

2023 年度

修 士 論 文

Seq2seq を用いた複合サービス推薦

指導教員： 村上 陽平

立命館大学大学院 情報理工学研究科
情報理工学専攻 博士課程前期課程
計算機科学コース

学生証番号： 6611220049-6

氏名： HONG Xinyu

Seq2Seq を用いた複合サービス推薦

HONG Xinyu

内容梗概

現在、インターネット上には多数の Web サービスが存在するが、個人や企業のニーズに合致するサービスを見つけるのは難しい。この問題を解決するには、要求に合致するように、既存の Web サービスを組み合わせた複合サービスを構築するサービス合成が有用である。これまでサービス合成の手法は数多く提案されており、主に 2 通りに分類される。一つは、複合サービスのプロセスに機能的に適合する原子サービスを同等なサービス群から選択する水平型サービス合成である。もう一つは、ユーザのゴールを満たす複合サービスのプロセスを組み立てる垂直型サービス合成である。しかしながら、前者は複合サービスを構成するプロセスを事前に与える必要があり、後者は各原子サービスの入出力に加えて、実行するための前提条件や効果までもが論理式で定義される必要がある。また、そもそもユーザがゴールを論理式で定式化することが困難である。

そこで、本研究では seq2seq モデルを使用して、自然言語によるユーザのゴール記述から、そのゴールを満たす複合サービスに必要なサービス集合を出力する手法を提案する。具体的には、複合サービスの説明文を入力、その複合サービスを構成する原子サービス集合を出力とする訓練データを用いて Seq2Seq モデルを学習する。Seq2Seq のエンコーダーには BERT モデルを、デコーダーには Transformer モデルを用いて、シーケンス生成タスクの訓練を行う。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。

訓練データの拡充

Web サービスレポジトリの Programmableweb から取得される複合サービスのデータは数が限られており、高精度のシーケンス生成モデルを訓練するには十分ではない。予測精度を向上させるために、取得されたデータを基に拡張し、訓練データを拡充する必要がある。

推薦ロジックの構築

Seq2seq モデルは、順序を持つ系列データを出力し、同じ値を繰り返し出力することが可能である。しかしながら、本研究の複合サービス推薦では、複合サービスを構成する原子サービスの系列データではなく、集合データを出力する必要がある。集合のため、出力に重複する Web サービスが存在してはならない。そこで、Seq2seq モデルを使用して、推薦候補となるサービスの

集合データを出力する推薦ロジックを構築する必要がある。

一つ目の課題に対して、複合サービスだけでなく、原子サービスも単一サービスから構成された複合サービスとみなして訓練データを拡充する。さらに、ProgrammableWeb のデータは、複合サービスで組み合わせられる原子サービスの実行順序が明記されていないため、同じ複合サービスの説明文であっても、異なる順序のサービス系列を生成できる。この手法により、データ量を大幅に拡充し、モデルは単一のサービスの機能だけでなく、異なるサービスの組み合わせから生じる潜在的なパターンを識別する能力も得ることができる。

二つ目の課題に対して、本研究は最も一般的な二つの出力選択方法を使用して比較実験を行う。出力選択方法に応じて、予測ロジックをそれぞれ適切に修正する必要がある。予測 token を選択した後、選択された token を記録し、後続の予測時に以前に選択された token に遭遇した場合、その確率を修正するか、または直接無視して他の token を選択する。

提案手法による訓練データの拡充と予測モデルのチューニングを Seq2Seq に適用し、ProgrammableWeb の 5604 件の訓練データで学習し、30489 件のテストデータを用いて評価を行い、提案手法の有効性を検証した。本研究の貢献は以下の通りである。

seq2seq モデルに基づくデータ処理

seq2seq モデルに基づくデータ処理：提案手法によるデータ処理を通じて、データ量を大幅に増加させ、データの種類を豊かにする、データは最初の 5604 件から 30489 件に拡充した。提案手法によって処理されたデータは、処理される前のデータよりも精度が向上する。

サービス組み合わせ推薦のための予測ロジック

サービス組み合わせ推薦のための予測ロジック：提案手法による予測ロジックの調整を行い、調整後の予測ロジックは、予測の適合率は 0.12 から 0.21 に上昇し、再現率は 0.09 から 0.96 に上昇しました。

Seq2Seq-Based Composite Service Recommendation

HONG Xinyu

Abstract

With the vast number of Web services on the internet, finding ones that meet specific personal or corporate needs can be challenging. Service composition addresses this by combining existing Web services to create composite services tailored to particular requirements. There are mainly two types of service composition: horizontal, which selects functionally appropriate atomic services, and vertical, which assembles services to achieve user goals. The former requires a predefined process, while the latter needs detailed definitions of service attributes. Formulating these requirements into logical expressions is often difficult for users.

In this research, we propose a method using the seq2seq model to generate a set of services necessary for a composite service that fulfills user goals described in natural language. Specifically, we train the Seq2Seq model using training data that takes the description text of composite services as input and outputs the set of atomic services that constitute those services. The Seq2Seq encoder employs the BERT model, and the decoder uses the Transformer model for training in the sequence generation task. There are two main challenges that need to be addressed for the realization of this method.

Expansion of Training Data

The data of composite services obtained from the Web service repository ProgrammableWeb is limited in quantity and not sufficient for training a high-precision sequence generation model. To enhance prediction accuracy, it's necessary to expand and augment the training data based on the acquired data.

Construction of Recommendation Logic

The Seq2seq model outputs ordered sequential data and can repeat the same value. However, in this study's composite service recommendation, it's necessary to output a set of data of atomic services constituting the composite service, rather than sequential data. As it's a set, no duplicate web services should exist in the output. Therefore, there is a need to construct recommendation logic using the Seq2seq model to output a set of service candidates.

To address the first challenge, the training data is expanded by considering not only composite services but also atomic services as composite services made of single services. Moreover, since the ProgrammableWeb data does not specify the execution order of atomic services in a composite service, it is possible to generate different service sequences with the same composite service description. This method significantly increases the amount of data, allowing the model to identify potential patterns arising from different service combinations, in addition to the functionalities of individual services.

To address the second challenge, this study conducts comparative experiments using the two most common output selection methods. Depending on the output selection method, it's necessary to appropriately modify the prediction logic. After selecting a prediction token, the selected token is recorded. If a previously selected token is encountered during subsequent predictions, its probability is either adjusted or directly ignored in favor of selecting another token.

The proposed method involves augmenting training data and tuning the prediction model for Seq2Seq. The model was trained with 5,604 training data items from ProgrammableWeb and evaluated using 30,489 test data items to verify the effectiveness of the proposed method. The contributions of this research are as follows.

Data Processing Based on Seq2Seq Model

The passage discusses data processing using a seq2seq model. The method significantly increases the quantity of data, enriching the variety of data types. Specifically, the data expanded from an initial count of 5,604 items to 30,489. The processed data, as a result of the proposed method, exhibits improved accuracy compared to the data before processing.

Predictive Logic for Service Combination Recommendation

The passage discusses the adjustment of predictive logic for recommending service combinations according to the proposed method. After the adjustment, the predictive logic aligns well with the experimental design of this study, indicating improved suitability and effectiveness in the context of the research.

目次

| | |
|--------------------------------|----|
| 第1章 はじめに | 1 |
| 第2章 複合サービス推薦 | 3 |
| 2.1 キーワードに基づく複合サービス推薦 | 3 |
| 2.2 説明文類似度に基づく複合サービス推薦 | 3 |
| 第3章 SEQ2SEQ に基づく複合サービス推薦 | 6 |
| 3.1 訓練データ | 6 |
| 3.1.1 データ構造 | 6 |
| 3.1.2 データ処理 | 7 |
| 3.1.3 データ拡充 | 8 |
| 3.2 BERT エンコーダ | 9 |
| 3.3 TRANSFORMER デコーダー | 11 |
| 3.4 単語リストの作成 | 13 |
| 3.5 損失関数 | 14 |
| 第4章 モデル予測 | 16 |
| 4.1 予測を生成する前の準備 | 16 |
| 4.2 予測手法 | 18 |
| 4.2.1 Greedy-search | 18 |
| 4.2.2 Beam-search | 20 |
| 第5章 評価 | 22 |
| 5.1 評価手法 | 22 |
| 5.1.1 適合率 | 22 |
| 5.1.2 再現率 | 24 |
| 5.2 実験データ | 26 |
| 5.2.1 データ拡充の結果 | 27 |
| 5.2.2 データセット分割 | 29 |
| 5.3 結果 | 31 |

| | |
|---|----|
| 第6章 考察 | 36 |
| 6.1 最大生成長 | 36 |
| 6.2 カテゴリを用いたモデルの訓練 | 38 |
| 6.3 BERT MUTI-LABEL CLASSIFIER を用いた複合サービス推薦 | 41 |
| 第7章 まとめ | 43 |
| 謝辞 | 45 |
| 参考文献 | 46 |

第1章 はじめに

現在、インターネット上には多数の Web サービスが存在するが、個人や企業のニーズに合致するサービスを見つけるのは難しい。この問題を解決するには、要求に合致するように、既存の Web サービスを組み合わせた複合サービスを構築するサービス合成が有用である。これまでサービス合成の手法は数多く提案されており、主に 2 通りに分類される。一つは、複合サービスのプロセスに機能的に適合する原子サービスを同等なサービス群から選択する水平型サービス合成である。もう一つは、ユーザのゴールを満たす複合サービスのプロセスを組み立てる垂直型サービス合成である。しかしながら、前者は複合サービスを構成するプロセスを事前に与える必要があり、後者は各原子サービスの入出力に加えて、実行するための前提条件や効果までもが論理式で定義される必要がある。また、そもそもユーザがゴールを論理式で定式化することが困難である。

そこで、本研究では seq2seq モデルを使用して、自然言語によるユーザのゴール記述から、そのゴールを満たす複合サービスに必要なサービス集合を出力する手法を提案する。具体的には、複合サービスの説明文を入力、その複合サービスを構成する原子サービス集合のペアを出力とする訓練データを用いて Seq2Seq モデルを学習する。Seq2Seq のエンコーダーには BERT モデルを、デコーダーには Transformer モデルを用いて、シーケンス生成タスクの訓練を行う。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。

訓練データの拡充

Web サービスレポジトリの Programmableweb から取得される複合サービスのデータは数が限られており、高精度のシーケンス生成モデルを訓練するには十分ではない。予測精度を向上させるために、取得されたデータを基に拡張し、訓練データを拡充する必要がある。

推薦ロジックの構築

一般的な Seq2seq モデルの予測結果は、順序を持ち、繰り返すことが可能である。しかし、本研究では順序に関心がなく、単に複合サービス説明文に適合する基本 Web サービスの集合を探し出すことに焦点を当てる。このサービス集合には重複する Web サービスが存在してはならない。したがって、予測のロジックに変更を加える必要がある。モデルを使用して予測生成を行う際には、相応の制約を設ける必要があり、一般的な予測トークンの選択方法

についても，本実験の要件に応じて適切に修正する必要がある．

以下，本論文では，2章において従来の研究の複合サービスの推薦手法について説明する．次に3章は本研究で用いた研究データや，seq2seqモデルの構造について説明する．続いて4章は訓練済みモデルを用いた予測ロジックについて説明する．5章は評価手法と評価結果について説明する．6章は考察を行い，7章は今後の展望や課題について述べて結論とする．

第2章 複合サービス推薦

本章では、現代情報技術分野においてますます重要になっている複合サービス推薦方法について探求する。複合サービス推薦とは、ユーザの具体的なニーズに基づいて最適なサービス組み合わせを提案する過程であり、この過程ではユーザのニーズを正確に理解し、多数のサービスから最も適切なものを選び出し、業務効率化と顧客満足度の向上に貢献する。

2.1 キーワードに基づく複合サービス推薦

キーワードに基づく複合サービス推薦は、直感的で広く応用されている方法であり、ユーザが入力したキーワードに基づいてサービスを識別し推薦することを目的とする。初めにユーザはそのニーズに基づいて一つまたは複数のキーワードを入力する。これらのキーワードは通常、必要なサービスの簡潔な説明である。次にシステムはユーザが入力したキーワードをもとに、サービスデータベースや索引で検索を行う。

この過程は、キーワードとサービス説明の間の関連性のマッチングを含む。その後、システムは一定の基準（例えば関連性など）に基づいて検索結果を並べ替え、ユーザに推薦する。この方法の最大の利点は以下の二つである：

易用性：ユーザはキーワードを入力するだけで検索を開始でき、操作は簡単で直観的である。

迅速な応答：テキストベースの検索アルゴリズムは一般に速いため、すばやく結果を返すことができる。

しかし、この方法の欠点も明らかである。この方法はキーワードの質に大きく依存し、検索の質はユーザが入力するキーワードの正確さや具体性に高度に依存する。また、複雑なユーザのニーズを処理するのが難しく、キーワードが一般的な場合、関連性の低い多数の結果を返す可能性がある。

2.2 説明文類似度に基づく複合サービス推薦

説明文類似度に基づく複合サービス推薦は、より精密で複雑なサービス推薦方法である。これは、サービスの説明文書とユーザのニーズとの間の類似度を分析し比較することでサービスを推薦する。まずユーザはサービスニーズを記述し、これは通常自然言語テキストの形式で行われる。次にサービスデータベース

