

G-11

生成 AI を用いた企業のニーズと研究室の持つシーズの マッチングシステム

A matching system using GraphRAG between corporate needs and research lab seeds

奥村 拓喜
Hiroki Okumura

佐野 睦夫 †
Sano Mutsuo

あらし

本稿では、大阪工業大学が主催する「ソーシャル・オープンイノベーションチャレンジ」（通称「ソイチャレ」）における企業と大学研究室の効果的なマッチングを支援するシステムを提案する。従来の産学連携マッチングは、情報の非対称性やヒューマンリソースの限界といった課題を抱えている。本研究では、企業が持つ具体的なニーズと、大学研究室が持つ技術的なシーズを結びつけるため、GraphRAG (Retrieval Augmented Generation) 技術を核としたシステムを開発する。具体的には、ソイチャレ参加企業の資料と大学研究室の概要情報から知識グラフを構築し、それに基づいて関連性の高いマッチング候補を導出、さらにその過程を可視化することで、関係者の意思決定を支援する。これにより、効率的かつ高精度な産学連携の促進に貢献することを目指す。

1. はじめに

1.1 研究背景

近年、グローバル化と技術革新の加速は、企業に持続的な成長のための新たな価値創造を強く要求している。この要求に応える上で、大学や研究機関が有する専門的な知識や先進的な技術シーズと、企業が抱える具体的な課題や潜在的なニーズを結びつける産学連携は、オープンイノベーションを推進する重要な手段としてその重要性を増している。

しかしながら、効果的な産学連携のマッチングを実現することは容易ではない。特に中小企業においては、自社のニーズを適切に言語化するリソースや、広範な大学研究の中から自社に合致するシーズを探索する手段に限りがある場合が多い。これにより、有用な連携機会が逸失されるといった課題が指摘されている。また、産学連携における情報収集や調整は、対面での面談が最も良い方法であると認識されているものの、これはコーディネーターの負担増や、参加者の時間的損失といった無視できない課題を伴う。これらの課題は、人的リソースに依存した従来のシステムでは解決が困難である。

このような背景から、効率的かつ精度の高いマッチングを可能にするシステムの開発が喫緊の課題となっている。本研究では、大阪工業大学で実施されている、地域社会の課題解決を目指す産学連携プログラム「ソーシャル・オープンイノベーションチャレンジ」（通称「ソイチャレ」）に焦点を当てる。ソイチャレは、企業や自治体、NPOなどが抱える課題（ニーズ）と、大阪工業大学の研究室が持つ技術や知見（シーズ）をマッチングさせることで、具体的なプロジェクトを推進することを目的としている。

1.2 研究目的

本研究の目的は、ソイチャレの枠組みにおいて、GraphRAG 技術を活用し、企業ニーズと研究室シーズを効率的かつ高精度にマッチングさせる意思決定支援システムを構築することである。具体的には、以下の3点を主要な目的とする。

- 資料からの情報抽出: ソイチャレ参加企業から提出される PDF や将来的には Word, PowerPoint といった多様な形式の資料から、企業ニーズの概要を効率的に把握する仕組みを構築する。
- GraphRAG による高精度なマッチング: 企業ニーズと研究室シーズの情報を統合した知識グラフを構築し、GraphRAG を用いて、それらの間の潜在的な関連性を抽出し、最適なマッチング候補を提示する。
- マッチング過程の可視化と意思決定支援: マッチング結果だけでなく、その導出に使用された知識グラフを可視化することで、ユーザーがマッチングの根拠を理解し、より質の高い意思決定を行えるよう支援する。

本システムを通じて、ソイチャレにおける産学連携の効率化と成功率向上に貢献することを目指す。

2. 関連研究

本研究には大まかに分けて 2 つの要素があり、「企業と研究機関のマッチング」と「Graph RAG」である。以下にこれらについての関連研究を取り上げる。

2.1 企業と研究機関のマッチングに関する先行研究

産学連携における企業ニーズと研究機関シーズのマッチングは、古くから多くの研究がなされてきた領域である。

山本[1]は、特許と論文のデータを抽出し、分析を行っている。そのうち、テキストマイニングを用いて、情報の可視化を実施している。

谷口[2]は、当時の産学連携に関する政策や状況を鑑み、新たな体系化の方法を提案、実践している。

堀江ら[3]は、テキストマイニングを用いて、特許および論文情報を統合解析、可視化することで、四国 4 県、特に

高知県および愛媛県における研究開発の動向を分析している。

内海[4]は、産学連携を通じたオープン・イノベーションにおいて、どのようなプロセスで技術ニーズとシーズがマッチングするのかを調べている。

澤田[5]は、産学連携における産と学の衝突や、「ニーズとシーズのマッチング」への問題点を提唱している。

このように、産学連携についての研究は多数あるが、多くのものが10年以上前の論文であり、その間に産学の在り方は大きく変化している。また、後述の Graph RAG を利用した産学マッチングシステムというものも筆者が観測した範囲では見当たらなかった。この2点を、本研究を行う理由であり新規性とした。

