

卒業論文

マルチエージェントシステムによる
Web サービス合成

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

飛澤 佑季

2025年1月9日

マルチエージェントシステムによる Web サービス合成

飛澤 佑季

内容梗概

近年、Web サービスは生活や産業のあらゆる場面で不可欠な存在となり、複数の Web サービスを組み合わせて利用するサービス合成の重要性が高まっている。しかし従来手法では、ユーザーの目的や要求を時相論理・述語論理といった形式的表現で記述する必要があり、専門知識を持たない利用者にとって大きな障壁となっていた。

本研究では、大規模言語モデル（LLM）が自然言語で要求を記述できる点に着目し、柔軟かつ直感的に利用可能なサービス合成の実現を目指す。一方で単一エージェント方式では、多数の API 仕様や専門知識を 1 つのプロンプトに集約する必要があり、プロンプト肥大化や入力制限により推論精度が低下しやすい。

そこで、複数の LLM エージェントを協調させるマルチエージェントシステムを導入し、エージェントに API 仕様を分散保持させることでプロンプト制約を緩和する。具体的には、契約ネットプロトコル（Contract Net Protocol: CNP）に基づき、マネージャが要求を公告し、入札エージェントが担当可否と API 案を入札、マネージャが統合して合成計画を確定する。

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。

タスク分解 自然言語で記述されたユーザ要求に対し、CNP に基づいてタスク分解を行う際に、要求の粒度や曖昧さに応じて分解方式をどのように選択すべきか、またその分解をマネージャ等のどの役割のエージェントが担うことが適切かを検証する必要がある。

入札エージェントの組織化 CNP では、入札エージェント数の増加に伴い公告・入札の通信量が増大し、スケーラビリティが低下しやすいという課題がある。このデメリットを緩和するため、入札エージェントに対してどの API を割り当てるべきか、また入札エージェント群のカテゴリ別階層化・集約を検討し、組織構造の違いが入札の精度や合成結果、ならびに通信負荷に与える影響を検証する必要がある。

1 つ目の課題に対しては、マネージャ主導・コントラクタ主導・協調型の 3 つのタスク分解プロトコルを比較する。マネージャ主導では、マネージャがタスク分解とカテゴリ分解を行ってタスク仕様を公告し、入札エージェントは担当可

否と API 案を入札する。コントラクタ主導では、入札エージェントが自身の担当 API 仕様に基づいてタスク分解とカテゴリマッピングを行い入札する。協調型では、マネージャがタスク分解を提示したうえで、入札エージェントが分解されたタスクに対してカテゴリマッピングを行って入札する。本研究では、要求の粒度や曖昧さといった要求タイプに応じて、これら 3 つのプロトコルのうちどれが適するかを明らかにし、分担の違いが合成計画の整合性および計画生成精度に与える影響を分析する。

2 つ目の課題に対しては、入札エージェントの組織化により解決を図る。具体的には、入札エージェントに割り当てる担当範囲として、単一 API を担当させる設計に加え、API カテゴリ単位で複数 API をまとめて担当させる設計も比較対象とする。各エージェントは、割り当てられた API（またはカテゴリに属する複数 API）の仕様のみを参照して推論・入札を行うようにし、推薦精度および通信負荷に与える影響を検証する。

本研究では、提案手法の有効性を検証するため、ProgrammableWeb の複合サービスデータを用いた評価を行った。具体的には、複合サービスデータ 400 件をテストデータとし、対応する 909 件の Web API を対象として、マネージャ主導・コントラクタ主導・協調型という三種類のタスク分解方式を適用してサービス合成を実行し、各方式の計画生成精度を比較した。生成されたサービス計画と正解計画を照合し、適合率、再現率、F1 スコアを算出することでタスク分解方式の性能を評価した。また、入札エージェントの組織化（単一 API 単位の割当と、カテゴリ単位で複数 API を割り当てる設計）についても、各エージェントの入札行動と推薦精度を比較し、役割分担の違いがサービス合成結果に与える影響を分析した。また本研究の貢献は以下の通りである。

タスク分解 マネージャ主導・コントラクタ主導・協調型の三つのタスク分解プロトコルを比較し、ユーザ要求の粒度に応じた適切な方式を分析した。適合率・再現率・F1 値で評価した結果、本手法は単一方式を上回る性能を示した。

入札エージェントの組織化 エージェント数と担当 API 数を変化させた実験を実施し、専門性の深さと分担効率の関係を評価した。その結果、API カテゴリを限定した専門エージェントを用いることで、単一エージェント方式と比較してサービス推論の正確性が向上し、大規模サービス集合に対してもスケーラブルに動作することが明らかとなった

Web Service Composition via a Multi-Agent System

Yuki Tobisawa

Abstract

Web services have become indispensable in daily life and industry, and the importance of composing multiple Web services to realize value-added applications is increasing. However, conventional approaches often require users to specify goals and constraints in formal languages such as temporal or predicate logic, which poses a barrier to non-expert users.