に保存されているサービスの説明文書を分析する。これらの文書にはサービスの詳細な説明が含まれている。次にテキスト類似度アルゴリズム（例えばコサイン類似度、ジャカード類似度など）を使用して、ユーザのニーズとサービスの説明文書との間の類似度を比較する。最後に計算された類似度に基づいてサービスを選別し並べ替え、ユーザに推薦する。

この方法はキーワードベースの方法と比較して、ユーザのニーズとサービスの機能をより正確にマッチさせることができる。また、この方法は、特に具体的なサービス特性や詳細なニーズに関わる場合など、より複雑なユーザのクエリを処理することができる。しかし、この方法の欠点も顕著である。この方法の効果は、サービスの説明文書の質と詳細さに大きく依存している。しかし、サービスの説明文はサービスの作者によって異なり、現在のインターネット上ではサービスの説明文に対する明確な要件がないため、各作者は自分の習慣に従って説明文を書いており、説明文の正確性は保証されていない。また、この方法では、既存の複合サービスからのみ選別・並べ替えを行い、ユーザのニーズにより適合する新たなサービス組み合わせを推薦することはできない。

複合サービス推薦システムの向上方策を探る中で、まずは二つの主流推薦方法、キーワードに基づく推薦と説明文類似度に基づく推薦を見直す。

キーワードに基づく推薦方法は、そのシンプルさと速い反応速度で広く利用される。ユーザがキーワードを入力することで迅速に関連サービスを検索でき、操作が簡単であるが、推薦の質はユーザの入力するキーワードの質に大きく依存する。一般的またはあいまいなキーワードは、不正確な推薦結果やユーザのニーズと合わない結果をもたらす可能性がある。

一方で、説明文類似度に基づく推薦方法は、サービスの説明とユーザのニーズの間の類似度を分析してサービスを推薦する。これにより、より精密なマッチング結果を提供することができるが、この方法の効率と正確さはサービスの説明の包括性と正確さに依存し、複雑であいまいなユーザのニーズを処理する際には効果が不十分なことがある。

これらの方法の限界を考慮し、シーケンス・トゥ・シーケンス (seq2seq) モデルを新たな解決策として導入する。Seq2seq モデルは、深層学習に基づくエンドツーエンドのモデルで、特に機械翻訳分野で広く用いられており、可変長の入出力シーケンスを扱い、複雑なシーケンス間のマッピング関係を学習する能力がある。これを複合サービス推薦に適用することで、複合サービスの詳細な説明

文を分析し、異なる基本サービスコンポーネント間の関係と組み合わせパターンを学習し、ユーザの具体的なニーズに合った基本サービスの組み合わせを推薦することができる。

Seq2seq モデルを用いる主な利点は、その強力な学習能力と柔軟性にある。モデルはサービス説明の深層的な意味情報を自動的に捉え、ユーザのニーズの微妙な違いを理解し、よりパーソナライズされた正確なサービス組み合わせの推薦を生成することができる。さらに、seq2seq モデルの自己反復能力により、推薦システムはユーザのフィードバックから学習を続け、推薦の精度とユーザの満足度を向上させることができる。

以上のように、キーワードと説明文類似度に基づく推薦方法はそれぞれに利点があるが、複雑で多様なユーザのニーズに直面した時、それらの限界も明らかになっている。Seq2seq モデルの導入は、複合サービス推薦における新たな視点と方法を提供し、深層学習技術の適用により、伝統的な方法の不足を克服し、よりインテリジェントでパーソナライズされた推薦サービスを提供することができ、複合サービス推薦システムの未来の発展の新たな方向を示している。

第3章 Seq2Seq に基づく複合サービス推薦

Seq2Seq (Sequence to Sequence) は、深層学習モデルの一種で、主にシーケンスからシーケンスへの変換タスクを処理するために使用される。これには、機械翻訳、テキスト要約、質問応答システムなどが含まれる。Seq2Seq モデルは通常、2つの部分から構成される：エンコーダとデコーダ。

本研究では、Seq2Seq モデルを使用して、複合ネットワークサービスの説明文をそのサービスを構成する基本 web サービスの ID 集合に変換するモデルを訓練する。このタスクは、実際にはシーケンスからシーケンスへの変換であり、本質的には伝統的な Seq2Seq アプリケーション (例えば機械翻訳) に似ている。

本章では、この方法を使用するためのデータ準備、モデル設計、単語表の構築、およびモデルの訓練について説明する。

3.1 訓練データ

3.1.1 データ構造

訓練データに用いる原子サービスのデータ構造と複合サービスのデータ構造を図1と図2に示す。

```
"サービス名前": [  
  [  
    サービスのカテゴリ  
  ],  
  サービスの説明文  
  [  
    このサービスを使って、組合せた複合サービス集合  
  ],  
  サービスID  
]
```

図1：原子サービスのデータ構造

```

"複合サービス名前": [
  [
    複合サービスのカテゴリ
  ],
  複合サービスの説明文
  [
    この複合サービスを組合せた原子サービスの集合
  ],
  複合サービスID
]

```

図 2 : 複合サービスのデータ構造

3.1.2 データ処理

本実験で必要とされるモデルの訓練は，ユーザのニーズを入力とし，それに
 応じた原子サービス集合のサービス ID を出力することを目指している。
ProgrammableWeb から取得したデータに対して適切なデータ処理を行うこと
 が必要である。取得した複合ネットワークデータには，複合ネットワークサービ
 スの名前，カテゴリ，説明文，その複合サービスを構成する基本サービス集合，
 およびユニークな識別子が含まれている。ニーズに基づいて，このデータに対し
 て 2 つの処理が必要である。まず，複合サービスの組み合わせのユニークな識別
 子と複合サービスの名前は不要であるため，削除する必要がある。カテゴリは後
 のテスト時に必要となる可能性があるため，保持する。次に，取得したデータに
 は複合サービスを構成する基本サービス集合の名前が既に含まれているが，本
 実験では最終的にこの基本サービス集合の各サービスに対応する ID 集合を出力
 することを目指しているため，この部分も変更する必要がある。変更したのデー
 タ構造は図 3 に示された。

```

{
  "mashup_desc": 複合サービスの説明文
  "services": [
    この複合サービスを組合せた原子サービスのID集合
  ],
  "category": カテゴリ
}

```

図 3 : 訓練データのデータ構造

3.1.3 データ拡充

データ量から見ると、この 5604 件のデータ量は明らかに不足しており、理想的なモデルを訓練するための支持を得るには足りない。そのため、データの拡張が必要である。本研究では、データセットを拡充するために 2 つの方法を使用している。まず、ProgrammableWeb から取得したデータを観察すると、多くの複合サービスは単一の web サービスで構成されていることがわかる。つまり、取得した基本 web サービスを単一の web サービスで構成される複合サービスとして見ることができる。基本 web サービスのデータには、web サービスの名前、カテゴリ、説明文、そのサービスを使用して構成される複合 web サービスの名前および ID が含まれている。これらの中から、サービスの説明文、カテゴリ、ID を抽出して訓練データとして使用することができる。この操作には以下のような利点がある：

データ量の増加：訓練セットを拡大することで、モデルの学習効果と汎化能力を改善する。

モデル理解の強化：単一のサービスで構成される複合サービスを含むことで、モデルは単一サービスと複合サービスの関係をよりよく学習し理解することができる。**ロバスト性の向上：**データの多様性を増やすことで、モデルは様々な入力状況に対してよりロバストに対応することができる。

複雑度の簡素化：単一サービスは複合サービスの特例として、より複雑な複合サービスの組み合わせを処理する際に、より堅固な基盤をモデルに提供する。

総じて、この方法はデータセットの豊かさと多様性を高め、モデルの性能と精度を向上させることができる。

Seq2Seq モデルを使用して生成されるシーケンスには順序が存在するが、本実験の目的はユーザのニーズを満たす複合サービスを構成する基本サービスを出力することであるため、実際には順序に対する要求はない。そのため、トレーニングセットのデータを同じ内容で順序を変更してデータセットを拡張することが可能である。例えば、複合サービス A のデータが図 4 のような場合：

```
{
  "mashup_desc": 説明文
  "services": [
    サービスb,
    サービスc
  ],
  "category": カテゴリ
}
```

図4：例のデータ構造

このような1つのデータを、サービスの順序を変更することで2つのデータに拡張することができる。新しいデータは図5のようになる：

```
{
  "mashup_desc": 説明文
  "services": [
    サービスc,
    サービスb
  ],
  "category": カテゴリ
}
```

図5：新しいデータ

サービスの順序を変更してデータを拡充する方法では、単一の基本サービスで構成される複合サービスには拡充を行わず、複数の基本サービスで構成される複合サービスにのみ拡充を行う。本実験では、実際の試験により、2つまたは3つの基本サービスで構成される複合サービスを拡充することが最も効果的であることが判明した。

3.2 BERT エンコーダ

BERTは、Transformerアーキテクチャに基づくエンコーダである。その設計は、テキストの深層双方向表現を生成するために主に用いられる。従来の一方向

または浅層の言語モデルとは異なり、**BERT** は双方向のトレーニングを通じて、より包括的に言語コンテキストを理解する。この構造により、テキスト分類、固有名詞認識、質問応答システムなど、さまざまな下流の自然言語処理タスクに特に適している。これらのタスクにおいて、**BERT** は通常、強力な特徴抽出器として使用され、テキストの深層意味特徴を抽出する。

本研究では、以下のいくつかの利点があるため、エンコーダとして **BERT** を使用する：

深層意味理解：**BERT** の双方向構造により、テキストのコンテキストをより深く理解することができ、入力テキストの微妙な意味の違いを捉えることができる。

強力な事前訓練知識：大量のデータで事前訓練された **BERT** は、これらの学習した一般的な言語知識を特定のタスクに適用し、モデルの全体的な性能を向上させることができる。

高品質な特徴表現：**BERT** が生成する特徴表現の品質が高く、言語の多様性と複雑性をよりよく捉えることができ、これは後続のデコードプロセスに不可欠である。

異なる文脈への適応：**BERT** の双方向特性により、さまざまな文脈や言語使用に適応することができ、複雑なサービス説明テキストの理解に特に重要である。

訓練時間の短縮：広範囲にわたって事前訓練されている **BERT** をエンコーダとして使用することで、特定のタスクに対する訓練時間やリソースの消費を減らすことができる。

総合すると、**BERT-BASE** をエンコーダとして使用することで、**Seq2Seq** モデルに強力な言語処理能力を提供し、複雑な自然言語タスクを処理する際のモデルのパフォーマンスを向上させることができる。**BERT** エンコーダが本研究の流れは図6に示された。

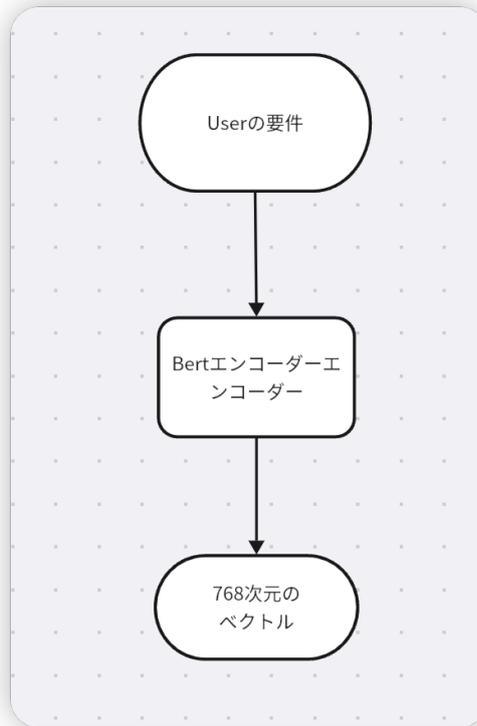


図6 : BERT エンコーダの流れ

3.3 TRANSFORMER デコーダー

Transformer デコーダーは Transformer モデルの重要な構成部分で、シーケンスの出力を生成するために特化している。自己注意力メカニズムなど、Transformer アーキテクチャの先進技術を活用し、シーケンスデータを効果的に処理する。従来のシーケンス生成方法と異なり、Transformer デコーダーは、生成されたシーケンスの全文脈情報を同時に考慮し、次のシーケンス要素をより正確に生成する能力がある。

このような Transformer デコーダーの構造と機能により、複雑な言語生成タスクにおいて優れた性能を発揮し、特に入力データと密接に関連し意味的に連続性のあるテキストを生成するのに長けている。このため、現代の自然言語処理分野において重要なツールとなっている。

本実験の seq2seq モデルでは、入力は複合サービスの説明文で、出力はその複合サービスを構成する基本サービスの ID 集合である。Transformer デコーダーを使用することの利点は以下の通りである：

正確なシーケンス生成 : Transformer デコーダーは文脈情報を十分に考慮するため、入力説明文に対応するサービス ID シーケンスをより正確に生成できる。

長距離依存の処理 : Transformer デコーダーは長距離依存関係を特にうまく処理するため、長文から特定のサービス ID を抽出するのに非常に有効である。

高速かつ効率的 : Transformer の並列処理能力により、大量のデータを処理する際にモデルが出力シーケンスをより迅速に生成でき、全体的な効率が向上する。

結果の一貫性の向上 : 他のデコーダーと比較して、Transformer は生成されたテキストの一貫性をよりよく維持でき、これはサービス ID シーケンスの正確性と論理性を保持するために非常に重要である。