2.2 Graph RAG (Retrieval Augmented Generation) に関する研究

Retrieval Augmented Generation (RAG) は、大規模言語モデル (LLM) の出力精度と信頼性を向上させるためのフレームワークとして近年注目されている。

Chuangtao[6]らは、従来の類似性に基づくスキーママッチング手法や、LLM の幻覚問題に対し有効な解決手段として、KG-RAG4SM という RAG を使用した有効な解決手段を提案している。

Qinggang[7]らは、専門分野に LLM を適用する方法として、GraphRAG を使用することで Graph 検索が強化できるかを検証している。

Liu[8]らは、キーワード生成とグラフを組み合わせることにより、教育現場での意思決定支援をおこなっている。

このように、現状 RAG を使った研究は海外研究が多く、国内では非常に少ない。産学連携をテーマにしたものも国内海外ともに少ないため、この点を本研究の新規性とした。

3. システムの詳細

3.1 システム概要

本研究で開発するシステムは、大阪工業大学が実施する「ソーシャル・オープンイノベーションチャレンジ (ソイチャレ)」における、企業ニーズと大学研究室シーズのマッチングを目的としている。本システムは、以下の基本的な流れで動作する：

1. 資料の入力：ソイチャレに参加する企業の資料 (PDF 等) と、大阪工業大学の研究室概要をまとめた JSON ファイルをシステムに入力する。
2. 概要の把握：入力された資料から、企業ニーズおよび研究室シーズの概要を把握する。
3. GraphRAG によるマッチング：把握した概要情報に基づき、GraphRAG を用いて関連性の高い研究室をマッチングし、提案する。
4. グラフの可視化：マッチング結果の導出に使用された知識グラフを可視化する。

以前の構想では、チャットボットによるニーズ・シーズの対話的抽出を考えていたが、GraphRAG を用いたマッチングメカニズムの確立を優先するため、現在は既存の資料からの情報抽出に重点を置いている。

3.2 システムの流れ

本システムは、以下の流れで動作する。

1. データ収集

ソイチャレ参加企業が提出する、自企業の特徴や、課題についての概要が書かれた PDF をシステムの所定のファイルに入れる。また、それとは別に大阪工業大学情報科学部の研究室について、それぞれの研究室の概要や特色、扱っている技術などを json ファイルにまとめ、システムの所定のファイルに入れる。

2. 解析

企業から得た資料から、概要をテキストとして抽出する。抽出はシステム内で ChatGPT の API を使用して行われる。この時点で、企業の要求書類の把握は完了している想定である。

3. グラフ構築

解析された企業ニーズと研究室シーズの概要情報に基づき、グラフを構築する。企業、研究室、関連キーワードなどがノードとなり、関連性がエッジとして定義される。この時点で構築されたグラフを、画像として出力し、可視化を行う。

4. GraphRAG マッチング

構築されたグラフから、関連性が高いもの同士をマッチングし、出力する。出力された結果どのような経緯でそれらがマッチングしたのかなどの理由を ChatGPT により説明する。

3.3 システム利用に伴う前準備

本システムを利用するにあたり、事前準備として用意しておく必要があるものがある。以下に示す。

企業資料: 大阪工業大学ソイチャレに参加する企業から提供される資料。現在は、PDF のみ読み取れるようになっていたため、powerpoint や word の形式で提出されている資料は、システムの利用者が PDF に変換してから利用する想定である。これらの資料から、企業が抱える課題、技術的な要望、目標、既存の取り組みなどのニーズを抽出する。

研究室情報: 大阪工業大学内の各研究室の概要をまとめた JSON ファイル。これには、研究室の専門分野、過去の研究実績、得意な技術、保有する設備、教員の専門性などのシーズ情報が含まれる。現状、この中身の情報は大学の研究室紹介および、研究室の HP の内容を ChatGPT に読み込ませ、指定の形式で出力させる方法を取っている。これは、システム利用者の知識やバイアスによる、研究室について間違った情報を入力してしまうリスクを避けるためのものである。