This study aims to enable flexible and intuitive service composition by leveraging large language models (LLMs), which allow users to describe requirements in natural language. A single-agent approach tends to aggregate many API specifications and domain knowledge into one prompt, which can lead to prompt bloating and degraded inference quality due to context-length limitations.

To mitigate this issue, we introduce a multi-agent LLM system in which API specifications are distributed across agents. Based on the Contract Net Protocol (CNP), a manager agent announces a user request, contractor agents bid with feasibility and candidate APIs, and the manager integrates bids to finalize a composition plan (Plan).

We address two challenges: (1) task decomposition from natural-language requirements under CNP, including how to select a decomposition protocol and which agent role should perform it; and (2) organizing contractor agents to reduce communication overhead while maintaining recommendation accuracy. We compare three task decomposition protocols—manager-led, contractor-led, and cooperative—and also compare assigning agents per API versus per API category.

For evaluation, we use 400 mashup services from ProgrammableWeb as test data and 909 corresponding Web APIs. We execute service composition under each protocol and measure plan-generation quality by comparing generated plans to ground-truth plans using precision, recall, and F1 score. We further analyze how different agent organizations affect bidding behavior, recommendation quality, and communication cost.

マルチエージェントシステムによる Web サービス合成

目次

第1章	はじめに	1
第2章	サービス合成	3
2.1	サービス合成手法	3
2.1.1	水平型サービス合成	3
2.1.2	垂直型サービス合成	4
2.2	LLM を用いたサービス合成	4
第3章	契約ネットプロトコルに基づくサービス合成	5
3.1	契約ネットプロトコル	5
3.2	システム概要	7
3.3	LLM に基づくエージェント	9
3.3.1	コントラクタエージェント	9
3.3.2	マネージャエージェント	9
3.4	プロトコル	10
3.4.1	マネージャ主導型タスク分解	10
3.4.2	コントラクタ主導型タスク分解	10
3.4.3	協調型タスク分解	12
第4章	評価	13
4.1	実験設定	13
4.1.1	実験データ	13
4.1.2	評価指標	15
4.2	実験結果	15
4.2.1	マネージャ主導タスク分解	16
4.2.2	コントラクタ主導タスク分解	16
4.2.3	協調型タスク分解	16
第5章	考察	17
5.1	単一エージェントとの比較	17
5.2	コントラクタエージェント数	17

5.3	今後の課題	17
第 6 章	おわりに	18
	参考文献	19

第1章 はじめに

近年、サービスコンピューティングの発展により、多種多様な Web サービスを組み合わせて新たな付加価値を提供する複合サービスが盛んに構築されている。翻訳サービスと音声認識サービスを連携させたリアルタイム音声翻訳や、地理情報サービスと天気情報サービスを統合した地図表示サービスなど、個々の Web サービスでは実現できない機能を柔軟に実装できる。しかし、公開されている Web サービスのうち、複合サービスに実際に利用されているものは一部にとどまり、ユーザが膨大な候補の中から目的に適したサービスを選択することは依然として大きな負担となっている。

従来の複合サービス合成は、水平型サービス合成と垂直型サービス合成の二つのアプローチに大別される。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 サービスの組み合わせを選択する枠組みであり、ワークフロー設計のコストが高いという課題を持つ。後者は、人工知能のプランニング技術を用いて、論理式で与えられたゴール状態に到達するサービス実行系列を生成する手法であるが、ユーザが時相論理や述語論理といった形式的表現で要求を記述しなければならず、エンドユーザにとって扱いやすいとは言い難い。この問題を軽減するため、自然言語で記述された要件定義からサービスの組み合わせを推定する手法や、WSDL やサービス説明文、サービスネットワークをテキストマイニングやグラフ埋め込みによって特徴ベクトル化し、機能ごとにクラスタリングする手法が提案されてきた。