以上のように、複雑なサービス説明文を処理し、精確なサービス ID 集合を生成するタスクにおいて、Transformer デコーダーを使用することは、効率的で正確かつ一貫した性能を提供する。本研究の流れは図 7 に示された。

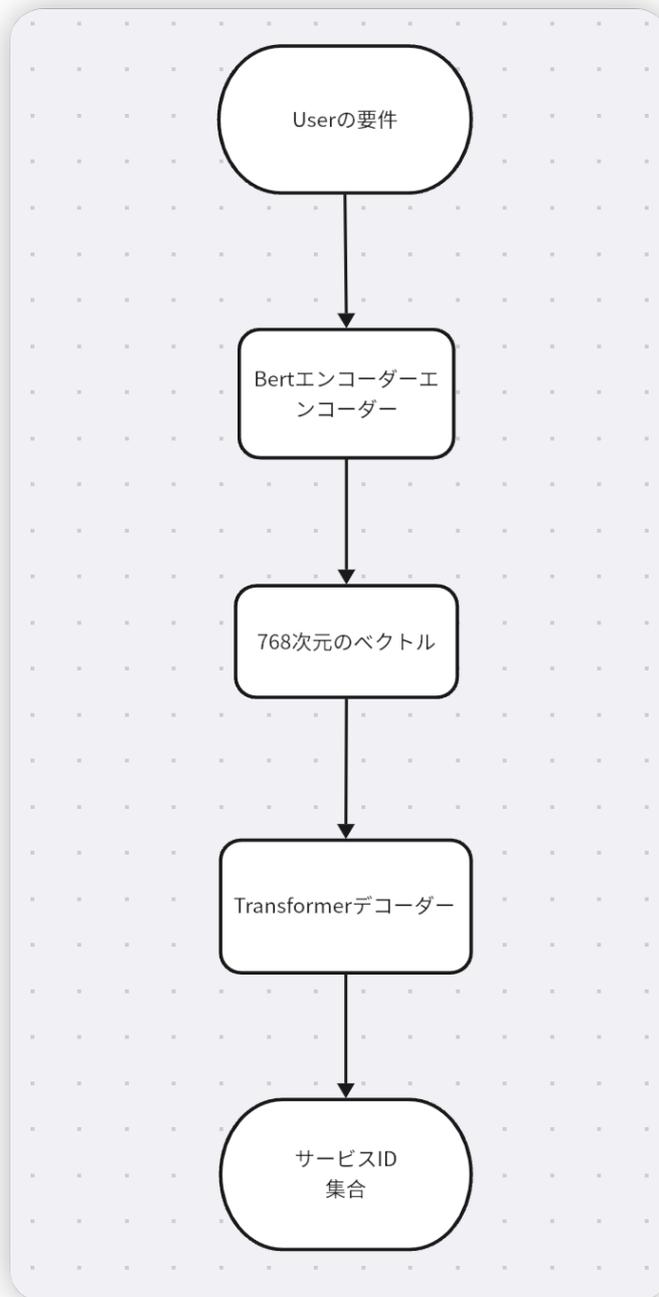


図7：本研究の流れ

3.4 単語リストの作成

本研究で訓練されたモデルは、最終的にユーザーのニーズに合ったID集合を出力する。この目標を達成するためには、最終出力の単語リストを自ら構築する

必要がある。モデルを使用して予測する際、選択される出力はすべて単語リストから選ばれる。

単語リストを初期化するには、まず 3 つの特別なマーカーを追加する：[PAD]は埋め込みに使用され、[BOS]は開始を示し、[EOS]は終了を示す。したがって、モデルの訓練時にも同様に、各トレーニングデータに開始と終了のマーカーを追加し、[PAD]を使用して各データの説明文を同じ長さにする。

表 2：単語リストの構造

| 単語 | ID |
|------------|-----|
| [PAD] | 0 |
| [BOS] | 1 |
| [EOS] | 2 |
| ServiceID1 | 3 |
| ServiceID2 | 4 |
| serviceID3 | 5 |
| serviceID4 | 6 |
| ... | ... |

3.5 損失関数

本研究では、交差エントロピー損失関数（Cross-Entropy Loss）を、神経ネットワークの訓練における主要な損失関数として採用している。交差エントロピー損失関数は一般的に分類問題に使用されるが、私たちのシーケンス生成タスクにも同様に適用可能である。これは、タスクをシーケンス内の各位置でのサービス ID の分類と見なすことができるためである。

具体的には、神経ネットワークモデルの目標は、与えられたサービス記述テキストに基づいて、そのサービスを構成する一連の基本サービス ID を予測することである。このプロセスは、シーケンス内の各時間ステップでの分類と見なすことができ、各時間ステップでは、事前に定義された単語リスト（可能なすべてのサービス ID を含む）から正しい ID を選択する必要がある。

交差エントロピー損失関数の役割は、モデルが予測したサービス ID シーケンスと実際のサービス ID シーケンスとの間の差異を計算することである。モデルが生成する各サービス ID 予測に対して、損失関数は予測確率と実際のデータと

の間の不一致の程度を評価し、モデル訓練過程でこの不一致を徐々に減少させる。また、シーケンスの長さの不一致を考慮して、入力シーケンスにパディング (padding) を施し、それらが統一された長さを持つようにした。損失を計算する際には、パディングされた位置の予測は無視され、モデルの性能に悪影響を及ぼさないようにする。

総じて、交差エントロピー損失関数は、予測誤差を量化し最小化する効果的な方法を提供し、モデルがより正確にサービス ID シーケンスを生成する方法を学習するための指針となる。

第4章 モデル予測

本章では，訓練済みの Seq2Seq モデルを使用して，一般的に使われる二つの予測方法を紹介する．

4.1 予測を生成する前の準備

Seq2Seq モデルを使用してシーケンスを生成する前に，いくつかの必須のステップがある．

BERT 初期化 : BertTokenizer.from_pretrained('bert-base-uncased') および BertModel.from_pretrained('bert-base-uncased') を使用して，BERT のトークナイザーとモデルを初期化する必要がある．

単語リストの読み込み : vocab_file ファイルを読み込むことで，単語リストをロードし，単語からインデックスへのマッピングとインデックスから単語へのマッピングを作成する必要がある．

Transformer モデルの定義 : モデルを訓練するときと同じの Transformer の定義が必要である．

訓練済みモデルのロード : 訓練済みのモデル状態をロードし，それを定義済みのモデルに適用する．

本研究の予測基本流れは図 8 に示された．

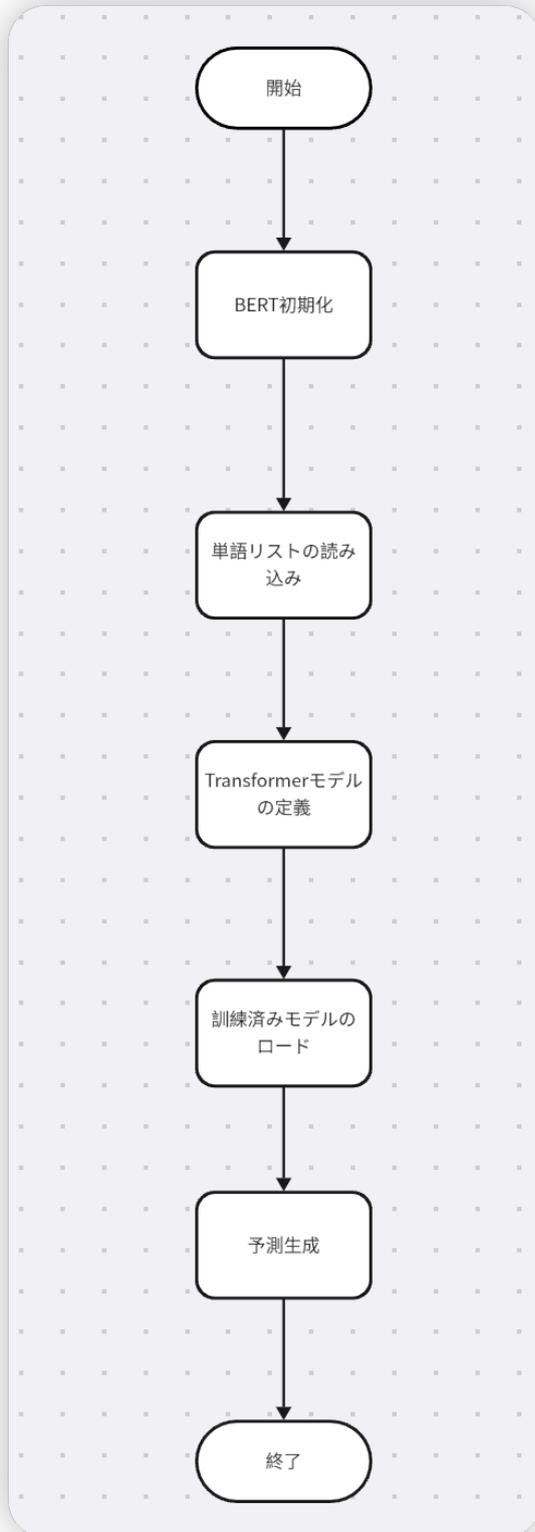


図 8 : 予測の基本流れ

4.2 予測手法

本実験の目的は、ユーザーのニーズを満たす基本サービスの組み合わせを予測することである。したがって、最初に従うべき原則は、予測される単語が重複しないことである。これは最も基本的な予測要求であり、実際の実験では、単語を予測するごとに、既に予測された単語をリストに加え、その確率を極めて低く設定する。これにより、後続の予測で再び選択されることはない。一般的なシーケンス生成の予測手法には 2 種類ある。一つは `greedy-search`、もう一つは `beam-search` である。

4.2.1 Greedy-search

グリーディサーチは、機械学習において、特に自然言語処理のタスクで広く使用される方法である。確率モデル（例えば `Seq2Seq` モデル）からシーケンスを生成するための単純だが効果的な手法である。

グリーディサーチの流れは以下の通り：

開始トークン：プロセスは、通常 NLP タスクで「文の始まり」を意味する初期入力から始まる、本研究の始まりは「**BOS**」になる。

モデル予測：このトークンを入力としてモデルが受け取り、次のトークンについての可能性の分布を生成する。

最も可能性の高いトークンの選択：この分布から、最も確率の高いトークンがシーケンスの次の要素として選ばれる、これはグリーディサーチの最も重要な部分である。

反復プロセス：次に、このトークンが次の予測のためのモデルの入力にフィードバックされる。このプロセスは繰り返し行われる。

シーケンス生成：最大シーケンス長に達するか、モデルが終了トークン（本実験は「**EOS**」）を生成するまで、このプロセスが続く。

最終出力：このプロセスを通じて生成されたトークンのシーケンスが最終的な予測として出力される。

この実験では、ユーザーのニーズに応える基本サービスセットを出力する必要があるため、出力されるシーケンスの各トークンは重複してはならない。そのため、貪欲探索を使用する際には、通常の貪欲アルゴリズムを少し変更する。通常の貪欲アルゴリズムでは、最も確率の高いトークンを継続的に選択して出力するが、本研究の出力要件を満たすために、出力される各トークンをある集合に追加する。そして、最も確率の高いトークンを選択する際に、まずそのトークン

が集合内にあるかどうかをチェックする。もし既に集合内にある場合は、そのトークンを出力せず、その単語を除いて次に確率の高いトークンを選択する。本研究の greedy-search のフローチャートは図9に示された。