3.4 GraphRAG を用いたマッチングメカニズム

本システムの中核は、GraphRAG によるマッチングメカニズムにある。企業資料から抽出されたニーズ情報と、研究室情報から抽出されたシーズ情報を基に、統一された知識グラフを構築する。例えば、企業ニーズ (例: VR や AR を用いた地域振興) と研究室シーズ (例: 「VR 技術」, 「AR を使用した研究」) は、それぞれ「VR」「AR」という単語単位でノードとなり、関連性のあるものとして接続

される。マッチングは、特定の企業ニーズに対して、研究室シーズから、関連するものがあつた場合に行われる。複数適応する研究室がある場合は複数マッチングする。その後、知識グラフから得られた情報を基に、企業ニーズと研究室シーズの関連度を評価し、マッチングの根拠となる詳細な説明を生成する。

4. 実験

本システムを制作するにあたり、システムが正しく実装するか、想定通りの挙動を示すか等を試すプレ実験を行った。実験環境と使用した資料などを示したのち、実験内容と結果について示す。

4.1 実験環境

使用したデバイスは大学で使用しているデスクトップ PC、開発環境、及び実行環境は Google Colab にて行った。

4.2 使用した資料

ソイチャレの管理をを担当している教員から、2025年度のソイチャレに参加している自治体や企業の説明用資料、及び対応研究室の表を提供していただいたため、それを使用する。今回はプレ実験と言う事で、対応表の中でもより研究室の専門性がより活かされていると担当教員が判断した14種類の組み合わせについて検証する。具体的な資料の数として、別の研究室が同じ企業の課題を担当している例もあるため、企業の資料は8種類使用する。それに対する研究室の json ファイルについては、対応表にある研究室8種類と、対応表にない研究室2種類、合計10種類の研究室を、3.3の項に示した方法で形式を合わせたものを使用する。

4.3 実験内容

上記の資料をシステムに入れ、3.2に示した手順を行い、実際にマッチング、及びグラフの可視化ができるかどうかを実験する。ある程度のマッチングができた場合、対応表と照らし合わせ、実際にマッチした研究室とこのシステムでマッチした研究室がどの程度一致しているかを調べる。この一致率が高いほど、信頼性のあるシステムというように評価する。

4.4 ドキュメント解析モジュール

企業の要求書類 (PDF) を読み込み、テキストデータを抽出する。抽出されたテキストは、ChatGPT API に入力され、その内容の要約および主要なキーワード抽出を行う。使用しているライブラリは PyPDF2、指示しているプロンプトは以下の通り。

以下の内容を簡潔に要約してください。(要約時)、以下の企業要求の要約から、重要なキーワードを5~10個程度抽出してください。カンマ区切りで回答してください。コア技術、応用分野、プロジェクトの目標に焦点を当ててください。(キーワード抽出時)

4.5 ナレッジグラフ構築モジュール

ドキュメント解析モジュールで得られた企業の要求に関する情報と、情報収集・整理モジュールで得られた研究室の情報を基に、ノードとエッジを持つナレッジグラフを構築する。ノードについては3種類で分けており、企業の資料、研究室、「VR」などのキーワードのようになる。エッジに関しては2種類で分けており、企業ノードとキーワ

ードノード間を「requires」、研究室ノードとキーワードノード間を「specializes_in」としている。ライブラリは NetworkX を使用している。

4.6 マッチングモジュール

構築されたナレッジグラフと ChatGPT API を連携させ、最適なマッチング候補を生成する。企業の要求に関連するキーワードをグラフ内で探索し、関連性の高い研究室ノードを抽出し、出した企業要求の概要、関連する研究室情報、およびマッチングのための質問プロンプトを ChatGPT API に送信する。質問プロンプトは以下の通り。

あなたは企業と研究室のマッチングを専門とするコンサルタントです。提供された企業の要求と研究室の情報を精査し、最も適合性の高い研究室を最大3つ提案してください。直接的なキーワードの一致だけでなく、**研究内容や技術の応用可能性、関連分野の専門性なども考慮し、類推して提案**してください。各提案について、その適合理由を具体的なキーワードや研究内容の関連性に基づいて詳細に説明してください。厳格な以下の形式で回答してください。もし適切な研究室が全く見つからない場合は、「適切な研究室は見つかりませんでした。」と回答してください。提案1: [研究室名]-理由: [具体的な適合理由を詳細に記述。例: 「AR技術の専門知識が観光促進に直接応用可能であるため。」] 提案2: [研究室名]-理由: [具体的な適合理由を詳細に記述] 提案3: [研究室名]-理由: [具体的な適合理由を詳細に記述]