一方、近年急速に発展している大規模言語モデル (Large Language Model; LLM) は、自然言語で記述された要求から計画を生成し、外部ツールや API を呼び出す能力を備えており、ユーザに形式的な記述を要求することなくサービス合成を行う枠組みとして期待されている。しかし、単一の汎用 LLM エージェントに広範なサービス領域を一括して扱わせる場合、個々のサービス仕様や実行環境に関する専門的な知識の理解が浅くなりやすい。また、多数の API 仕様や利用例を一つのプロンプトに詰め込む必要があるため、プロンプトの膨張とコンテキスト長制限に起因する性能劣化が生じるという問題がある。

本研究では、これらの課題を解決するため、複数の LLM エージェントからなるマルチエージェントシステムに基づくサービス合成手法を提案する。各エージェントを特定の API カテゴリや機能領域に特化させることで、プロンプト内

で扱う知識を局所化しつつ、専門性の高いサービス推薦を実現することを目指す。さらに、タスク割当てメカニズムとして契約ネットプロトコルを導入し、マネージャエージェントとコントラクタエージェントがタスク分解と入札・応札を通じて協調的にサービス合成を行う枠組みを設計する。本論文では、マネージャ主導・コントラクタ主導・協調型という三種類のタスク分解プロトコルを比較し、サービス合成精度や単一エージェント方式との性能差、エージェント数や担当 API 数が与える影響を評価することで、LLM ベース・マルチエージェントによるサービス合成の有効性と課題を明らかにする。

そこで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。

タスク分解 自然言語で記述されたユーザ要求から、契約ネットプロトコルに基づいて適切なタスク分解を行う。要求の粒度や曖昧さに応じて、マネージャ主導・コントラクタ主導・協調型といった異なるタスク分解戦略を切り替えつつも、一貫したサービス合成結果を得られる仕組みが必要となる。

LLM エージェントの役割分担 エージェントを特定の API カテゴリや機能領域に特化させることでプロンプトの膨張を抑えつつ、サービス仕様や実行環境に関する専門的な知識を十分に反映させる必要がある。また、タスク分解プロトコルの違いがエージェント間の情報共有や推論過程に与える影響を明らかにすることも求められる。

第2章 サービス合成

本章では、複合サービスの実現に向けたサービス合成手法について説明する。複合サービスとは、インターネット上に存在する複数の Web サービスを組み合わせて、新たなサービスのことである。近年では、さまざまな Web サービスが API を通じて連携可能となっており、ユーザ要求に応じた柔軟なシステム開発が可能となっている。

複合サービスの代表例として Expedia を挙げる。Expedia は航空券、宿泊、移動手段など複数の API を統合することで、旅行計画・予約を一括提供している。航空券の検索では航空会社情報を提供する API、予約確定時には決済 API、旅の計画閲覧には地図 API を利用し、単独では実現不可能な利便性を提供している。このような複合サービスを開発するためには、ユーザ要求に応じて必要な機能を提供できる Web サービスを適切に発見する必要がある。しかし膨大な数の Web サービスの中から目的に合致したものを探索することは容易ではない。本章では、サービス合成手法について説明する。

2.1 サービス合成手法

従来のサービス合成手法では大きく、水平サービス型と垂直サービス型の 2 つに分類できる。以下でそれぞれについて説明する。

2.1.1 水平型サービス合成

水平型サービス合成とは、あらかじめ与えられたワークフロー（タスク列）に対し、各タスクを実行可能な Web サービス候補の中から、合成全体として望ましい組合せを選択する枠組みである。Zeng らは、同一の機能を提供するサービスが多数存在する状況を想定し、機能差ではなく QoS（非機能的特性）に基づいてサービスを選別する合成方式を提案し [1]。具体的には、価格・応答時間・可用性などの非機能情報から各サービスの QoS ベクトルを定義し、ユーザが与える制約や嗜好（重み付け）を用いて候補を評価する。その上でサービス選択をローカル最適化とグローバル最適化の二通りで扱う。ローカル最適化では、各タスクごとに他タスクとの関係を考慮せず、制約を満たす範囲でスコアが最大となるサービスを選ぶため計算が軽い一方、合成全体の QoS が最適になる保証は弱い。これに対しグローバル最適化では、合成全体の QoS を目的としてタスク間の影響も含めて最適な割当を求めることで、合成結果の品質向上

