大規模言語モデル（LLM）による Few-shotクラスタリングの可能性

VISWANATHAN Vijay　GASHTEOVSKI Kiril　LAWRENCE Carolin　WU Tongshuang　NEUBIG Graham

要　旨

半教師ありクラスタリングは従来の教師なしクラスタリングとは異なり、ユーザーがデータに意味のある構造を与えられるようにします。これは、クラスタリングアルゴリズムをユーザーの意図（インテント）に合致させることに役立ちます。しかし、既存のアプローチでは、クラスターを改善するためにエキスパートによる大量のフィードバックを必要としていました。本稿では、大規模言語モデル（LLM）がエキスパートのガイダンスを強化し、クエリ効率の高いFew-shot半教師ありテキストクラスタリングを実現できるかどうかを検討します。最初に、LLMがクラスタリングの改善に非常に効果的であることを示します。次に、LLMをクラスタリングに組み込む際の3つの段階として、クラスタリング前（入力特徴の改善）、クラスタリング中（クラスターに制約を提供）、クラスタリング後（LLMを使用した修正）の検討を行います。LLMを最初の2つの段階に組み込むことでクラスター品質が定常的に大きく改善し、またLLMによって所望のクラスターを生成する際にユーザーがコストと精度の兼ね合いを考慮できるようになることが明らかになりました。

KeyWords　クラスタリング／少数ショット／Few-shot／人間中心

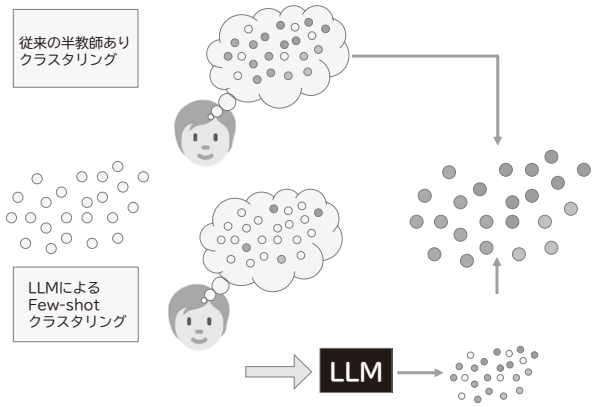
1. はじめに

教師なしクラスタリングは、不可能なタスクを行うことを目的としています。すなわち、ドメインエキスパートのニーズを特定せずに、そのニーズを満たすようにデータを整理することです。クラスタリングはその性質上、基本的には条件が決まっていない問題です。リッチ カルアナ氏の「Clustering: probably approximately useless?」[1]によれば、条件が決まっていないことにより、クラスタリングは「おそらくほぼ役に立たない」とされています。

一方で、半教師ありクラスタリングは、ドメインエキスパートがクラスタリングアルゴリズムを誘導できるようにすることでこの問題を解決しようとしています[2]。これまでの研究では、エキスパートとクラスタリングアルゴリズムとの間で異なるタイプの相互作用が導入されています。例えば、手作業で選ばれたシードポイントでクラスターを初期化する[3]、ペアワイズ制約を指定する[4) 5]、特徴のフィードバックを提供する[6]、クラスターを分割または統合する[7]、または1つのクラスターを固定し、残りを精緻化する[8]などがあります。これらのインタフェースはすべて、エキスパートが最終的なクラスターを制御できることを示しています。ただし、これにはエキスパートの労力がかなり必要です。例えば、トイデータセットに対するシミュレーション[8]では、分割/統合、ペアワイズ制約、ロック及び精緻化の相互作用について、どのクラスタリングアルゴリズムでも、ユーザーの仕様と一致するクラスターを生成するのに20～100回のフィードバック相互作用が必要でした。大規模な実世界データセットでは、可能なクラスターが多いため、インタラクティブなクラスタリングアルゴリズムに必要なフィードバックコストが膨大になる可能性があります。

本研究では、LLMを人間の意思決定のノイズシミュレーションとして使用する最近の一連の研究に基づき[9) -11]、半教師ありテキストクラスタリングに対して異なるアプローチを提案します。具体的には、「エキスパートがLLMに望む相互作用（例えば、ペアワイズ制約）のデモンストレーションをいくつか提供した後、LLMがクラスタリングアルゴリズムを指示することは可能か」という調査課題への回答を試みます（**図1**）。