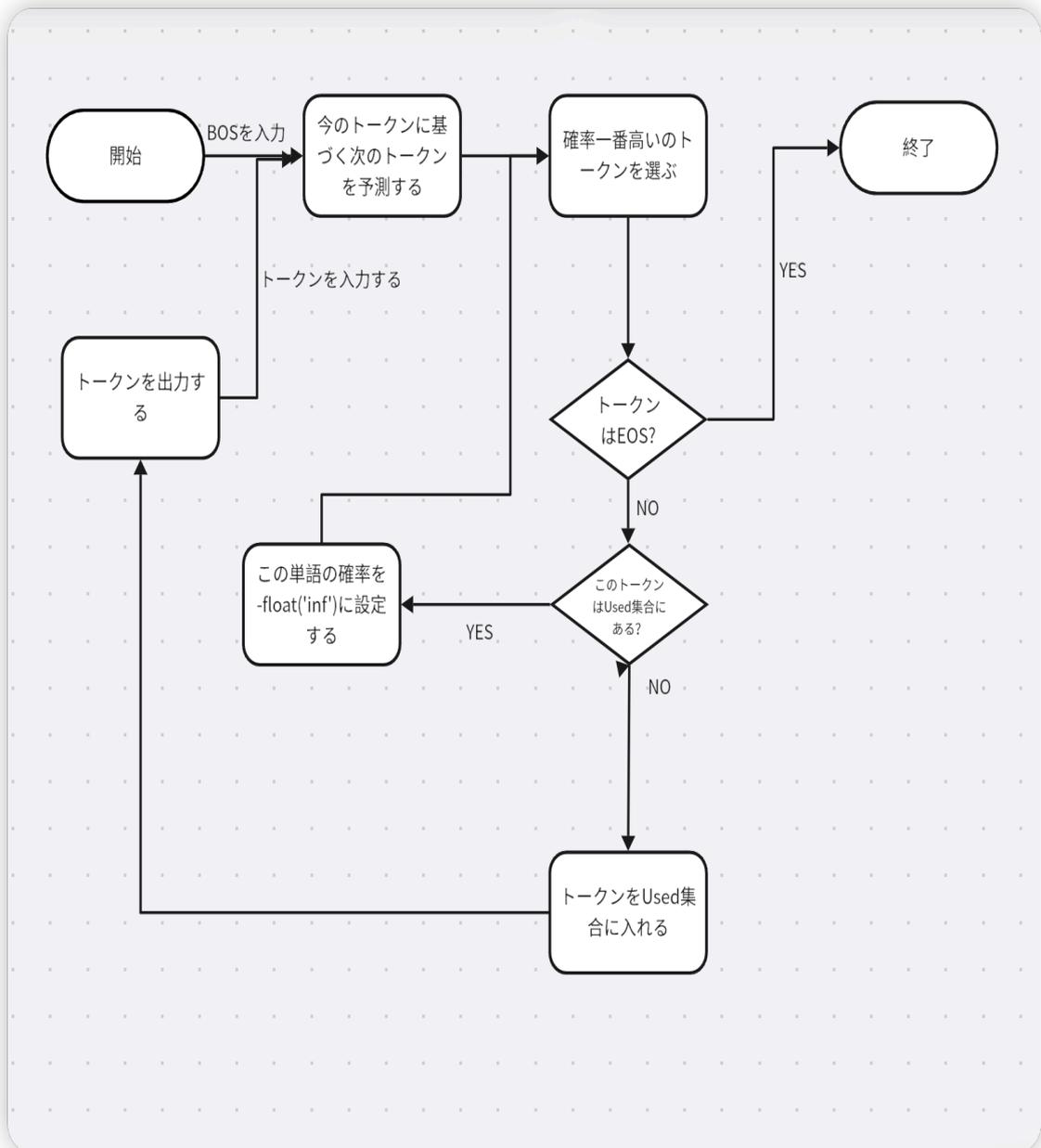


図9 : greedy-search のフローチャート

4.2.2 Beam-search

ビームサーチは、自然言語処理におけるシーケンス生成のためのアルゴリズムであり、特に機械翻訳、音声認識、テキスト生成などの分野で用いられる。効率と品質のバランスを取る戦略であり、最も確率が高い出力シーケンスを見つけることを目的とする。単純な貪欲探索（各ステップで最も確率が高い選択をする）と比べ、ビームサーチは複数の可能な候補シーケンスを考慮する。

ビームサーチの流れは以下の通りである：

幅（ビーム幅）：ビームサーチの核心は「ビーム幅」であり、各ステップでいくつの候補シーケンスを保持するかを決定するパラメーターである。小さいビーム幅は貪欲探索に近く、大きいビーム幅ではより多くの候補を考慮するが、計算コストが高くなる。

シーケンス生成：開始記号（本実験は「[BOS]」）から始め、各ステップで全ての可能な次の記号を生成し、各新しいシーケンスの確率を計算する。

最良候補の保持：各ステップで、アルゴリズムは確率が最も高い「ビーム幅」のシーケンスのみを保持し、他のシーケンスは破棄される。このようにして、アルゴリズムは常に現時点で最も確率が高い一連のシーケンスのみを考慮する。

終了まで繰り返す：シーケンスの最大長に達するか、全ての候補シーケンスが終了記号（本実験は「[EOS]」）に到達するまで、このプロセスを繰り返す。

最良シーケンスの選択：最後に、全ての完了した候補シーケンスの中から最も確率が高いシーケンスを最終的な出力として選択される。

本実験の目的は、ユーザーのニーズを満たす基本サービスの集合を出力することである。したがって、出力されるシーケンスの各トークンは重複してはならない。このため、**beam-search** を使用する際にはいくつかの変更が必要である。通常の **beam-search** では、各ステップで総スコアを最も高くするトークンを保存し、**beam_size** の数だけシーケンスを候補として保存する。しかし、実験の要件に従って、トークンを選択する前にまず重複しないことを保証する必要がある。シーケンスの前半部分で既に出現したトークンは、後の選択で無視され、そのトークンを飛ばして総スコアを最も高くする別のトークンを選び、シーケンスを保存する。最終的には、得点が最も高いトークンを選択して結果として出力する。本研究の **beam-search** のフローチャートは図 10 に示された。

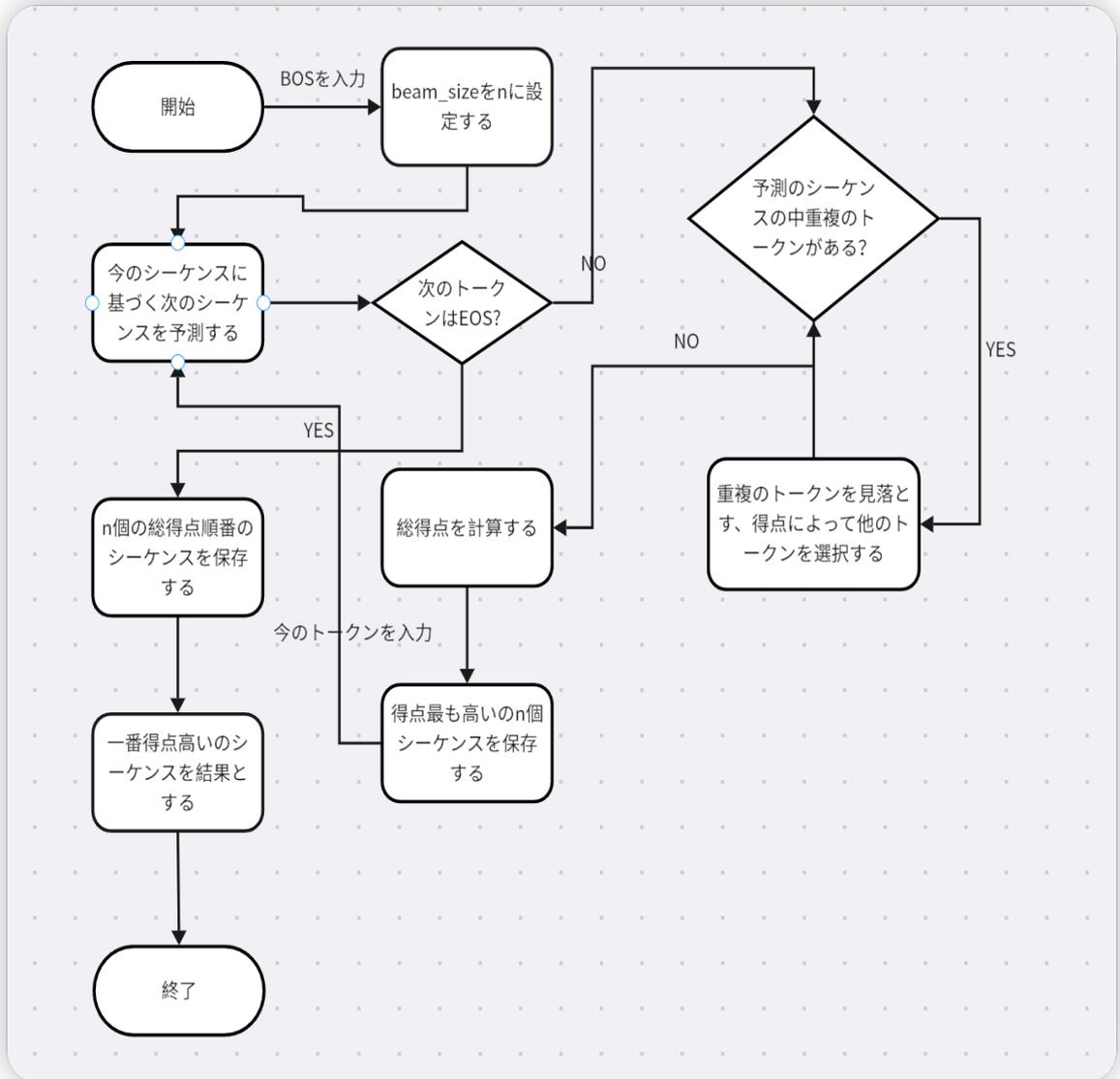


図 1 0 : beam-search のフローチャート

第5章 評価

第3章では、モデル訓練の流れについて述べられている。第4章では、モデルの具体的な使用方法について説明されている。本章では、本研究で使用される2種類の評価手法について主に述べるとともに、その結果を展示する。

5.1 評価手法

本実験で使用される評価手法は、最も一般的な正解率と再現率である。しかし、本実験の目的は、ユーザーが望む複合サービスの要求を入力し、ユーザーの要求を満たすことができる基本サービスの集合を得ることであるため、使用される手法は正解率と再現率であるが、正解と再現の定義は、本実験の要求を満たすために若干の変更を加える。

5.1.1 適合率

適合率 (Precision) は、特に分類タスクや推薦システムにおいて、よく使用される評価指標である。その定義は非常に直感的で、モデルが正しく予測した結果が、その総予測結果に占める割合を測定するために主に使用される。本実験において、適合率は、モデルが予測した結果の中で、どれだけユーザーの要求を満たす基本サービスであるかを意味する。適合率の公式は図11に示された。

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}}$$

図11：適合率の公式

True Positives (TP): モデルが正と正しく予測した数。

False Positives (FP): モデルが正と誤って予測した数。

本実験のテストデータは、500条の複合サービスの説明文と、それらを構成する基本サービスの集合である。伝統的な正解率を用いて計算する場合、予測されたサービスIDと複合サービスの実際の構成IDを比較することになる。しかしこのような比較は明らかに問題がある。なぜなら、本研究の目的は、ユーザーの要求を満たす基本サービスの集合を見つけ出すことであり、特定の要求を満

たす基本サービスは複数存在し、一つに限定されるべきではないからである。例えば、あるユーザーが地図機能を要求している場合、ユーザーの要求を満たす基本サービスは、`google_map` だけに限定されるべきではなく、同様の機能を持つ `bing_map` も正しいとされるべきである。普通の適合率の例は図 1 6 に示された。本研究の適合率は図 1 3 に示された。

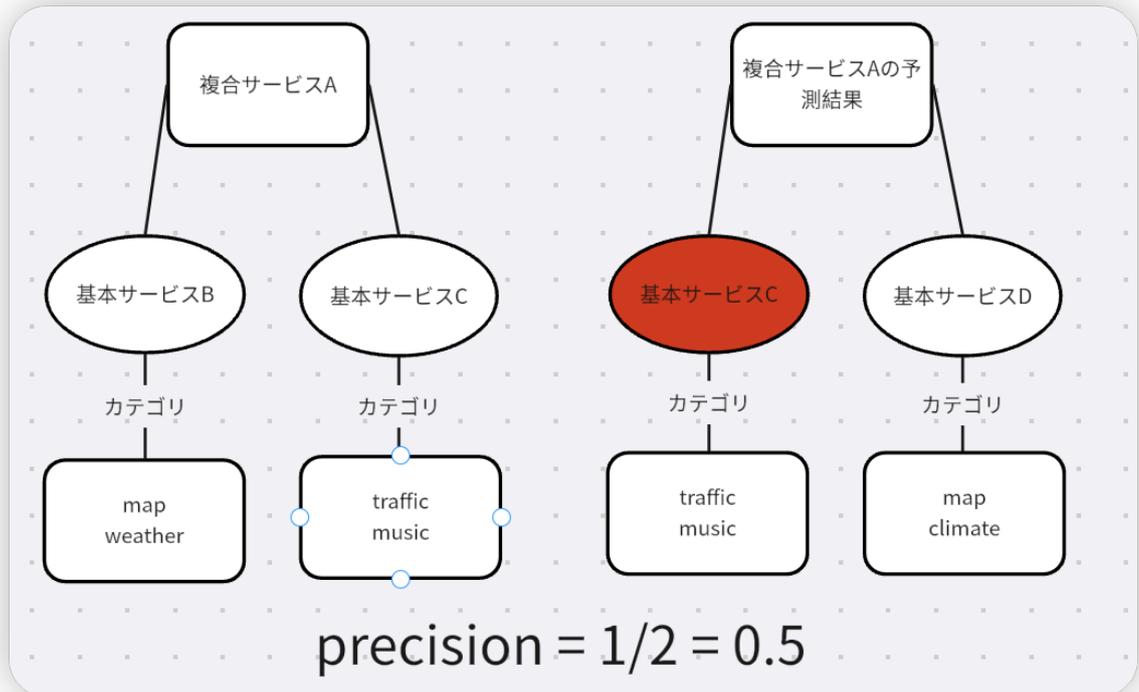


図 1 2 : 適合率

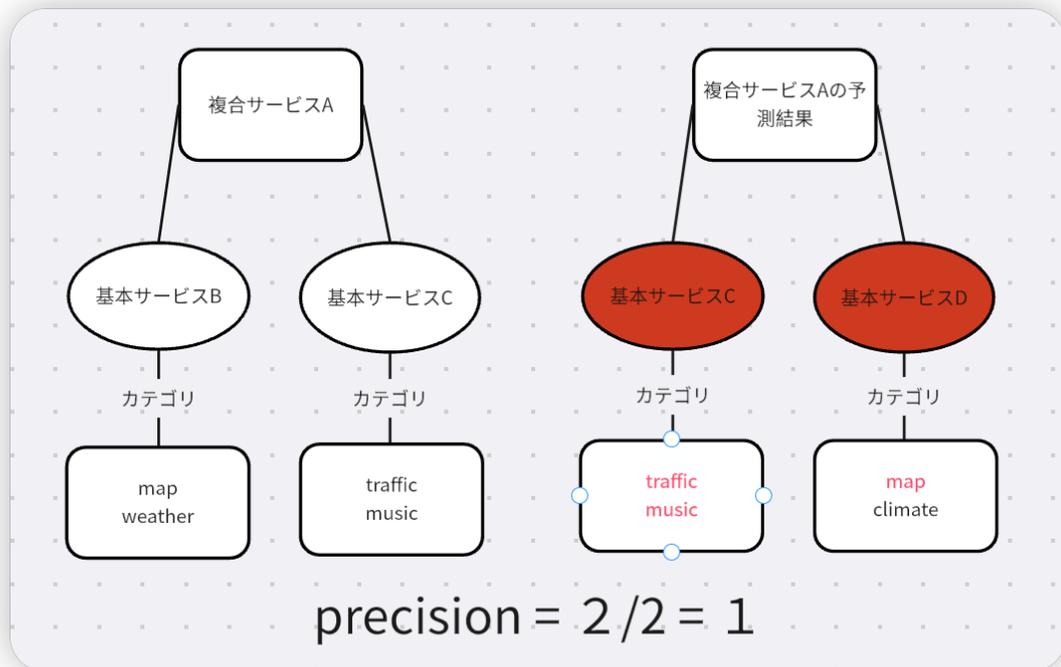


図 1 3 : 本研究の適合率

5.1.2 再現率

再現率 (Recall) は、分類タスク、情報検索、推薦システムなどで広く使用されている重要な評価指標である。その重点は、モデルが関連するすべてのインスタンスを正確に識別する能力を測定することにある。本実験では、再現率は、ユーザーの要求がどれだけ満たされたかを示す。再現率の公式は図 1 4 に示された。