を狙う。

2.1.2 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て、ユーザが指定したゴール状態に到達するための Web サービスの実行系列（合成手順）そのもの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して捉え、オープンで不確実性を含む Web 環境に適した合成アルゴリズムを提案した [2]。提案手法の要点は二つに整理できる。第一に、セマンティック型マッチングである。Web サービス間では、出力と入力データ型が厳密に一致しないことが多く、単純なラベル一致に基づく接続判定では合成可能性を過小評価する。そこで、出力型が他サービスの入力型や最終的に達成すべき目標型と意味的に関連するかを、ラベルが指す概念の一般化関係や型構造も含めて判断し、異種スキーマ間の非互換性を緩和する。第二に、探索と実行を統合した逐次的な合成である。サービスの実行結果や入出力値が事前に確定しない状況では、合成計画を最初から完全に構築することが難しいため、候補サービスを選択して実行し、得られた結果を観測して次のステップを更新する手順を繰り返す。このように探索と実行を交互に行うことで、不完全な情報の下でも計画を修正しながら目標達成へ近づく柔軟性を確保する。以上より、本研究は、型の意味的整合性に基づく接続判定と、試行錯誤を許容する逐次実行型の探索戦略を組み合わせ、垂直型サービス合成を実現する先行研究として位置付けられる。

2.2 LLM を用いたサービス合成

LLM を用いたサービス合成では、自然言語で記述されたユーザ要求から、適切な Web サービスを選択・組み合わせる手法である。従来機能要件を論理ベースで記述し、形式的なフロー設計が必要であったが、LLM の自然言語理解能力を活用することで、機能要件を推定し、それに基づいてタスク分割をすることができる。そして API 使用に関しても、LLM が API ドキュメントを解析し、適切な呼び出し方法を生成することで、ユーザは専門知識なしにサービス合成を実現できる。

第3章 契約ネットプロトコルに基づくサービス合成

本章では、LLMを用いたマルチエージェントシステムに基づくサービス合成手法について説明する。マルチエージェントを用いる中で本研究では契約ネットプロトコル (Contract Net Protocol; CNP) を採用し、エージェント間の協調メカニズムとして利用する。

3.1 契約ネットプロトコル

契約ネットプロトコル (Contract Net Protocol; CNP) は、Reid G. Smith によって提案された分散問題解決のための高水準通信プロトコルであり、分散環境におけるタスク割当てと協調計算を、交渉に基づくメッセージ交換によって実現する枠組みである。CNP では、タスクを外部に委託したい側のエージェントをマネージャ (manager)、タスクの実行を引き受ける側のエージェントをコントラクタ (contractor) と呼び、両者の間でタスクの告知、入札、契約締結、結果報告といった一連のやり取りを明示的なプロトコルとして規定している。

典型的な CNP の相互作用は、次のようなステップで構成される。

1. タスク告知 (task announcement / call for proposals)

マネージャは、自身が保持するタスクの内容 (目的, 前提条件, 必要資源, 締切, 評価基準など) を記述した告知メッセージをブロードキャストし、コントラクタ候補に対して提案を呼びかける。

2. 入札 (bidding / proposal submission)

告知を受け取った各コントラクタは、タスクを遂行可能かどうか、現在の負荷や利用可能資源を考慮して評価する。実行可能であれば、予想コスト、所要時間、達成可能な品質などを含む入札 (プロポーザル) をマネージャに送信し、実行困難と判断した場合は辞退メッセージを返す。

3. 契約締結・割当て (award / contract)

マネージャは、収集した複数の入札を比較し、評価基準に基づいて最も望ましいコントラクタを選定する。選定されたコントラクタには契約成立を示すアワードメッセージを送り、その他のコントラクタには不採択を通知する。これにより、当該タスクに関する一時的なマネージャ-コントラクタ関係が形成される。

4. 実行と結果報告 (task execution / status reporting)

コントラクタは割り当てられたタスクを実行し、中間状態や完了結果をステータス報告メッセージとしてマネージャに送信する。マネージャはこれらの報告に基づいてタスクの進捗を把握し、必要であれば再割当てなどの制御を行う。

このように、CNP は単なるタスク割当てアルゴリズムというよりも、タスク告知から結果報告に至るまでのエージェント間通信の流れとメッセージ種別を明確に定義したコミュニケーション・プロトコルとして位置づけられる。各エージェントは、あるタスクに関してはマネージャとして振る舞い、別のタスクに関してはコントラクタとして振る舞うことができるため、システム全体としては固定的な中央管理者を持たない柔軟な階層構造を動的に形成できるという特徴を持つ。また、タスク告知の範囲や入札の条件を調整することで、探索空間の絞り込みや通信量の制御が行える点も利点とされている。本研究では、このCNP をエージェント間の協調メカニズムとして採用し、サービス合成におけるタスク（機能分割や API 選択など）の割当てに応用する。

