分別の再現率の比較

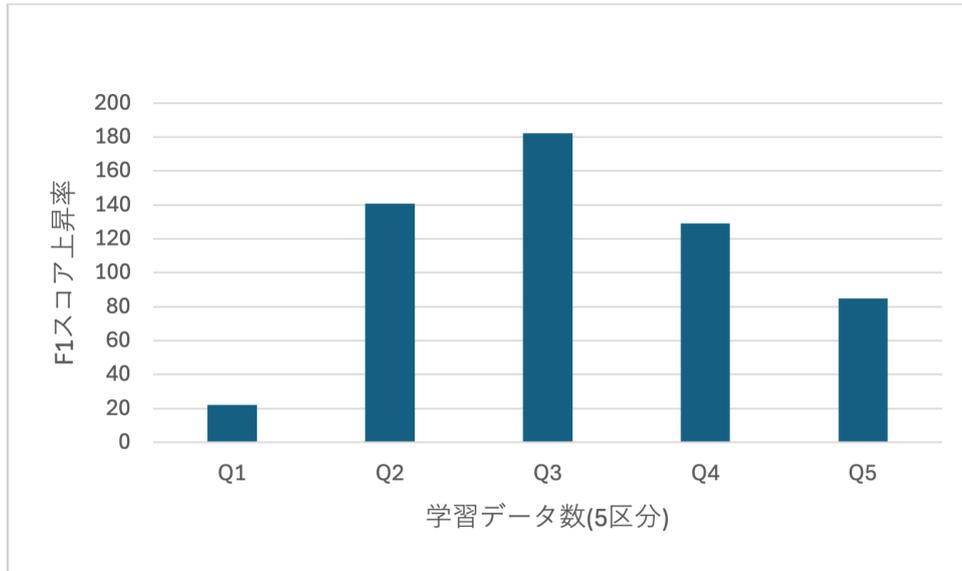


図 15: 多段階マルチラベル分類における学習データ数区分別の F1 スコアの比較

常に高く、再現率の減少を抑えることができたため F1 スコアが 180% も上昇した。多段階マルチラベル分類は一段階マルチラベル分類よりも適合率と再現率のバランスが取れていることがわかる。再現率は下がったが適合率の上昇率が高いため F1 スコアも高くなった。なので多段階マルチラベル分類のほうが精度が高いと言える。

## 6.2 カテゴリの階層化

カテゴリのクラスタリング精度とカテゴリの階層化から多段階マルチラベル分類について考察する。本研究ではカテゴリのクラスタリングを BalancedK-means を用いて行った。これはクラスタサイズを統一するという制約付きのクラスタリングである。そのため制約のないクラスタリング手法に比べてクラスタリング精度が低くなってしまふ。そのためクラスタリング精度が低くなったと考える。なのでカテゴリの階層化は適切にできていないと考える。しかしカテゴリの階層化は適切にできていないが多段階マルチラベル分類では精度が高くなった。カテゴリを階層化するために擬似的に大カテゴリと中カテゴリを設けたがその精度が高くななくても上手く分類することができているということである。学習データでこのテキストはこのカテゴリであると決め打ちすることでその与えられたカテゴリで正確に分類するように学習が進んだため精度が高くなったと考える。さらにラベルの不均衡の問題も解消された。表 4 は各モデル

表 4: 各カテゴリ分類モデルの学習データの不均衡比の平均

モデル	不均衡比
一段階	643
大カテゴリ	5.60
中カテゴリ平均	7.79
小カテゴリ平均	50.11

の学習データの不均衡比である。一段階マルチラベル分類の不均衡比は 643 で非常に高い値であり不均衡なデータであると言える。しかし大カテゴリ分類モデルの学習データの不均衡比は 5.60 であり低い値であるため不均衡ではない。中カテゴリ分類モデルは 8 個のモデルがある。その不均衡比の平均は 7.79 であり、低い値であるため不均衡ではない。小カテゴリ分類モデルは 32 個ある。その不均衡比の平均は 50.11 であり、大カテゴリと中カテゴリに比べると不均衡ではあるが一段階よりは不均衡ではない。そのため多段階マルチラベル分類では不均衡なデータでの学習を抑えることができたと言える。これが多段階マルチラベル分類の精度が上がった一因であると考え。カテゴリの階層化はラベルの不均衡の解消となったため精度の向上に繋がった。しかしより精度の高い予測にするためにはカテゴリの階層化を正確にやることは必要であると考え。

## 第7章 おわりに

本研究では、マルチラベル分類を用いたサービス推薦を提案した。まずカテゴリの階層化を行い、大カテゴリ、中カテゴリ、小カテゴリを作成した。それぞれのカテゴリに対してモデルを作成した。作成したモデルを組み合わせて、多段階マルチラベル分類モデルを構築した。多段階マルチラベル分類を用いることで一段階マルチラベル分類よりも推薦精度が高くなること適合率、再現率、F1スコアを用いて示した。本研究の貢献は以下の通りである。

### 膨大な予測ラベル数の解消

膨大な予測ラベルを段階的にクラスタリングしてカテゴリを階層化することで、ラベル不均衡問題を解消した。さらに、この階層構造に対応した多段階マルチラベル分類モデルを構築し、ラベル数の多い分類タスクにおいて、一段階マルチラベル分類と比較して精度を向上させることに成功した。

### カテゴリの階層化

152個のカテゴリを8個の大カテゴリと32個の中カテゴリを作成し、カテゴリの階層化を行った。カテゴリの階層化を行う際に使用したクラスタリングの精度を検証した。カテゴリの階層化を行うことでラベルの不均衡を解消し、マルチラベル分類の精度を向上させることが分かった。

今後の展望として、カテゴリの階層化のクラスタリング精度を改善することによってより多段階マルチラベル分類の精度が向上するかを検証する。またサービス推薦まで行い、ユーザにとって適切なサービスが推薦できているかを検証する必要がある。最後に、サービス推薦モデルを用いることによりユーザにとって最適な構成サービスを簡単に選択する事ができるようになり、複合サービスの開発がより一層進むことが期待される。そのためには日々開発される原子サービスと複合サービスを随時学習し、ユーザにとってより良い推薦サービスを得られるようにする必要があると考える。

## 謝辞

本研究を行うにあたり, 熱心なご指導, ご助言を賜りました指導教官の村上陽平教授を並びに松本賢司さんに深く感謝を申し上げます.

## 参考文献

- [1] Liangzhao Zeng, Boualem Benatallah, Anne H. H. Ngu, Marlon Dumas, Jayant Kalagnanam, Henry Chang: "QoS-aware middleware for Web services composition," *IEEE Transactions on Software Engineering*, vol. 30, no. 5, pp. 311-327 (2004).
- [2] Mark A. Carman, Luciano Serafini, Paolo Traverso: "Web Service Composition as Planning," *ICAPS 2003 Workshop on Planning for Web Services*, pp. 1636-1642 (2003).
- [3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova: "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *arXiv preprint*, arXiv:1810.04805 (2018).
- [4] Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean: "Efficient Estimation of Word Representations in Vector Space," *arXiv preprint*, arXiv:1301.3781v3 (2013).
- [5] Mikko I. Malinen, Pasi Fränti: "Balanced K-Means for Clustering," *Structural, Syntactic, and Statistical Pattern Recognition*, pp. 32-41 (2014).
- [6] David L. Davies, Donald W. Bouldin: "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224-227 (1979).