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}}$$

図 1 4 : 再現率の公式

True Positives (TP) : モデルが正しく正例として識別したサンプルの数である。

False Negatives (FN) : モデルが誤って負例として識別したが、実際には正例であるサンプルの数である。

本実験では、再現率の計算も通常の再現率とは異なる。適合率と同様に、本研究の要求を満たすためにいくつかの変更を加える。一般的な再現率の計算では、

複合サービスの実際の基本サービスのカバー率を直接計算する. 例えば, ある複合サービス A が基本サービス B と基本サービス C で構成されているとし, 予測結果が基本サービス C と基本サービス D であった場合, このように直接計算すると, カバーされた基本サービスはサービス C だけ, 再現率は **0.5** となる. しかし実際には, 基本サービス D と基本サービス B はカテゴリーが重なっているの
 で, カバーとみなすことができる. したがって, 本実験ではこの例の再現率は 1 となる. 普通の再現率は図 1 5 に示された. 本研究の再現率は図 1 6 に示された.

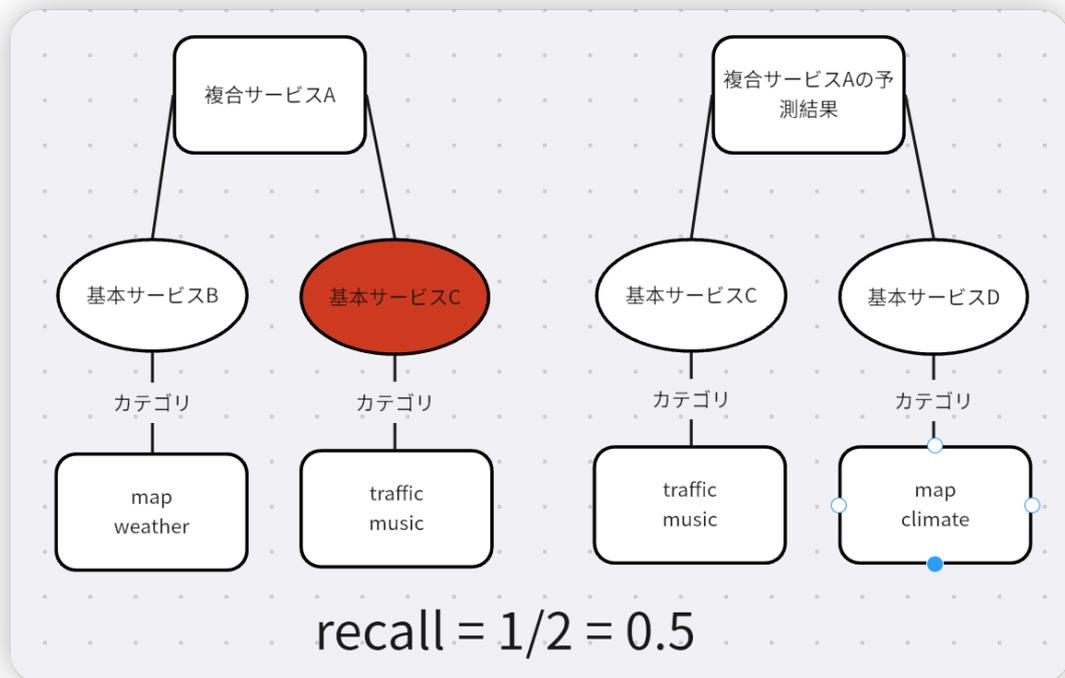


図 1 5 : 再現率

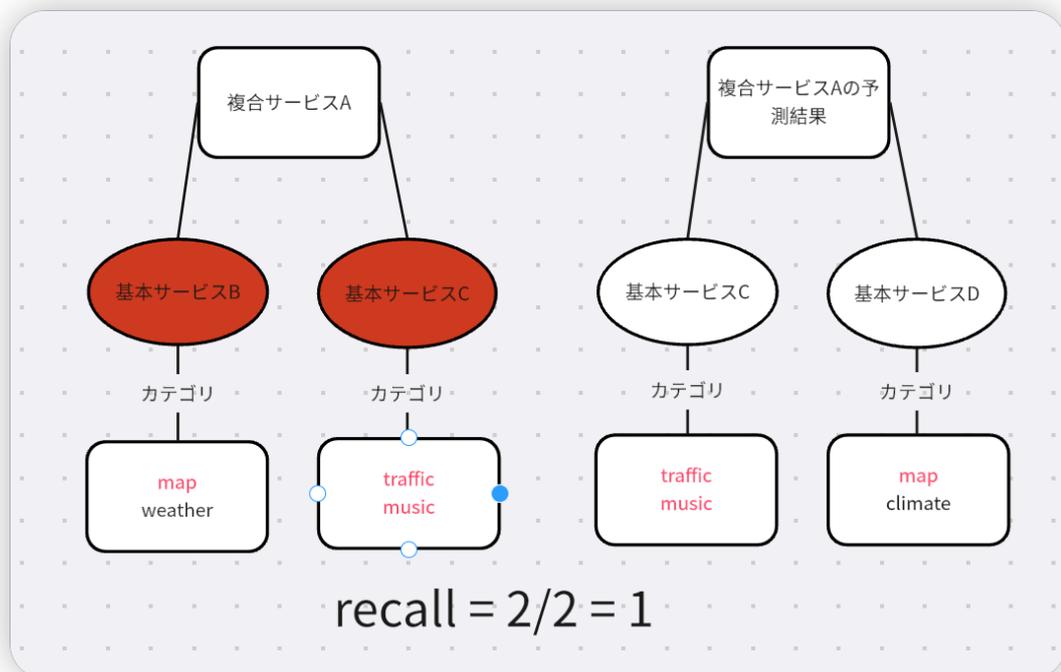


図 1 6 : 本研究の再現率

5.2 実験データ

ProgrammableWeb は, API, Mashups (複合サービス), Web サービスに関連する情報を提供する著名なオンラインサービスである. 開発者には, ソーシャルメディア, 金融, 地図サービスなど, さまざまな業界や分野をカバーする豊富な API ディレクトリが提供されている. 本実験の全データは ProgrammableWeb から取得され, 最終的に説明文を持つ基本 web サービスは 21526 件, 複合ネットワークサービスは 5604 件であった.

複合サービスは, 原子サービスで構成されているため, この 5, 604 件の複合サービスには, 1 つの基本サービスで構成されているものや, 複数の原子サービスで構成されているものもある. 組み合わされた原子サービス数ごとの複合サービスの分布を図 1 7 に示す.

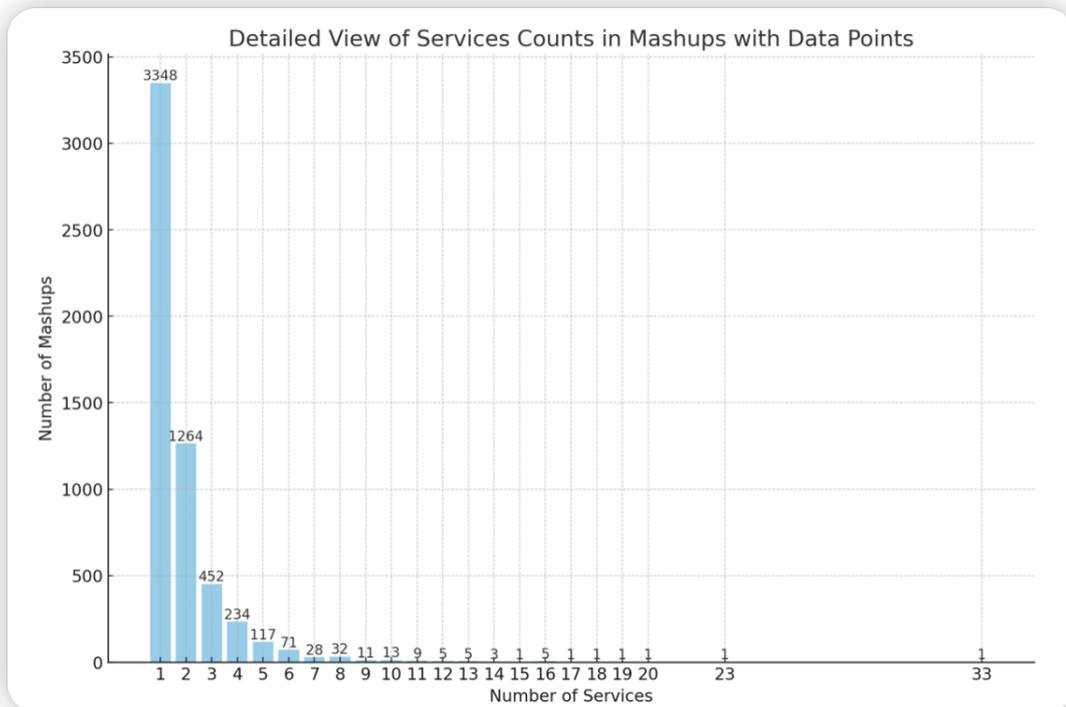


図 1 7 : 複合サービスで組み合わせられるサービス数

5.2.1 データ拡充の結果

一回目のデータ拡充でデータを拡張した後のデータ総量は **27130** 件に達し、そのうち **21526** 件は基本 web サービスを用いて拡充されたデータで、残りは元の複合サービスのデータである。この方法でデータ拡充したの結果は図 1 8 に示された：

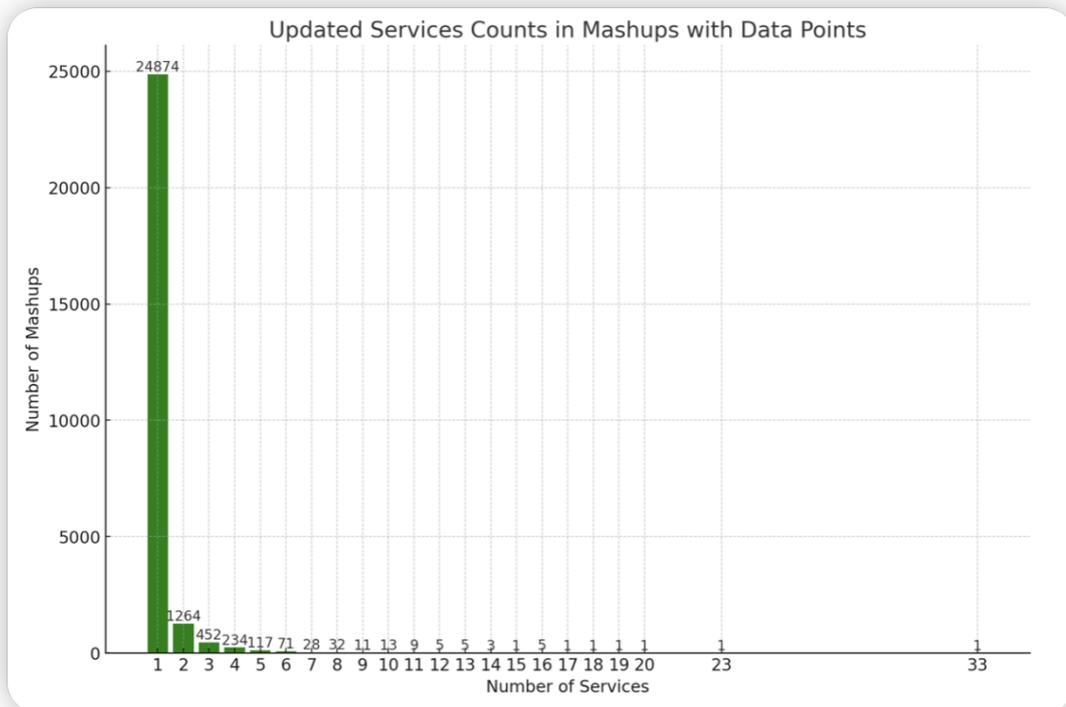


図 1 8 : 一回目のデータ拡充後の分布

2 回目のデータ拡充で訓練データを拡充する, テストデータは拡充しない, この 2 番目の方法でデータは 30489 件に拡充された. 最後拡充したの結果は図 1 9 に示された.