図 1: CNP のシーケンス図

3.2 システム概要

本研究で提案するサービス合成システムは、LLM を用いたマルチエージェントシステムに基づき、自然言語で与えられた複合サービスの要件から、利用すべき Web サービス API の組み合わせを推薦することを目的としている。

従来の LLM による API 推薦では、要件理解から API 選択までを単一の推論として一括に処理する設計が多い。しかしこの設計では、多数の API 説明文や仕様を単一プロンプトへ集約する必要があり入力肥大化しやすい、そして要件や API 群の変動により推論の再現性が低下しやすい。さらに推論過程がブラックボックス化し、誤りが生じた際にどの段階で破綻したかを切り分けにくいといった課題が生じる。各ステップに目的・参照情報・出力形式を与えて推論を誘導する Guided Reasoning Chains の考え方に基づき、サービス合成推論を LLM の内部処理に閉じず、外部化可能な推論タスクの系列として捉える。

本システムは、ユーザ要件を受理しタスクを編成するマネージャエージェント、各 API カテゴリや個別 API に専門化した複数のコントラクタエージェント、および各エージェントが参照する API 説明文から構成される。エージェント間の協調枠組みとして Contract Net Protocol (CNP) を採用し、タスクの公告 (Call for Proposals)、入札 (Proposals)、選択 (Award) という流れのもとで推論タスクを分担し、最終的な API 推薦結果を得る。CNP を採用する理由は以下の通りである。

1. 専門性に基づく自律的な可否判断を反映するためである。

本研究では、コントラクタエージェントを API カテゴリまたは個別 API に専門化させ、各エージェントが保持する局所的知識に基づいて「担当可能性」と「候補 API」を判断する設計をとる。CNP は、公告に対して各コントラクタが自律的に入札・辞退を返す手順を持つため、専門性に依存した可否判断を自然に取り込める。

2. 複数候補を比較可能な形で収集し、最終決定を体系化するためである。

サービス合成では、要求に適合する API 候補が複数存在し得るため、候補の列挙に留まらず、根拠や評価観点を揃えて比較し、最終的な採択を行う必要がある。CNP は複数の入札を収集し、マネージャが採択する流れを規定するため、候補の収集から統合・決定までを一貫した枠組みとして扱える。

3. 責任分担の差分をプロトコルとして明示し、方式間の比較を可能にするためである。

本研究の焦点は、推論タスクの責任分担（どの推論ステップを誰が担当するか）が、推薦精度および通信負荷に与える影響を明らかにする点にある。CNP は公告・入札・選択という役割とメッセージの流れを明確に定義しているため、責任分担の差分をプロトコルとして記述でき、方式間の比較・検証を体系的に行える。

その上で、本システムにおけるサービス合成の推論過程を、以下の 4 ステップからなる共通フローとして定義する。

1. コア機能分解：要件文からユーザが最終的に達成したい目的を推定し、必要となる機能単位を抽出する。
2. カテゴリマッピング：抽出した機能を実現し得る API カテゴリ（例：決済、地図、予約管理など）を推定する。
3. API マッチング：推定カテゴリに属する API 説明文と要件に基づいて、適合する候補 API を抽出する。
4. API 選択：抽出した候補 API を比較し、要件適合性の観点から採用 API を確定する。

ここで、(3) は候補 API 集合の抽出（候補提示）、(4) は候補集合からの最終採択として区別する。

この 4 ステップは単一のエージェントが一括で実行するのではなく、CNP による公告・入札・選択の枠組みにおいて、マネージャとコントラクタの間で分担される推論タスクとして扱われる。マネージャは要件文の受理、推論タスクの定義と公告、入札結果の集約と整合性確認、および API 群の確定を担う。コントラクタは担当の API 知識に基づき、公告されたタスクに対する判断と候補 API、および根拠を入札として提示する。

本研究では、推論タスクの責任分担（どのステップを誰が担当するか）を CNP 上の設計変数とみなし、マネージャ主導型、コントラクタ主導型、協調型の 3 方式を設計・比較する。この比較により、推論の分割粒度と専門化の度合いが、API 推薦精度および推論コストに与える影響を明らかにする。

3.3 LLMに基づくエージェント

本研究では、自然言語で記述されたユーザ要求および API 説明文を入力として、サービス合成のための API 群を生成する推論主体として、LLM に基づくエージェントを用いる。

本システムにおける LLM エージェントは、マネージャエージェントと複数のコントラクタエージェントから構成される。各エージェントは、CNP に基づく公告・入札・選択の枠組みの中で、自身に割り当てられた推論タスクを実行し、その結果を明示的なメッセージとして外部に出力する。これにより、サービス合成推論は LLM の内部推論として閉じるのではなく、外部から観測・比較可能な推論タスクの系列として実現される。