また、テキストクラスタリングプロセスの3つの段階であるクラスタリング前、クラスタリング中、及びクラスタリング後におけるLLM活用の可能性の検討を行います。クラスタリング前には、テキスト表現を拡張することでLLMを活用します。各例に対して、LLMを使用してキーフレーズを生成し、これらのキーフレーズをエンコードしてベース



従来の半教師ありクラスタリングでは、ユーザーはクラスターに大量のフィードバックを提供します。一方、本研究のアプローチでは、ユーザーは少量のフィードバックでLLMにプロンプトを与えます。その後、LLMはクラスターに対して大量の擬似フィードバックを生成します。

図1　従来の半教師ありクラスタリングとLLMによるFew-shotクラスタリング

の表現に追加します。クラスタリング中には、クラスター制約を追加することでLLMの組み込みを行います。半教師ありクラスタリングのための古典的なアルゴリズムを採用することで、LLMをペアワイズ制約の擬似オラクルとして使用します。そして、ペアワイズ制約の擬似オラクルを使用して信頼度の低いクラスター割り当てを修正することにより、クラスタリング後にLLMを使用することを検討します。すべてのケースにおいて、ユーザーが記述してLLMにプロンプトを提供することによって、ユーザーとクラスタリングアルゴリズムの相互作用が可能になります。

これらの3つの方法を、エンティティの標準化、意図（インテント）によるクエリのクラスタリング、トピックによるツイートのグループ化という3つのタスクにわたり、5つのデータセットにおいて評価しました。その結果、文書の埋め込みに対する従来のK-Meansクラスタリングと比較して、LLMを使用して各文書の表現を豊かにすると、対象のすべてのデータセットにおいてあらゆる指標でクラスター品質が経験的に向上することが明らかになりました。LLMがペアワイズの類似性を判定できる場合には、LLMをペアワイズ制約の擬似オラクルとして使用することは非常に効果が得られる可能性がありますが、効果を発揮するにはより多くのLLMクエリが必要です。しかし、LLMによる修正がもたらす利点は限定的です。重要なのは、LLMはわずかなコストで、人間のオラクルを使用した従来の半教師ありクラスタリングの性能に近づくことが可能ということです。

更にその簡単さにおいて、近年のディープラーニングに基づくテキストクラスタリング手法[12) 13)]とは一線を画しています。本研究の3つの方法のうち2つ（LLMを文書の表現を拡張するために使用する方法及びLLMをクラスタリングの出力を修正するために使用する方法）は、どのようなテキストクラスタリングアルゴリズムに対してどのようなテキスト特徴のセットを使用しても、プラグインとして追加できます[*1]。LLMプロンプトのどの側面がクラスタリングの挙動に最も影響を与えるかを調査した結果、単に指示を（デモンストレーションなしで）使用するだけでも、大きな付加価値があることが分かりました。これにより、自然言語による指示をクラスタリングアルゴリズムに統合することに向けた将来の研究方向に、モチベーションを与えることが期待されます。

2. LLMの組み込み手法

第2章では、クラスタリングにLLMを組み込むために本研究で使用する手法について紹介します。

2.1 LLMキーフレーズ拡張によるクラスタリング

通常、エキスパートはクラスタリングを行う前に、各文書のどの側面をとらえたいかを既に把握しています。これらの重要な要素をクラスタリングアルゴリズムにゼロから抽出させるのではなく、これらの側面を事前に全体的に強調する（それによってタスクの重点を指定する）ことが重要です。このことを実現するために、LLMを使用してクラスタリングのニーズに関連するエビデンスでテキスト表現を強化・拡張し、各文書のテキスト表現をタスク依存にします。具体的には、キーフレーズを生成するLLMに各文書を通過させます。これらのキーフレーズは埋め込みモデルによってエンコードされ、キーフレーズの埋め込みは元の文書の埋め込みに連結されます。

GPT-3（具体的には、gpt-3.5-turbo-0301）を使用してキーフレーズを生成します。LLMには指示で始まる短いプロンプトを与えます。例えば、「オンラインバンキングいく

[*1] その一方で、ペアワイズ制約クラスタリングでは、基礎となるクラスタリングアルゴリズムとしてK-Meansを使用する必要があります。

つかのクエリを、それらが同じインテントを表現しているかどうかに基づいてクラスタリングしようとしています。各クエリについて、インテントを表現できる包括的なキーフレーズのセットをJSON形式のリストとして生成してください」のプロンプトです。この指示の後には、**図2**の上半分の例に示されているようなキーフレーズのデモが4つ続きます。