4.7 結果表示・可視化モジュール

構築したナレッジグラフを可視化する。システム全体のマッチング状況を示す全体グラフと、マッチングした場合は、その企業に特化した個別マッチングサブグラフの2種類を出力する。グラフにおいて、ノードはそれぞれ企業の資料を水色、研究室の資料を緑色、キーワードを赤色で表示する。エッジは、4.5で示した文言を赤文字で表示する。

4.8 実験結果

4.3の内容に基づいて、実験を行った結果を以下に示す。1回目

1回目の実験のみ、システムの動作確認のため4.2で示した資料の数より少ない資料で行っているが、その他条件は同じである。

マッチング結果に関しては、全て「適切な研究室は見つかりませんでした。」となった。出力されたグラフの画像は、多くのノードが、文字化けしてしまっており、何のキーワードがグラフ上でより近くなっているかわからない状態であった。

2回目

前回の結果を踏まえ、日本語対応をするため、環境の変更やフォントの変更などを行ったが、問題は改善しなかったため、1回目と同じ条件で、企業資料の名称と、jsonファイルの中身を英語にすることによって対応した。その結果、マッチング結果に関して、1つの企業資料と2つ研究室がマッチングした。その結果出力された画像が以下の2つである。

のため、大規模言語モデルである ChatGPT の推論能力を活用した以下のアプローチを試みた。

企業ニーズの技術的解決策への変換: 企業の概要と主要キーワードから、情報技術を活用した具体的な解決策や応用分野の視点を ChatGPT に提案させた。これにより、企業の抽象的なニーズを研究室の専門性に合致しやすい技術的視点へと変換する試みを行った。

キーワードの自動置き換えによる接続性の強化: 企業が提案する技術的解決策のキーワードについて、システム内の全研究室が持つキーワードリストの中から、ChatGPT に最も意味的に近い単一のキーワードを特定させ、そのキーワードに置き換えて知識グラフに組み込む手法を導入した。これは、同義語や関連語、略語の差異を ChatGPT が吸収し、知識グラフにおける企業と研究室の接続性を向上させることを目的とした。

本アプローチは、今まで GraphRAG にて完全一致のみマッチングしていたところを、ChatGPT の推論能力を活かし、ある程度類推してマッチング提案が行えるような意図で行っている。このアプローチを元に、追加実験を行った。

実験結果

実験の結果、いくつかの企業資料でマッチング結果が提案されるようになった。さらに、3 組の組み合わせが、実際のマッチングと同じ結果が出力された。これは、本システムが目的とする「類推的マッチング」が、限定的ながらも実現し始めたことを示している。以下に出力例を示す。

デバッグ: 企業 '10 14' に見つかった潜在的な研究室数: 2 研究室名: ['System Information Laboratory', 'Interaction Design Laboratory']

デバッグ: GPT からの生応答:

提案 1: Interaction Design Laboratory - 理由: 大阪水都コンソーシアムが求める観光促進や水辺の活性化において、コミュニケーションロボットが重要な役割を果たす可能性があります。この研究室は人間の感情や行動を理解する技術に焦点を当てており、観光客の心理や行動を理解し、より豊かな体験を提供することができます。さらに、障害者支援や認知症ケアなど、観光客の多様なニーズにも対応できる可能性があります。

提案 2: System Information Laboratory - 理由: 大阪水都コンソーシアムの持つ課題に対して、最適化やシミュレーション、機械学習といったシステム情報の研究が有効であると考えられます。特に、IoT 技術を活用した水辺エリアの効率的な運営やビジターエクスペリエンスの向上、行動理解に基づくサービス提供などに応用が可能です。システム工学のアプローチにより、水辺エリアの持続可能な開発や観光施設の効率向上に貢献できるでしょう。

また、出力されたグラフ画像においては、日本語出力設定時と同様に、ノードラベルの日本語部分が文字化けしていることが確認された。図 4 とほぼ同じ内容であるため、画像は省略する。また、マッチングしているにもかかわらず、個別のグラフの画像化はされなかった。

7. 実験 2 の考察

実験 2 では、完全一致の問題を解決するため、ChatGPT

を使用した推論を目的として行った。結果、多くの組み合わせでマッチングが行われるようになり、本システムの目的に近づいたと言える。ただし、グラフの可視化などについて、このマッチング方法と連携しきれていないため、画像が出力されない。それに加え、今回の試行は類推の可能性を示したが、マッチング結果に偏りが見られる。特定の研究室が多くマッチングしたり、全く出てこない吸湿もある。これは、マッチング方法の他に、json ファイルを作成する際に参考にしてしている研究室の情報量に差があることも原因として考えられる。