3.3.1 コントラクタエージェント

コントラクタエージェントは、個別 API に専門化された LLM エージェントであり、マネージャから公告されたタスク仕様を入力として受け取る。各コントラクタは、自身が参照可能な API 説明文のみを用いて推論を行い、当該タスクを担当可能か否かを判断する。

担当可能と判断した場合、コントラクタエージェントは、適合すると考えられる API 候補と、その判断根拠を入札としてマネージャに返す。ここで提示される API 候補は、システム概要で定義した推論ステップ (3) API マッチングに相当し、候補 API 集合の抽出結果として位置付けられる。一方、担当不可能と判断した場合には、辞退メッセージを返すことで入札を行わない。

3.3.2 マネージャエージェント

マネージャエージェントは、ユーザ要求を受理し、サービス合成全体を統括する役割を担う LLM エージェントである。マネージャは、システム概要で定義した推論過程に基づいて、推論タスクを構成し、CNP に従ってコントラクタエージェントへ公告する。

入札を受信した後、マネージャエージェントは、複数のコントラクタから提示された API 候補と根拠を比較・統合し、要求全体との整合性を確認した上で、最終的に採用する API 集合を確定させる。この処理は、推論ステップ (4) API 選択に対応し、候補集合からの最終採択を意味する。

このように、マネージャエージェントは、分散した推論結果を集約する集中意思決定の役割を担い、コントラクタエージェントは専門性に基づく局所的推

論を担うことで、分散推論と統合的意思決定を両立させている。

3.4 プロトコル

本研究では、CNP に基づくサービス合成において、推論タスクの責任分担（どの推論ステップをどのエージェントが担当するか）を設計変数として扱う。具体的には、システム概要で定義した 4 ステップの推論過程（コア機能分解、カテゴリマッピング、API マッチング、API 選択）を、マネージャエージェントとコントラクタエージェントの間でどのように分担させるかによって、異なるサービス合成プロトコルを設計する。

本節では、推論タスクの分担方針が異なるマネージャ主導型、コントラクタ主導型、協調型の 3 方式を定義し、各方式における公告内容、入札内容、および最終的な API 群の確定方法を明確にする。

3.4.1 マネージャ主導型タスク分解

マネージャ主導型では、サービス合成に必要な推論の大部分をマネージャエージェントが担当する。4 ステップの推論過程における役割分担は以下の通りである。

- マネージャの担当：(1) コア機能分解、(2) カテゴリマッピング、(4) API 選択
- コントラクタの担当：(3) API マッチング

本方式では、マネージャエージェントがユーザ要求を解析し、必要なコア機能および対応する API カテゴリを事前に確定した上で、それらをタスク仕様としてコントラクタに公告する。公告には、対象となる機能、想定カテゴリが含まれる。

各コントラクタエージェントは、公告された機能及びカテゴリに基づき、自身が参照可能な API 説明文を用いて API マッチングを行い、適合すると判断した API 候補（自身）とその根拠を入札として返す。マネージャは、収集した入札を比較・統合し、要求全体との整合性を確認した上で、最終的な API 集合を確定させる。

3.4.2 コントラクタ主導型タスク分解

コントラクタ主導型では、推論の大部分をコントラクタエージェントに委ねる。4 ステップの推論過程における役割分担は以下の通りである。

- マネージャの担当：(4) API 選択

- コントラクタの担当：(1) コア機能分解, (2) カテゴリマッピング, (3) API マッチング

本方式では, マネージャエージェントは, ユーザ要求文そのものをタスクとして公告し, 詳細な分解やカテゴリ推定は行わない. 各コントラクタエージェントは, 自身の担当 API 知識に基づいて要求文を解釈し, コア機能分解およびカテゴリマッピングを行った上で, 適合する API 候補と根拠を入札として提示する.

マネージャエージェントは, 複数のコントラクタから提出された入札を集約し, 提案内容の整合性を確認した上で, 最終的に採用する API 集合を確定させる.