その後、LLMによって生成されたキーフレーズを単一のベクトルにエンコードし、このベクトルを元の文書のテキスト表現と連結します。より優れたエンコーダを利用して、LLMから知識を分離するために、元のテキストと同じエンコーダを使用してキーフレーズをエンコードします[*2]。

このアプローチは、マールテン ラエディット氏らの「IDAS: Intent Discovery with Abstractive Summarization」[14]による同時期の研究と同様のものであり、そこでは著者らは教師なしでインテントを発見するためにキーフレーズを生成しています。

2.2 擬似オラクルペアワイズ制約クラスタリング

半教師ありクラスタリングにおける最も一般的なアプローチは、恐らくペアワイズ制約クラスタリングと言えるでしょう。このアプローチでは、ユーザーの持つ抽象的なクラスタリングインテントが具体的なフィードバックから暗黙的に導出されるように、リンクする必要があるかまたはリンクできないポイントのペアを、オラクル（例えば、ドメインエキスパート）が選択します[15]。つまり、ユーザーは、どの種類のポイントをグループ化するかを概念的に説明し、そして最終的なクラスターがこのグループ化に従うようにします。本研究ではこのパラダイムを使用して、LLMがクラスタリング中にエキスパートによるガイダンスを強化する可能性を検討します。その際、LLMを擬似オラクルとして使用します。

分類するペアを選択するために、エンティティの正規化と他のテキストクラスタリングタスクに対して異なる戦略を採用しています。テキストクラスタリングの場合、Explore-Consolidateアルゴリズム[4]を適用して、まず埋め込み空間から多様なペアを（リンクできないポイントのペアを特定するために）収集します。次に、既に選択されたポイントの近傍のポイントを（リンクが必要なポイントのペアを見つけるため）収集します。クラスターが非常に多くリンクが必要なポイントが非常に少ないエンティティを正規化するために、埋め込み空間で最も近いが異なるポイントのペアのみサンプリングします。

LLMに簡潔なドメイン固有のプロンプトを与え、その後にテストセットのラベルから取得したペアワイズ制約のデモを最大4回行います。これらのペアワイズ制約を使用して、PCKMeansアルゴリズム[4]によってクラスターを生成します（**図3**）。このアルゴリズムは、任意の制約に違反するクラスター割り当てに対して、ハイパーパラメータwで重み付けされたペナルティを適用します。そして、シイカーヤシシュトゥ氏らの「CESI: Canonicalizing Open Knowledge Bases Using Embeddings and Side Information」[16]に従い、このパラメータを各データセットの検証分割で調整します。擬似オラクルのペアワイズ制約の信頼性が低いという可能性があるため、クラスターの初期化には、先行研究[4]のようにペアワイズ制約の近傍構造を直接に使用せず、代わりにK-Means++[17]を使用します。