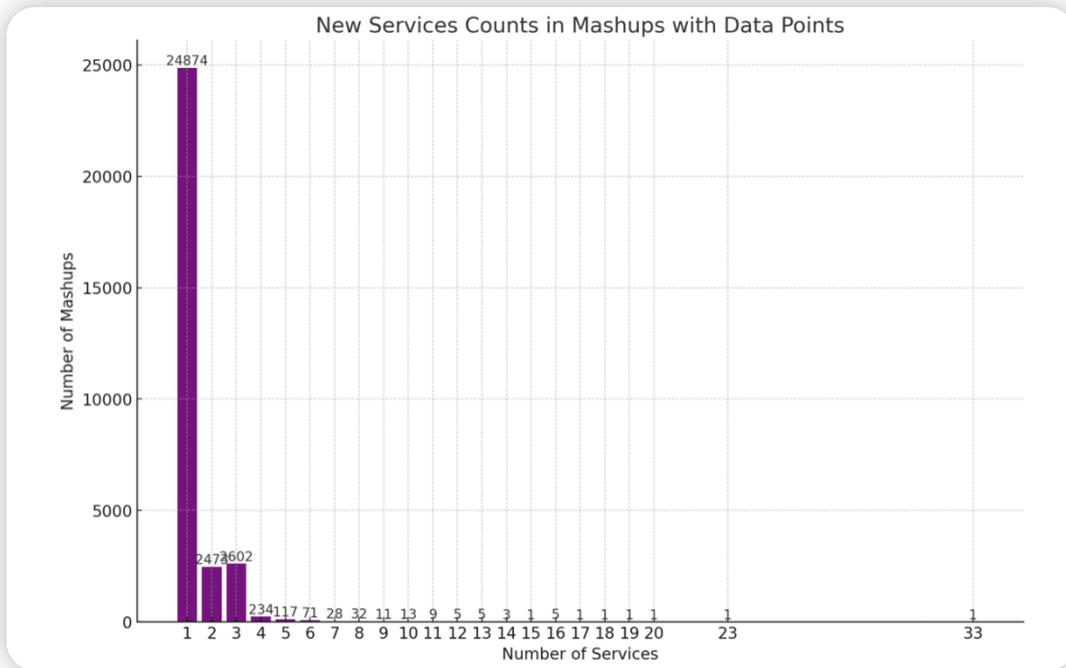


図 19 : 2 回目のデータ拡充した後の分布

5.2.2 データセット分割

本実験では合計 30489 件のデータがあり，そのうち 29989 件が訓練データ，500 件がテストデータである。

訓練データの構成は図 20 に示された，テストデータは図 21 に示された。

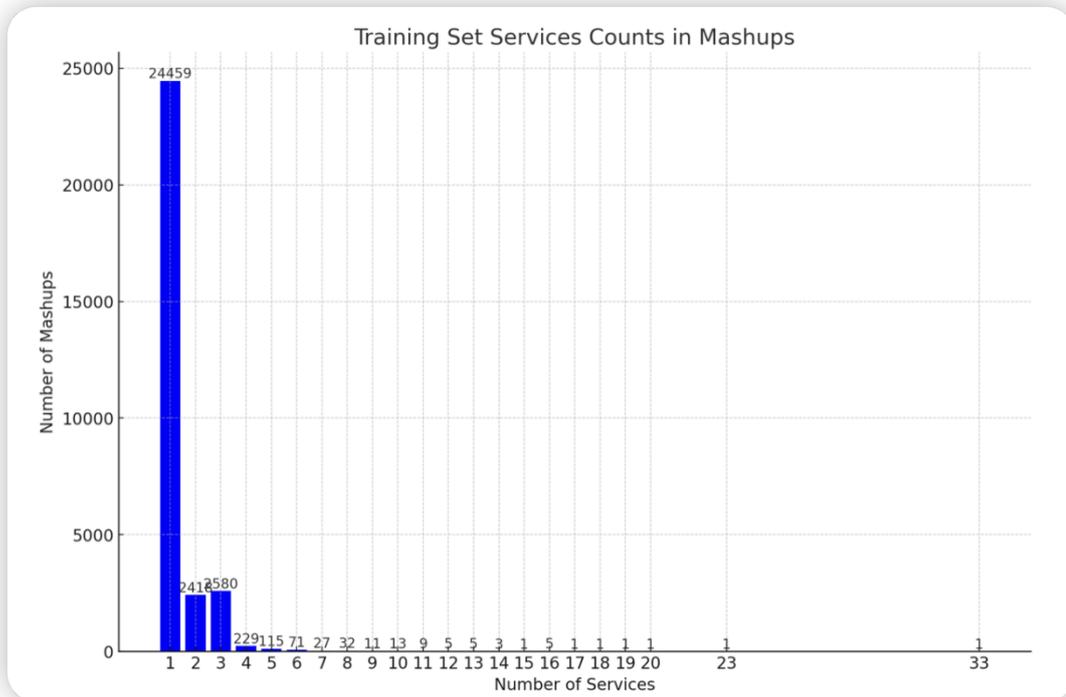


図 2 0 : 訓練データの構成

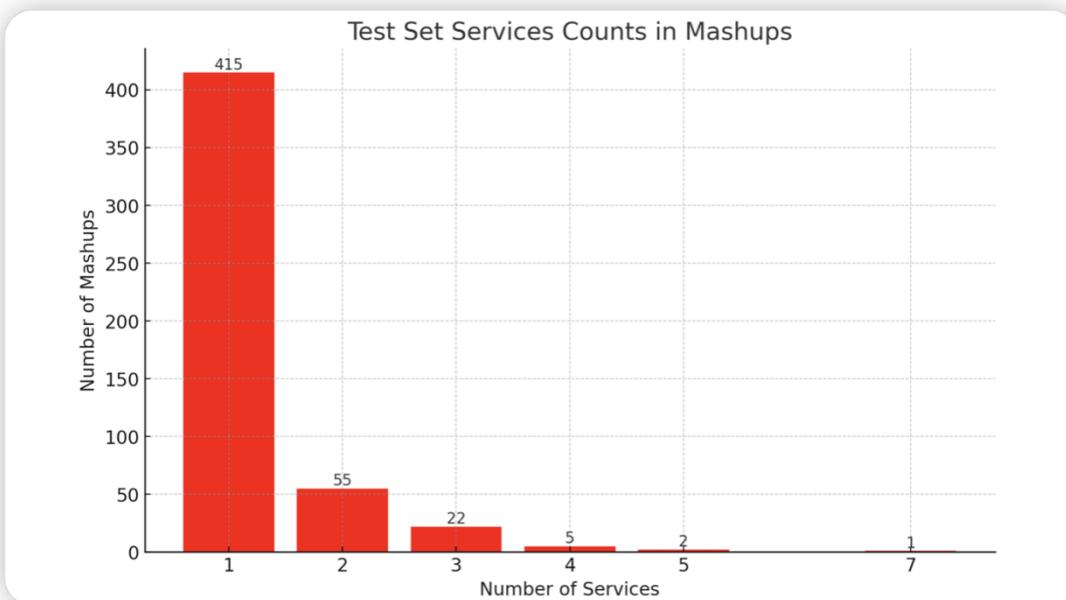


図 2 1 : テストデータの構成

5.3 結果

5.1 節で説明された評価指標を使用し、Seq2Seq に基づく複合サービス推薦の結果を以下に示す。また、ビームサーチとグリーディサーチを使用した結果は別々に表示される。データ拡充を行わない原始データで訓練されたモデルの結果も、比較のために表示される。

1. データ拡充しないの原始複合サービスのデータの結果

Programmableweb サイトから取得した原始の 5604 件の複合サービスデータを使用し、その中の 5104 件を訓練データとして、残りの 500 件をテストデータとして使用した。訓練されたモデルを評価した結果は、下表 3 に示されている。

表 3：データ拡充しないの原子複合サービスのデータの結果

| Token search | 適合率 | 再現率 | 平均長 |
|---------------|---------|---------|--------|
| Beam-search | 0.06400 | 0.14325 | 1.9291 |
| Greedy-search | 0.15960 | 0.40857 | 2.3860 |

2. 訓練データに原子サービスを追加する初回のデータ拡充の結果

21526 件の基本サービスデータを追加し、1 つの基本サービスから構成される複合サービスのデータセットとして機能させることで、拡充後のデータセットの総データ数は 27130 件となる。その中で 26630 件が訓練データとして使用され、残りの 500 件の原始複合サービスデータはテストデータとして使用される。テストの結果は下表 4 に示されている。

表 4：訓練データに原子サービスを追加する初回のデータ拡充の結果

| Token search | 適合率 | 再現率 | 平均長 |
|---------------|---------|---------|--------|
| Beam-search | 0.06080 | 0.52790 | 8.5183 |
| Greedy-search | 0.18651 | 0.75414 | 17.794 |

3. 第二回データ拡充したの結果

21526 件の基本サービスデータを追加し、原始の 5605 件の複合サービスデータから 500 件をテストセットとして選出し、残りの 5106 件のデータに対してデータ拡充を行う。2 つまたは 3 つの基本サービスから構成されるデータセットは、順序を変更することにより新たなデータを生成し、最終的に訓練セットに加

えて訓練を行う。総訓練データ数は **30489** 件で、その結果は下表 5 に示されている。

表 5 : 第二回データ拡充した結果

| Token search | 適合率 | 再現率 | 平均長 |
|---------------|----------------|----------------|---------------|
| Beam-search | 0.10792 | 0.86652 | 15.886 |
| Greedy-search | 0.20607 | 0.96461 | 38.73 |

表 3～表 5 の結果から見ると、最良の結果はデータを **2** 回拡充した後に訓練されたモデルを使用し、グリーディサーチによる予測生成を行った場合である。適合率は **0.20607**、再現率は **0.96461** である。表 3～表 5 の適合率の結果は図 2 2 に示された。再現率の結果は図 2 3 に示された。