8. 今後の課題

本システムについて、今後解決が必要な課題点を以下に示す。

第 1 に、文字化けとマッチング問題の解決が挙げられる。考察の項に示したように、現状堂々巡りの状態であり、この問題が解決しない限りマッチングとグラフの可視化の問題が解決しない。文字化けを解消しても、文字が密集し過ぎて可視化の意味をなしていないのも問題である。別の日本語対応の方法などの解決策が求められる。

第 2 に、マッチング方法の検証が挙げられる。実験 2 において、マッチング数の増加と結果の一致率が向上したが、アプローチ自体はかなり強引な面も強く、この方法で適切かを担当教員と話し合うなどして検証するがある。

最後に、一致率の検証と向上がある。上記の 2 つの課題が解決したのち、一致率の向上を図る必要がある。GRAPHRAG のファインチューニングを行い、マッチングの精度を向上させていく。

9. まとめ

本研究で提案した生成 AI を用いたマッチングシステムは、大阪工業大学の「ソーシャル・オープンイノベーションチャレンジ (ソイチャレ)」における企業ニーズと研究室シーズの連携を効率化するための新たな試みである。本システムは、企業の要求と研究室の情報を知識グラフとして構築し、ChatGPT を活用することで、最適なマッチング候補を提案する。

実験の結果、システムの基本的な機能は確認できたものの、いくつかの重要な課題が明らかになった。特に、グラフ可視化における日本語の文字化け問題と、マッチングにおける「キーワード完全一致」への依存が挙げられる。日本語での出力設定は、特定の事例でマッチングを可能にしたが、これは限定的なキーワードの一致によるものであり、グラフ表示の文字化けが再発した。一方、英語での出力設定は文字化けを解消したが、マッチング精度が著しく低下した。この結果は、日本語の原資料を扱う場合における、大規模言語モデルの言語処理の複雑性と、グラフ RAG の推論における語彙の一貫性の重要性を示唆している。

今後、本システムのさらなる発展には、これらの課題への取り組みが不可欠である。特に、グラフ可視化における日本語表示の根本的な解決策の確立、および生成 AI による「類推的なマッチング」能力の強化は、システムの信頼性と実用性を高める上で重要である。さらに、より多様な研究室情報や企業課題を取り込み、実際の運用を通じて蓄積されたデータと知見を活用することで、ソイチャレにおける産学連携の成功率向上にどれだけ貢献できるかを実証す

ることが求められる。

参考文献

- [1] 山本 外茂男,産学連携のマッチング性分析におけるテキストマイニングの有効性, 情報の科学と技術, 59 巻 6 号 p. 291-297.2009
- [2] 谷口 邦彦,産学官連携によるイノベーション創出プログラムの構築 I: 新たな価値創造に向けたニーズとシーズのマッチング(PM イノベーション,日本発信の仕組み P2M コンセプト明確化と発展的展開-イノベーションを促進し,競争力再生と人材育成に貢献する-),国際 P2M 学会研究発表大会予稿集, 2009 春季
- [3] 堀江 進之助, 上田 昌平, 佐藤 暢, 野田口 真也, 落合 圭, 山口 祐穂, 坂内 悟, 特許情報と文献情報の統合解析による技術情報の可視化-技術情報から見た高知県及び愛媛県における研究開発及び産学連携の現状とその動向-, 情報プロフェッショナルシンポジウム予稿集,2009
- [4] 内海 京久, オープン・イノベーションの技術ニーズ・シーズマッチングー産学連携による MEMS 新事業創造事例の分析ー, AAOS Transactions J-STAGE 早期公開論文,2025
- [5] 澤田 芳郎, 大学モデルと産学連携コンフリクト,産学連携学, 1 巻 1 号 p. 5-8,2004
- [6] Chuangtao Ma Sriom Chakrabarti Arijit Khan and Bálint Molnár Knowledge Graph-based Retrieval-Augmented Generation for Schema Matching 2025
- [7] Qinggang Zhang Shengyuan Chen Yuanchen Bei Zheng Yuan Huachi Zhou Zijin Hong Junnan Dong Hao Chen Yi Chang Xiao Huang A Survey of Graph Retrieval-Augmented Generation for Customized Large Language Models 2025
- [8] Liu Zhuchenyang AcademicRAG: Knowledge Graph Enhanced Retrieval-Augmented Generation for Academic Resource Discovery 2025