2.3 LLMを使用したクラスタリングの修正

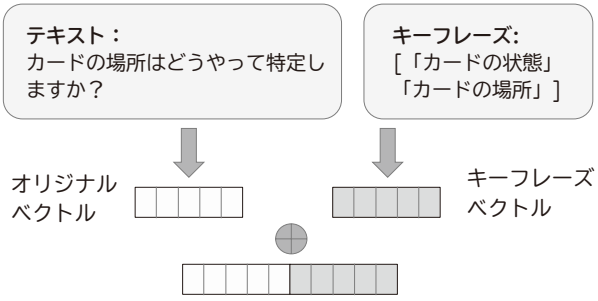


図2　キーフレーズ埋め込みとの連結による文書表現の拡張

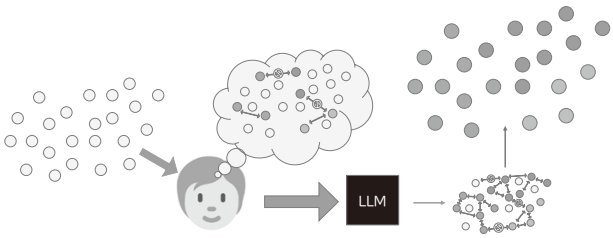
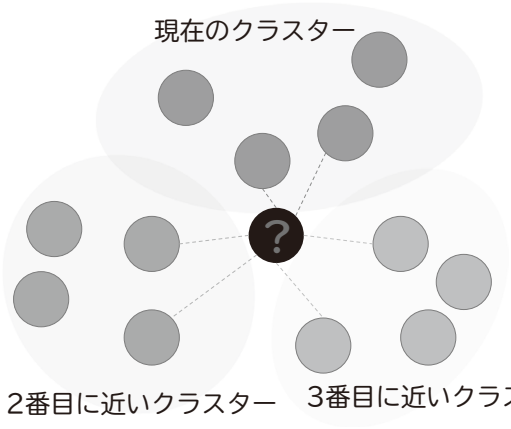


図3　最大4つの有効なペアワイズ制約のLLMを使用した特定データセットのペアワイズ制約の生成

[*2] 例外的なケースとして、エンティティのクラスタリングが挙げられます。ここでは、BERTエンコーダはWikipediaの文章をクラスタリングするために特化されていますので、キーフレーズのクラスタリングをサポートするためにDistilBERTを使用しています。



クラスタリングを実行した後、信頼度の低いポイントの特定を行います。これらのポイントについて、現在のクラスター割り当てが正しいかどうかをLLMに問い合わせます。LLMからの応答が否定的だった場合、代わりにこのポイントを上位5つの最も近いクラスターのいずれかにリンクすべきかどうかをLLMに確認し、それに応じてクラスタリングを修正します。

図4　LLMを使用したクラスタリングの修正手順

最後に、既存のクラスターセットがあるが、局所的な変更を最小限に抑えてその品質を向上させたいという状況を検討します。この状況を実現するために、第2章2節と同じペアワイズ制約の擬似オラクルを使用します。この手順を**図4**で示します。

信頼度の低いポイントを特定するために、最も近いクラスターと2番目に近いクラスターの間のマージンが最も小さいポイントkを見つけます（実験ではk=500と設定）。各クラスターは、埋め込み空間内でそのクラスター中心点（セントロイド）に最も近いエンティティによってテキスト的に表現されます。信頼度の低い各ポイントに対しては、そのポイントが現在割り当てられているクラスターの代表的ポイントと正しくリンクされているかどうかまずLLMに問い合わせます。もしLLMが、このポイントは現在のクラスターにリンクされるべきでないと予測した場合、埋め込み空間で2番目に近い4つのクラスターを近接度によりソートし、再ランキングの候補とみなします。現在のポイントを再ランキングするために、このポイントを各候補クラスターの代表的ポイントにリンクすべきかどうかLLMに問い合わせます。LLMからの応答が肯定的だった場合、このポイントを新しいクラスターに割り当て直します。LLMがすべての代替案に対して否定的に応答する場合、既存のクラスター割り当てを維持します。

3. タスク

3.1 エンティティの正規化

タスク： エンティティの正規化では、名詞句のコレクション$M = \{m_i\}_1^N$を、m_1及びm_2が同じエンティティを指す場合にのみ、$m_1, m_2 \in C_j$となるようなサブグループ$\{C_j\}_1^K$にグループ化する必要があります。例えば、名詞句“President Biden”（m_1）、“Joe Biden”（m_2）、及び“the 46th U.S. President”（m_3）が同じエンティティを指している場合、これらは1つのグループ（例えば、C1）にクラスタリングされることになります。名詞句のセットMは、通常OIE（Open Information Extraction）システムによって生成された「オープンナレッジグラフ」のノードです[*3]。関連するエンティティリンキングのタスク[18) 19)]とは異なり、対象のすべてのエンティティがキュレーションされたナレッジグラフ、地名辞典や百科事典に含まれているとは仮定していません。

エンティティの正規化は、半教師ありクラスタリングの課題に取り組むのに有用です。ここでは、数百または数千のクラスターがあり、クラスターごとに比較的少数のポイントしかなく、難しいクラスタリングタスクとなっています。

データセット： 次の2つのデータセットを使用して実験を行います。

- OPIEC59k[20)]には22,000の名詞句（2,138の一意のエンティティ表層形式を含む）があり、これらは490のグラウンドトゥルースクラスターに属しています。名詞句はMinIE[21) 22)]によって抽出され、グラウンドトゥルースのエンティティクラスターは同じWikipediaの記事にリンクするアンカーテキストです。
- ReVerb45k[16)]には15,500のメンション（12,295の一意のエンティティ表層形式を含む）があり、これらは6,700のグラウンドトゥルースクラスターに属しています。名詞句はReVerb[23)]システムの出力であり、「グラウンドトゥルースクラスター」のエンティティクラスターはエンティティをFreebaseナレッジグラフに自動的にリン