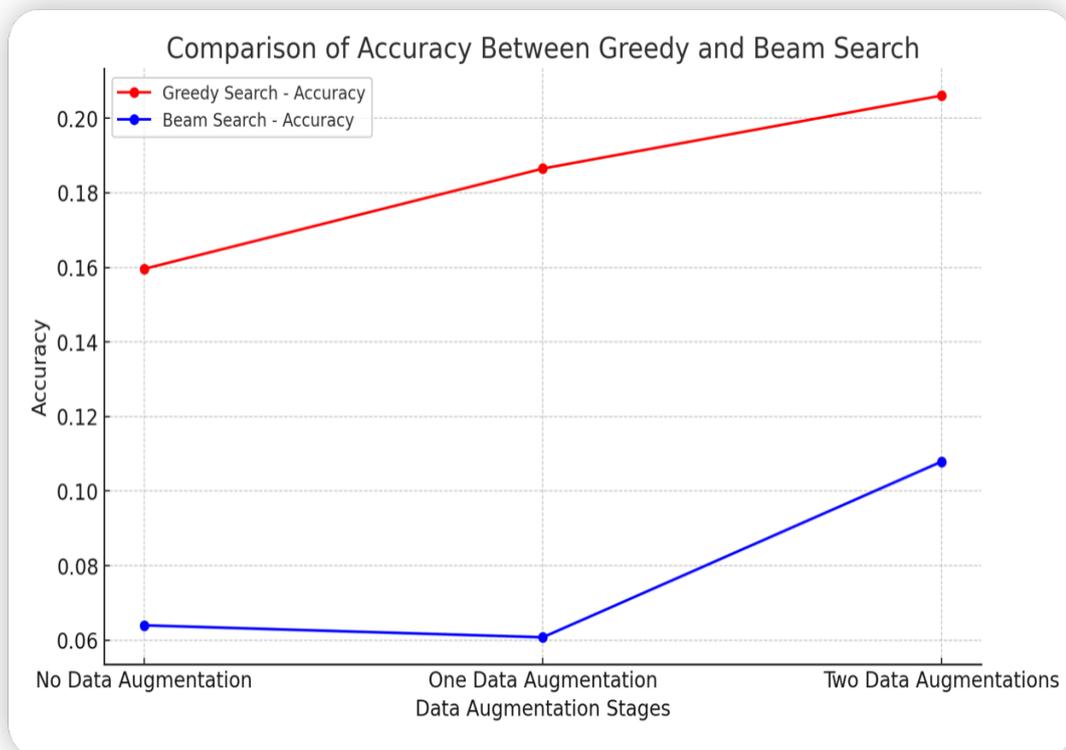


図 2 2 : 表 3～表 5 の適合率結果

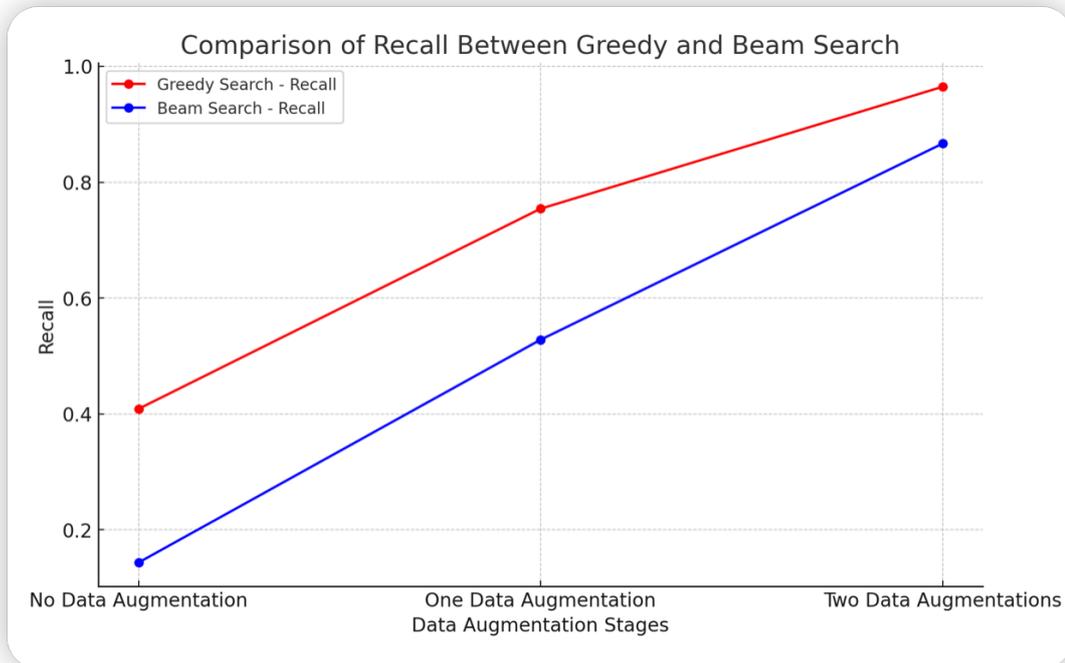


図 2.3 : 表 3 ~ 表 5 の再現率結果

図 2.3 から見える, 各データ拡充操作の後, グリーディサーチであってもビームサーチであっても, 適合率と再現率が共に向上していることが分かる. この結果から, 本研究の実験において, グリーディサーチを使用する方がビームサーチを使用するよりも効果が大きいことが示されている.

上記の表 3 ~ 表 5 のデータは, 第 5 章で紹介する適合率と再現率を用いて計算される. この方法を使用するのは, 本研究の目的により適合するためである. 通常の適合率と再現率を用いて予測結果の適合率と再現率を計算した結果は, 表 6 ~ 表 8 に示される.

4. データ拡充しないの原始複合サービスのデータの結果(通常の場合)

表 6 : データ拡充しないの原始複合サービスのデータの結果(通常の場合)

| Token search | 適合率 | 再現率 |
|---------------|---------|--------|
| Beam-search | 0.0 | 0.0 |
| Greedy-search | 0.09126 | 0.3077 |

5. 訓練データに原子サービスを追加する初回のデータ拡充の結果(通常
の計算方)

表 7 : 訓練データに原子サービスを追加する初回のデータ拡充の結果(通常
の計算方)

| Token search | 適合率 | 再現率 |
|---------------|---------|---------|
| Beam-search | 0.00932 | 0.21642 |
| Greedy-search | 0.02838 | 0.48570 |

6. 第二回データ拡充したの結果(通常
の計算方)

表 8 : 第二回データ拡充したの結果(通常
の計算方)

| Token search | 適合率 | 再現率 |
|---------------|---------|---------|
| Beam-search | 0.00646 | 0.45790 |
| Greedy-search | 0.02788 | 0.77837 |

通常的な計算方法を使用すると、結果が悪化することが分かる。なぜなら、伝
統的な計算方法では、正確な同一カテゴリーの基本サービスを無視するため
である。

以上の結果からも、Beam Search を使用した結果は Greedy Search を使用し
た結果より劣ることが見て取れる。実際に、Greedy Search は Beam Search の特
殊な形態であり、Beam Search の beam-size を 1 に設定すると Greedy Search に
なる。では、なぜ Beam Search の結果が Greedy Search より劣るのか、以下の
推測がある。

序列長 : 長い序列を処理する際、Beam Search はより多くの候補解を考慮す
るため、高品質の解を見つける機会が Greedy Search よりも良い。しかし、序列
が比較的短い場合、Greedy Search はすでに最適解に十分近いため、探索幅を増
やしても顕著な改善は得られないかもしれない。

過学習 : モデルが訓練データに過学習している場合、他の理由として考えら
れる。この場合、モデルは訓練データのパターンを繰り返す傾向があり、Greedy
Search はこれらのパターンを捉えるのにちょうど良いかもしれない。Beam
Search は多様性を導入することで、この過学習の効果を減少させる可能性があ

り，その結果，テストセットでのパフォーマンスが **Greedy Search** より劣ることになる。

実装の詳細：Beam Search の実装は Greedy Search より複雑であり，実装上の細かい違いが最終的なパフォーマンスに影響を与える可能性がある。

これらの問題を解決する方法には，評価関数を最適化することや，訓練過程で正則化技術を使用して過学習の問題を軽減することが含まれる。さらに，サンプリング方法 (**Sampling**) や，序列の長さと言質のバランスをとるために罰則項を追加するなど，他の高度な検索戦略を検討することもできる。

第6章 考察

本章では、第五章の結果に基づき、第五章で得られた最良の結果の出力長さについて考察する。seq2seq モデルによって生成される予測シーケンスの長さは固定ではないため、この章では予測の最大長さを 10 単位で制限し、最大長さを固定した後の各区間の適合率と再現率を計算する。正しい予測の区間範囲情報を得るため、正しい予測の集中区間情報を観察する。正しい予測の集中区間情報を考察した後、データの形式を変更し、サービス ID を使用せず、サービスの分類を使用して同じ構造の seq2seq モデルを訓練し、その結果を観察し比較する。

6.1 最大生成長

予測生成を行う際には、まず[BOS]を開始として生成を行い、トークンが[EOS]という終了信号を見つけるまで生成を続ける。しかし、予測生成を行う際には最大生成長を設定することが一般的である。本研究では、最大生成長として 128 を使用している。この数字は、本実験の目的に対して非常に大きいものであるため、最大生成数を調整することで予測結果を改善できるかどうかを観察する必要がある。ここでは、第五章で示された最良の結果を使用する。これは、二次データの拡充を行い、greedy-search の結果を使用したものである。適合率と再現率に加えて、予測平均長の統計も新たに表 9 に加えられている。

表 9：予測平均長を統計する結果

| 最大生成長 | 適合率 | 再現率 | 予測平均長 |
|-------|---------|---------|-------|
| 128 | 0.20607 | 0.96461 | 38.73 |

今回の考察では、最大生成長を変更して観察を行う。まず、最大生成長を 10 に設定して観察を開始し、その後の各観察で最大生成長を 10 ずつ増加させる。10, 20, 30...128 というように進める。その結果は、表 10 に示されている。適合率と再現率の観察結果は図 2 4 に示されている。

表 1 0 : 最大生成長の観察結果

| 最大生成長 | 適合率 | 再現率 | 予測平均長 |
|-------|---------|---------|--------|
| 10 | 0.16379 | 0.90440 | 9.99 |
| 20 | 0.20456 | 0.95680 | 19.692 |
| 30 | 0.21174 | 0.96320 | 28.16 |
| 40 | 0.20906 | 0.96420 | 33.332 |
| 50 | 0.20896 | 0.96460 | 36.602 |
| 60 | 0.20675 | 0.96460 | 38.336 |
| 70 | 0.20612 | 0.96460 | 38.704 |
| 80 | 0.20607 | 0.96460 | 38.73 |
| 90 | 0.20607 | 0.96460 | 38.73 |
| 100 | 0.20607 | 0.96460 | 38.73 |
| 110 | 0.20607 | 0.96460 | 38.73 |
| 120 | 0.20607 | 0.96460 | 38.73 |
| 128 | 0.20607 | 0.96460 | 38.73 |

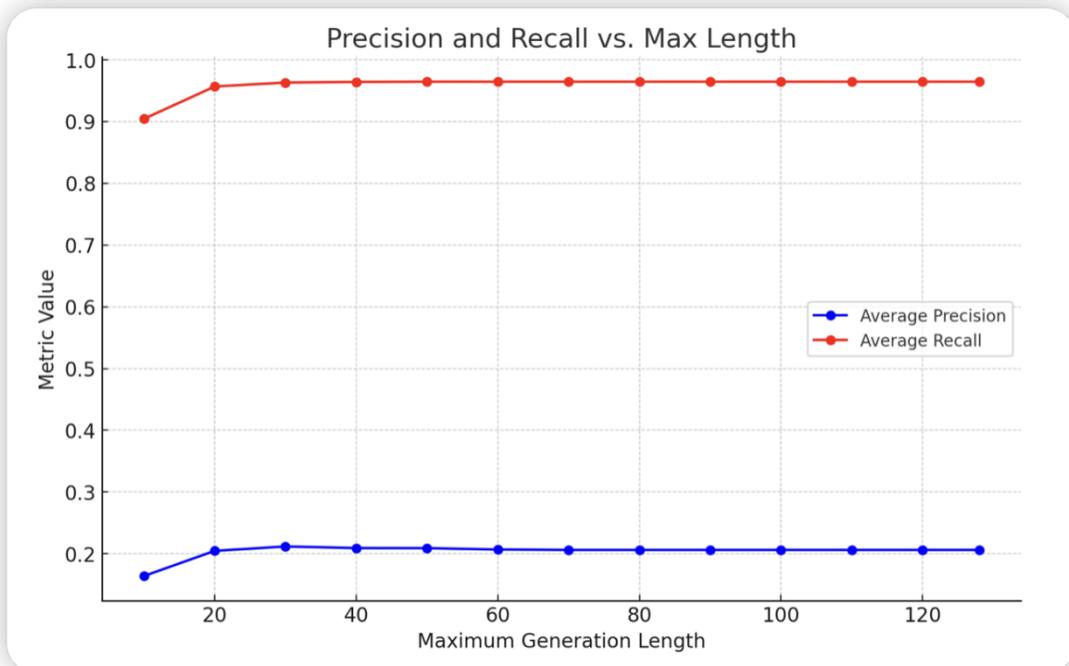


図 2 4 : 適合率と再現率の観察結果

図 2 4 から明らかに、最大長さの設定が 20 を超えると、適合率への影響は非常に小さくなる。これは、ほとんどのモデルが生成する正確な予測が 20 の長さ以内に集中していることを示している。

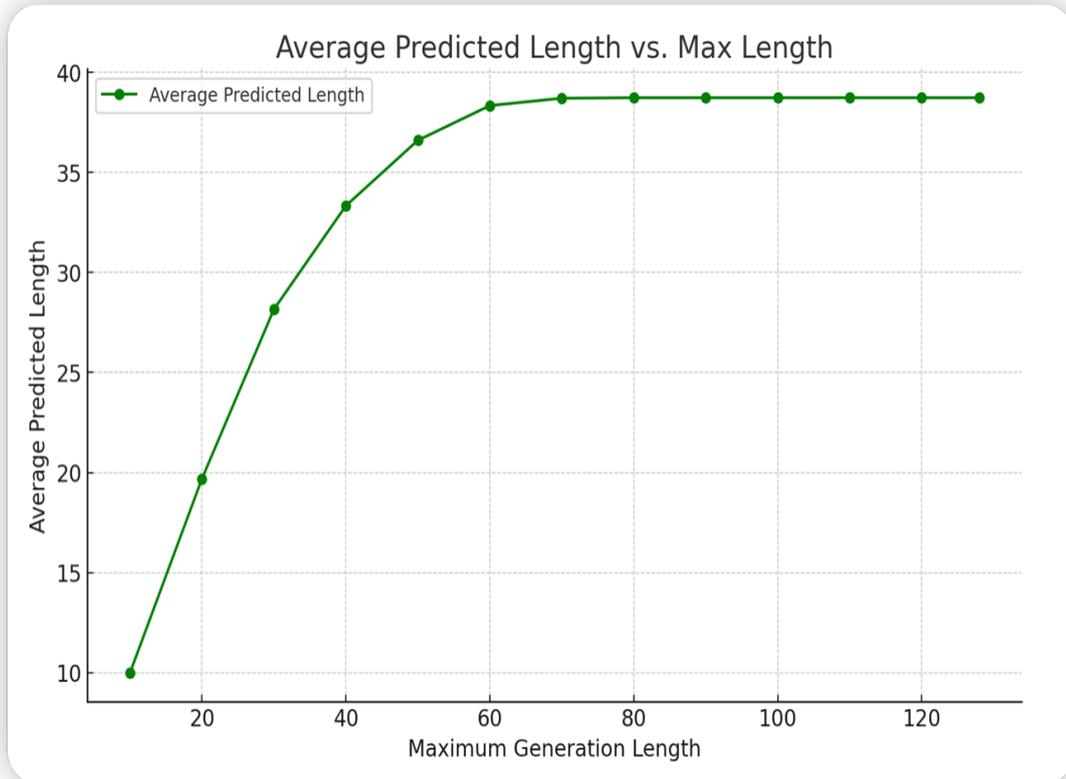


図 2 5 : 平均長の観察結果

図 2 5 によると、最大生成長の増加に伴い、予測生成の平均長さは、最大生成長が 60 に設定された時点でほとんど変化しなくなる。これは、ほとんどの予測生成の長さが 60 を超えないことを意味する。また、図 2 3 と比較すると、20 を超える長さでの予測生成のほとんどは意味がないことがわかる。なぜなら、ほとんどの説明文に適した正確な予測は、長さ 20 以内の予測に集中しているからである。