図 2: マネージャ主導プロトコルのシーケンス図

図 3: コントラクタ主導プロトコルのシーケンス図

3.4.3 協調型タスク分解

協調型では、マネージャとコントラクタが推論タスクを分担し、両者の役割を組み合わせる。4 ステップの推論過程における役割分担は以下の通りである。

- マネージャの担当：(1) コア機能分解
- コントラクタの担当：(2) カテゴリマッピング, (3) API マッチング, (4) API 選択

本方式では、マネージャエージェントがユーザ要求からコア機能分解を行い、抽出した機能単位をタスクとして公告する。各コントラクタエージェントは、公告された機能に対してカテゴリ候補を推定し、担当 API 知識に基づいて API マッチングおよび選択を行い、採用すべき API 候補と根拠を入札として返す。

マネージャは、各入札を統合し、要求全体との整合性を確認した上で、最終的な API 集合を API 群として確定させる。

図 4: 協調型プロトコルのシーケンス図

第4章 評価

4.1 実験設定

本研究では、提案する CNP ベースのマルチエージェント方式について、タスク分解プロトコル（マネージャ主導・コントラクタ主導・協調型）と入札エージェントの組織化（単一 API 単位／カテゴリ単位）を比較する。いずれの設定においても、ユーザ要求（マッシュアップ要件）を入力として API 群を生成し、正解計画との照合によって計画生成精度を評価する。

4.1.1 実験データ

本研究では、サービス推薦モデルを構築するために、ProgrammableWeb に掲載されている複合サービス（Mashup）データおよび原子サービス（Web API）データを使用した。複合サービスデータは、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（使用 API の一覧）、および ID を値として保持する。複合サービスカテゴリおよび構成サービスは複数存在する場合があるため、いずれもリスト型として扱う。原子サービスデータは、原子サービス名をキーとし、サービスカテゴリ、原子サービス説明文、サービス提供者、サービス利用者、および ID を値として保持する。カテゴリ情報は複数付与されており、そのうち先頭要素をメインカテゴリ、それ以外をサブカテゴリとして扱う。これらの属性はいずれも ProgrammableWeb 上でユーザにより付与されたものである。複合サービスデータには欠損が含まれるため、前処理としてフィルタリングを行った。フィルタリング条件は、「複合サービスが参照する構成サービス（使用 API）が、原子サービスデータに存在しない」レコードを除外することである（データ構造は図2・図3に示す）。この処理の結果、複合サービスデータは 4284 件となった。

なお、本節では実データ例ではなく、データ構造の構成概念（スキーマ）を Listing として示す。原子サービス（Web API）データの構成概念を Listing 1, 複合サービス（Mashup）データの構成概念を Listing 2 に示す。

また、複合サービスデータには欠損が含まれるため、「複合サービスの構成サービス（使用 API）が原子サービスデータに含まれない」レコードを除外するフィルタリングを行った。

Listing 1: 原子サービス（Web API）データの構成概念

```
1 {
```

```

2   "<atomic_service_name>": [
3       [
4           "<main_category>",
5           "<sub_category>",
6           "...",
7       ],
8       "<service_description>",
9       [
10          "<provider>",
11          "...",
12      ],
13      [
14          "<consumer>",
15          "...",
16      ],
17      <service_id>
18  ]
19 }

```

Listing 2: 複合サービス (Mashup) データの構成概念

```

1 {
2   "<mashup_service_name>": [
3       [
4           "<category>",
5           "...",
6       ],
7       "<service_description>",
8       [
9           "<used_api_name>",
10          "...",
11      ],
12      <service_id>
13  ]
14 }

```

4.1.2 評価指標

提案手法（CNP に基づくマルチエージェント方式）で生成される API 推薦結果の品質を評価するため、適合率（Precision）、再現率（Recall）、および F1 スコアを用いる。各プロトコル（マネージャ主導・コントラクタ主導・協調型）について、マッシュアップ要件から生成した Plan と、ProgrammableWeb 上で当該マッシュアップに紐づく使用 API 集合（正解集合）を照合し、適合率、再現率、F1 スコアの算出を行った。

照合は API 名一致に基づく集合評価として行い、正解集合を G 、生成された推薦 API 集合を P とすると、 TP, FP, FN はそれぞれ以下で定義される。

$$TP = |P \cap G|, \quad FP = |P \setminus G|, \quad FN = |G \setminus P| \quad (1)$$

このとき、適合率、再現率、F1 スコアは式 (1)–(3) から算出される。

$$\text{適合率} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{再現率} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{F1 スコア} = 2 \cdot \frac{\text{適合率} \cdot \text{再現率}}{\text{適合率} + \text{再現率}} \quad (4)$$

ここで、 TP （True Positive）は推薦した API のうち正解に含まれる API 数、 FP （False Positive）は推薦したが正解に含まれない API 数、 FN （False Negative）は正解に含まれるが推薦できなかった API 数を表す。各マッシュアップについて算出した適合率・再現率・F1 スコアを用いて、各プロトコルの API 推薦精度を比較した。