[*3] オープン情報抽出（OIE）は、スキーマフリーな方法で自然言語テキストから表層形式（主語、関係、目的語）トリプルを抽出するタスクです。

クさせたものです。ラベリングエラーを取り除くために、手動でフィルタリングを行ったウェイ シェン氏らの「Multi-View Clustering for Open Knowledge Base Canonicalization」[20]のデータセットのバージョンを使用します。

正規化の指標：ウェイ シェン氏らの「Multi-View Clustering for Open Knowledge Base Canonicalization」[20]が使用している標準的な指標に従います。

- マクロ適合率と再現率
  - 適合率（Prec）：予測されたクラスターのすべての要素が同じゴールドクラスターに属している割合はどれくらいか
  - 再現率（Rec）：ゴールドクラスターのすべての要素が同じ予測クラスターに属している割合はどれくらいか
- ミクロ適合率と再現率
  - 適合率（Prec）：予測クラスターの過半数と同じゴールドクラスターに属するポイントの数はどれくらいか
  - 再現率（Rec）：ゴールドクラスターの過半数と同じ予測クラスターに属するポイントの数はどれくらいか
- ペアワイズ適合率と再現率
  - 適合率（Prec）：予測されたリンクが実際にゴールドクラスターによってリンクされたポイントのペアの数はどれくらいか
  - 再現率（Rec）：ゴールドクラスターによってリンクされたポイントのペアのうち、予測されたリンクもされているペアの数はどれくらいか

最後に、各ペアの調和平均を計算して、マクロF1、ミクロF1、及びペアワイズF1を得ます。

3.2 テキストクラスタリング

タスク：次に、短いテキスト文書のクラスタリングのケースを検討します。このクラスタリングタスクは文献で広く研究されています[24]。

データセット：この設定では、以下の3つのデータセットを使用します。

- Bank77[25]には、77あるインテントカテゴリのオンラインバンキングアシスタントに対する、3,080のユーザークエリが含まれています。
- CLINC[26]には、「スコープ外」のクエリを削除[13]した150あるインテントカテゴリのタスク指向対話システムに対する、4,500のユーザークエリが含まれています。
- Tweet[27]には、89あるカテゴリの2,472のツイートが含まれています。

指標：先行研究[12]に従い、テキストクラスターをグラウンドトゥルースと比較するために、正規化した相互情報量及び精度（ハンガリアンアルゴリズム[28]を使用して、グラウンドトゥルースと予測クラスターの最良のアラインメントを求めることにより得られる）を使用します。

4. ベースライン

4.1 埋め込み上のK-Means

本研究の手法は、K-Means++クラスター初期化[17]を用いてエンコードしたデータの、K-Meansクラスタリング[29]のベースライン上に構築されています。使用する特徴とクラスターセンターの数は、主に先行研究に従って、タスクごとに選択します。

エンティティの正規化：先行研究[16][20]に従い、エンティティへの言及（例：「古代ギリシャ人により紀元前600年にマルセイユが建都されて以来」）それぞれを、特定の言及コンテキストに関係なく、一意の表層形式（例：「マルセイユ」）をグローバルに表現することによってクラスタリングします。一意の表層形式をクラスタリングした後、このクラスターマッピングを個々の言及（個々の文から抽出されたもの）に合成して、言及レベルのクラスターを得ます。

「マルチビュークラスタリング」アプローチ[20]に基づき、各名詞句をインターネットからのテキスト言及とOIEシステムから抽出された「オープン」なナレッジグラフを使用して表現します。彼らはBERTエンコーダ[30]を使用して、エンティティが発生するテキストコンテキストを表現し（「コンテキストビュー」）、TransEナレッジグラフエンコーダ[31]を使用して、オープンナレッジグラフ内のノードを表現します（「ファクトビュー」）。上記論文の著者らは、共参照エンティティからの弱い教示を使用してBERTエンコーダをファインチューニングし、ナレッジグラフ上でのデータ拡張を使用してナレッジグラフ表現を改善することにより、こ