6.2 カテゴリを用いたモデルの訓練

本実験の研究目的は、複合サービスの説明文を入力し、基本サービスのサービス ID 集合を出力することである。出力が複合サービスの ID 集合ではなく、その複合サービスのカテゴリである場合、そのカテゴリを用いてマッチングサ

ービスを探す可能性もある。したがって、本節ではこのようなケースについての実験と比較を行う。カテゴリを用いたモデルの訓練データのデータ構造は図 26 に示された。

```
{
  "mashup_desc": 説明文
  "services": [
    サービスc,
    サービスb
  ],
  "category": カテゴリ
}
```

図 26 : データ構造

モデルの訓練において、ターゲットは元のサービスから図 25 に示されるカテゴリに変換する必要がある、基本の流れは図 27 に示される。カテゴリで構成される全ての単語リストを作成し、訓練を行う。このような訓練により、モデルは説明文を入力とし、その説明文のカテゴリを出力することができる。

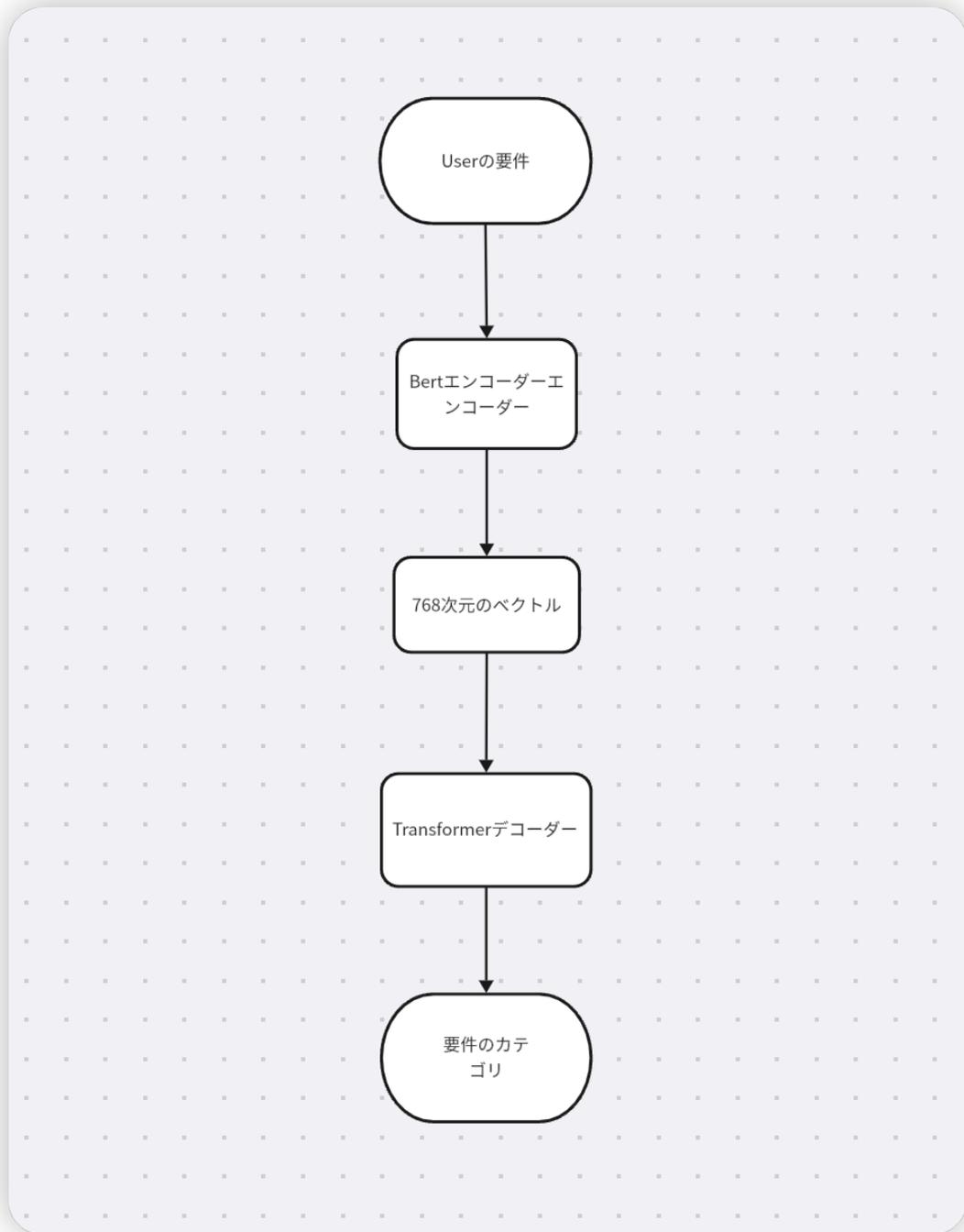


図 2 7 : カテゴリを出力の流れ

このモデルの評価には、依然として適合率と再現率を使用する。出力がカテゴリであるため、通常の適合率と再現率を直接使用して評価することができる。一度拡張されたデータを使用した結果は表 1 1 に示される。

表 1 1 : カテゴリを出力の結果

| Token search | 適合率 | 再現率 | 平均長 |
|---------------|----------|---------|---------|
| Beam-search | 0.02364 | 0.20209 | 126.739 |
| Greedy-search | 0.021414 | 0.40362 | 41.25 |

表 1 1 から、このモデルをカテゴリで訓練する場合の適合率と再現率は非常に低いことが分かる。したがって、結果の生成にはサービス ID を使用する方が結果が良い。

6.3 Bert multi-label classifier を用いた複合サービス推薦

この研究では、seq2seq モデルを使用してモデルを訓練し、訓練されたモデルを使用して具体的なサービスセットを推薦する。seq2seq で生成されたシーケンスは順序がありますが、この研究の目的はサービスセットを推薦することであり、順序は結果に影響を与えないため、BERT の multi-label classifier タスクを使用してサービスセットを推薦するののも一つの方法である。

多ラベル分類は、各インスタンス（例えば Web サービスの説明文）が複数のカテゴリやラベルに同時に割り当てられる機械学習のタスクである。これは、各インスタンスが一つのカテゴリにのみ割り当てられる従来の単一ラベル分類タスクとは異なる。多ラベル分類は多くの実用的な応用で非常に有用である。この研究の目的に従い、複合サービスを構成するサービスセットを予測するためには、実際のアトミックサービスをラベルとして扱う必要がある。例えば、複合サービスがアトミックサービス A とアトミックサービス B から構成されている場合、その複合サービスの説明文に対応するラベルは A と B である。この複合サービスの説明文と対応するアトミックサービスのラベルを使用して、BERT の multi-label classifier タスクで効果的なモデルを訓練することも可能である。

初歩実験の後、サービス ID をラベルとして使用する場合でも、カテゴリをラベルとして使用する場合でも、すべてのテストセットで予測された結果が完全に同じであることがわかった。これは、モデルがラベルを正確に学習して区別していないことを示しており、訓練データ内での出現頻度が高いラベルの確率を高め、出現頻度が低いラベルの確率を低くしている。この状況が発生する原因として以下の 2 つの推測がある：

過学習：モデルが訓練データに過学習しており、未知のデータに対して一般化できない。これは、モデルが複雑すぎるか、訓練データが少なすぎるためかもしれない。ラベルの数が増加するにつれて、可能なラベルの組み合わせの数が指数関数的に増加し、すべての可能な組み合わせを直接モデリングすることが非現実的になる。

ラベルの不均衡：あるラベルが他のラベルよりもはるかに頻繁に出現し、モデルがより一般的なラベルを予測する傾向にある。

したがって、サービスセットの推薦に多ラベル分類の方法を使用する場合、まずデータの問題を解決する必要がある。多数のラベルが存在するため、必要なデータ量は非常に大きくなるが、実際のインターネット上にはそんなに多くのデータが存在しないため、これは非常に解決が難しい問題である。したがって、この研究では最終的に **seq2seq** モデルを使用してサービスセットの推薦を行う。しかし、将来的にラベルが多すぎる問題やデータ量が不足している問題を解決できれば、多テキスト分類を使用してサービスセットの推薦を行うことは可能である。

第7章 まとめ

本研究では seq2seq モデルを使用して、自然言語によるユーザのゴール記述から、そのゴールを満たす複合サービスに必要なサービス集合を出力する手法を提案された。seq2seq モデルを基に、本実験では二つのデータ拡張方法と予測ロジックの設計を提案した。実験結果から、提案された二つのデータ最適化方法が予測結果の精度を大幅に向上させたことが分かる。提案されたデータ拡張手法を使用する前の元のデータでの適合率と再現率はそれぞれ 0.16 と 0.41 であったが、提案手法を用いた後の適合率と再現率は 0.21 と 0.96 である。本研究の貢献は以下の通りである：

訓練データの拡充

複合サービスだけでなく、原子サービスも単一サービスから構成された複合サービスと見なして訓練データを拡充する。さらに、ProgrammableWeb のデータは、複合サービスで組み合わせられる原子サービスの実行順序が明記されていないため、同じ複合サービスの説明文であっても、異なる順序のサービスシーケンスを出力にとることができる。このような方法により、データ量を大幅に拡大し、モデルは単一のサービスの機能だけでなく、異なるサービスの組み合わせから生じる潜在的なパターンを識別する能力も得ることができる。そして最後の実験がこの二つの手法は有効であると示した。

推薦ロジックの構築

本研究は最も一般的な二つの出力選択方法である Greedy-Search と Beam-Search を使用して実験比較を行う。出力選択方法に応じて、推薦ロジックを構築する必要がある。具体的には、ある出力を選択した後、選択された出力を記録し、後続の予測時に以前に選択された出力が予測された場合、その確率を修正して他の出力を選択する。ProgrammableWeb のデータを用いた評価結果から、Greedy-search の方が Beam-search よりも推薦精度が高いことが示された。

本研究では、推薦ロジック内で、Greedy-search と Beam-search の二種類の探索手法を用いたが、今後は他のアルゴリズムや手法を探求し、例えば Random Sampling とか Top-k Sampling などの探索方法を試し、さらに精度の高い推薦手法を実現するの可能性がある。今回の研究では ProgrammableWeb のデータを使用した。将来的にはより多様なデータソースを統合し、モデルの汎用性と

ロバスト性を高めることが重要である.

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に深謝申し上げます。また普段からお世話になっている社会知能研究室の皆さまにも感謝の意を表します。

参考文献

- [1] Chen L, Wu J, Jian H, et al. Instant recommendation for web services composition[J]. *IEEE Transactions on Services Computing*, 2013, 7(4): 586-598.
- Beizer, B.: *Towards a New Theory of Sequential Switching Networks*, *IEEE Trans. Computers*, Vol. C-19, pp. 936-956 (1970).
- [2] Wang H, Gu M, Yu Q, et al. Adaptive and large-scale service composition based on deep reinforcement learning[J]. *Knowledge-Based Systems*, 2019, 180: 75-90.
- [3] Fan, Guodong, et al. "What Is Next? A Generative Approach for Service Composition Recommendations." 2022 IEEE Smartworld, Ubiquitous Intelligence & Computing, Scalable Computing & Communications, Digital Twin, Privacy Computing, Metaverse, Autonomous & Trusted Vehicles (SmartWorld/UIC/ScalCom/DigitalTwin/PriComp/Meta). IEEE, 2022.
- [4] Zeng, Kungan, and Incheon Paik. "Web Service Composition Sequence Learning Based on Neural Language Networks." 2020 IEEE 14th International Conference on Big Data Science and Engineering (BigDataSE). IEEE, 2020.
- [5] Ter Beek M, Bucchiarone A, Gnesi S. Web service composition approaches: From industrial standards to formal methods[C]//Second International Conference on Internet and Web Applications and Services (ICIW'07). IEEE, 2007: 15-15.
- [6] Fan G, Chen S, He Q, et al. Service Recommendations for Mashup Based on Generation Model[J]. *IEEE Transactions on Services Computing*, 2023.
- [7] Meng, Shunmei, et al. "KASR: a keyword-aware service recommendation method on mapreduce for big data applications." *IEEE Transactions on Parallel and Distributed Systems* 25.12 (2014): 3221-3231.
- [8] Liu, Liwei, et al. "Using context similarity for service recommendation." 2010 IEEE Fourth International Conference on Semantic Computing. IEEE, 2010.
- [9] Yang Y, Qamar N, Liu P, et al. Servenet: A deep neural network for web services classification[C]//2020 IEEE international conference on web services (ICWS). IEEE, 2020: 168-175.

[10]Li B, Nong X, Hou Y, et al. Multi-Label Web Service Classification Using Neural Networks[C]//2023 IEEE International Conference on Control, Electronics and Computer Technology (ICCECT). IEEE, 2023: 540-544.