4.2 実験結果

各方式で生成された Plan を正解計画と照合し、計画生成精度を比較した。その結果、要求タイプ（粒度・曖昧さ）によって有利なタスク分解方式が異なる傾向が確認され、単一方式に固定するよりも、要求タイプに応じて方式を切り替える運用が有効であることが示された。また、入札エージェントの組織化により、専門性と分担効率のトレードオフが観測され、API カテゴリを限定した専門エージェントを用いることで、単一エージェント方式と比較して推薦の正

確性が向上し、大規模サービス集合に対してもスケーラブルに動作することが確認された。

4.2.1 マネージャ主導タスク分解

マネージャ主導方式の結果を示す。タスク分解をマネージャが一貫して行うことで、合成計画の整合性を保ちやすい一方、要求が曖昧な場合には、コントラクタ側の知識を十分に引き出せない場合がある。

4.2.2 コントラクタ主導タスク分解

コントラクタ主導方式の結果を示す。コントラクタが担当領域の API 仕様に基づいて分解を提案するため、局所的な知識を反映した提案が得られる一方で、提案間の統合にはマネージャ側での整合性確認が重要となる。

4.2.3 協調型タスク分解

協調型方式の結果を示す。マネージャによるタスク分解とコントラクタによるカテゴリ分解を組み合わせることで、整合性と柔軟性の両立を図る。

第5章 考察

5.1 単一エージェントとの比較

単一の汎用 LLM エージェントに多数の API 仕様を集約する方式では，プロンプト肥大化と文脈長制約により推論精度が低下しやすい．提案手法では API 仕様を分散保持させることで，担当領域の文脈を局所化し，推薦の正確性を維持しながらスケールさせる設計の有効性を確認した．

5.2 コントラクタエージェント数

エージェント数と担当 API 数を変化させた場合，専門性の深さ（担当範囲の狭さ）と分担効率（エージェント数の少なさ）にトレードオフが生じる．カテゴリ単位の担当は，単一 API 単位と比較して通信量を抑えつつ，一定の専門性を維持できる設計として有効である．

5.3 今後の課題

今後の課題として，(1) 公告範囲の動的制御や入札の事前フィルタリング等による通信効率化，(2) 要求タイプ推定に基づくプロトコル自動選択，(3) Plan の入出力整合性検査の強化（型・スキーマの導入）などが挙げられる．

第6章 おわりに

本論文では、LLM を用いたマルチエージェントシステムに基づく Web サービス合成手法を提案した。契約ネットプロトコルに基づくタスク分解メカニズムを導入し、マネージャ主導・コントラクタ主導・協調型という三種類のタスク分解プロトコルを設計・比較した。また、入札エージェントの組織化として、単一 API 単位の割当てに加え、カテゴリ単位で複数 API をまとめて担当させる設計を比較した。

ProgrammableWeb の複合サービスデータを用いた評価により、要求タイプ（粒度・曖昧さ）に応じてタスク分解方式を選択することが、単一方式に固定する場合よりも有効であることを確認した。さらに、API 仕様を分散保持する専門エージェントを用いることで、単一エージェント方式と比較して推薦の正確性を高めつつ、大規模サービス集合に対してもスケーラブルに動作することを示した。

今後は、通信効率化、プロトコル自動選択、Plan の整合性検査の強化などに取り組むことで、より実用的で柔軟なサービス合成システムの実現を目指す。

参考文献

- [1] Zeng, L., Benatallah, B., Ngu, A. H. H., Dumas, M., Kalagnanam, J. and Chang, H.: QoS-aware middleware for Web services composition, *IEEE Transactions on Software Engineering*, Vol. 30, No. 5, pp. 311–327 (2004).
- [2] Carman, M. A., Serafini, L. and Traverso, P.: Web Service Composition as Planning, *ICAPS 2003 Workshop on Planning for Web Services*, pp. 1636–1642 (2003).
- [3] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *arXiv preprint* (2018). arXiv:1810.04805.
- [4] Mikolov, T., Chen, K., Corrado, G. and Dean, J.: Efficient Estimation of Word Representations in Vector Space, *arXiv preprint* (2013). arXiv:1301.3781v3.
- [5] Malinen, M. I. and Fränti, P.: Balanced K-Means for Clustering, *Structural, Syntactic, and Statistical Pattern Recognition*, pp. 32–41 (2014).
- [6] Davies, D. L. and Bouldin, D. W.: A cluster separation measure, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-1, No. 2, pp. 224–227 (1979).