れらのエンコーダを改善します。各エンティティの2つのビューは、表現を生成するために組み合わされます。

前述の論文では、1つのビューで計算されたクラスター割り当てを使用してもう一方のビューでクラスターのセントロイドを初期化する、交互マルチビューのK-Means手順が提案されています。一定回数の反復後、もしビューごとのクラスタリングが一致しない場合、著者らは「コンフリクト解決」手順を実行して、両方のビューで低い慣性を持つ最終的なクラスタリングを求めます。本研究が副次的に貢献したことの一つは、このアルゴリズムを簡略化したことです。著者らが微調整したエンコーダを使用し、各ビューからの表現を連結し、K-Means++初期化[17]を使用して、共有ベクトル空間でK-Meansクラスタリングを行うだけで、元の論文で報告された性能を達成できることを確認しました。

最後に、クラスターセンターの数に関しては、Log-Jumpメソッド[20]に従い、OPIEC59k及びReVerb45kについて、それぞれ490と6,687のクラスターを選択します。

インテントクラスタリング：Bank77及びCLINCデータセットに関して、ウーウェイ ジャン 氏らの「ClusterLLM: Large Language Models as a Guide for Text Clustering」[13]に従い、Instructorエンコーダで各ユーザークエリをエンコードします。エンコーダをガイドするために簡単なプロンプト「Represent utterances for intent classification」を使用します。先行研究に従い、CLINC及びBank77については、それぞれ150と77のクラスターを選択します。

ツイートクラスタリング：デジャオ ジャン 氏らの「Supporting Clustering with Contrastive Learning」[12]に従い、文の類似度分類用にファインチューニングされたDistilBERT[32]のバージョンを使用して各ツイートをエンコードします[33] *4。その際89のクラスターを使用します[12]。

4.2 対照学習を使用したクラスタリング

第2章で紹介した方法に加えて、以前に発表されたSCCL[12]とClusterLLM[13]の2つのテキストクラスタリング手法も採用します。これら両方の方法は、ディープエンコーダの対照学習を使用してクラスターを向上させ、本研究で提案される方法よりも大幅に複雑で計算集約的です。SCCLは深い埋め込みクラスタリング[34]と教師なし対照学習を組み合わせて、テキストから特徴を学習します。ClusterLLMは、LLMを使用して学習された特徴を改善します。階層的クラスタリングを実行した後、これらの方法ではまた、LLMからの3重フィードバック（「ポイントAはポイントBよりもポイントCに似ていますか?」）を使用して、クラスターの階層からクラスターの粒度を決定し、フラットなクラスターセットを生成します。これらのアプローチとの比較を効果的にするために、本研究ではSCCL及びClusterLLMの先行研究で報告されたものと同じエンコーダを使用します。Bank77及びCLINCに対してはInstructor[35]、ツイートに対してはDistilBERT（文の類似度分類用にファインチューニングされたもの）[32] [33]です。

5. 結果

5.1 結果の概要

表1にエンティティ正規化の実験結果を、そして**表2**にテキストクラスタリングの結果を示します[*5]。ここから、テキスト表現を拡張するためにLLMを使用することが最も効果的であり、両方の正規化データセットで最高水準の結果を達成し、すべてのテキストクラスタリングデータセットでK-Meansベースラインを大幅に上回っていることが分かります。LLMによって擬似ラベリングされ

表1　LLMをエンティティ正規化に統合する手法の比較

データセット/手法		OPIEC59k				ReVerb45k			
		マクロF1	ミクロF1	ペアF1	平均	マクロF1	ミクロF1	ペアF1	平均
最適クラスタリング		80.3	97.0	95.5	90.9	84.8	93.5	92.1	90.1
CMVC		52.8	90.7	84.7	76.1	66.1	87.9	89.4	81.1
K-Means		53.5±0.0	91.0±0.0	85.6±0.0	76.7	69.6±0.0	89.1±0.0	89.3±0.0	82.7
独自	キーフレーズクラスター	60.3±0.0	92.5±0.0	87.3±0.0	80.0	72.3±0.0	90.2±0.0	90.0±0.0	84.2
独自	PCKMeans	58.7±0.0	91.5±0.0	86.1±0.0	78.7	72.0±0.0	88.5±0.0	87.0±0.0	82.5
独自	LLMによる修正	58.7	91.5	85.2	78.4	69.9	89.2	88.4	82.5

CMVCはマルチビュークラスタリング法[20]を指し、K-Meansは同じ手法を簡略化して再実装した本研究の手法を指します。
該当する場合、標準偏差は異なるシードでクラスタリングを5回実行して得られます。

*4　このモデルはHugging Faceのdistilbert-base-nli-stsb-mean-tokensです。
*5　第4章で述べたように、エンティティ正規化を行う際、同じエンティティ表層形式（例：“Marseille”）を含む言及がいくつかあればそれらを同じクラスターに割り当てます。これは先行研究[16][20]に従ったものです。このアプローチは多義的な名詞句に対して削減不可能なエラーを引き起こす可能性があります（例：“Marseille”はサッカークラブのOlympique de Marseille、または都市のMarseilleを指す可能性がある）。

表2　LLMをテキストクラスタリングに統合するための手法の比較

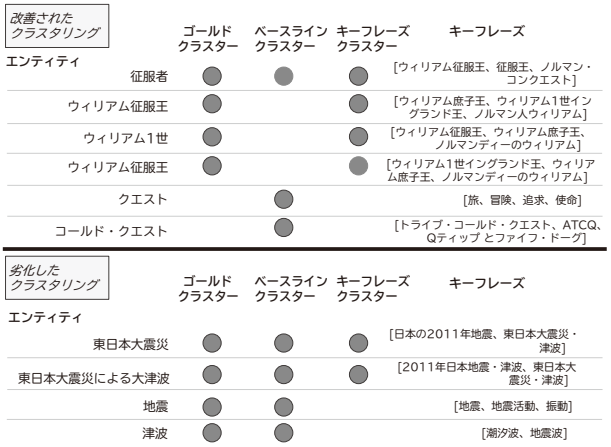
データセット/手法		Bank77		CLINC		Tweet	
		精度	正規化相互情報	精度	正規化相互情報	精度	正規化相互情報
SCCL		-	-	-	-	78.2	89.2
クラスターLLM		71.2	-	83.8	-	-	-
K-Means		64.0±0.0	81.7±0.0	77.7±0.0	91.5±0.0	57.5±0.0	80.6±0.0
独自	キーフレーズクラスター	65.3±0.0	82.4±0.0	79.0±0.0	92.6±0.0	62.0±0.0	83.8±0.0
独自	PCKMeans	59.6±0.0	79.6±0.0	79.6±0.0	92.1±0.0	65.3±0.0	85.1±0.0
独自	LLMによる修正	64.1	81.9	77.8	91.3	59.0	81.5

“SCCL”はデジャオ ジャン氏らの「Supporting Clustering with Contrastive Learning」[12]の手法を指し、“ClusterLLM”はウーウェイ ジャン氏らの「ClusterLLM: Large Language Models as a Guide for Text Clustering」[13]の手法を指します。本研究の実験では、これらの手法と同じ基本エンコーダを使用しています。可能な場合、クラスタリングを異なるシードで5回実行することにより標準偏差を得ています。

表3　各クラスタリングアルゴリズムの出力をグラウンドトゥルースと対応させた後の改善または劣化したクラスターの数

手法	改善	劣化
K-Means	0	0
キーフレーズクラスター	168	83
PCKMeans	155	82
LLMによる修正	102	51

対応するグラウンドトゥルースクラスターとのジャッカード類似度によって測定。
各アルゴリズムは490のクラスターを生成しました。



各エンティティのキーフレーズをエンコードしてクラスタリングした後に、変更されたクラスターを特定します。クラスターには、エンティティ名とエンティティに関するテキストコンテキストの両方を提供していますが、ここでは表示上の制限からテキストコンテキストを省略しています。

図5　キーフレーズの拡張例

た20,000のペアワイズ制約が提供された場合、ペアワイズ制約K-Meansは5つのデータセットのうち3つで（OPIEC59kで現行の最高水準を上回る）強力なパフォーマンスを発揮します。次では、各手法が効果を持つ、あるいは持たない要因について、より詳細な分析を行います。

5.2 実例と主要な要因

各LLMベースの変更がクラスタリングプロセスに与える影響を定性的に調査するために、OPIEC59kデータセットを使用して、さまざまなクラスタリング戦略から得られたクラスターとK-Meansベースラインから得られたクラスターの比較を行います。

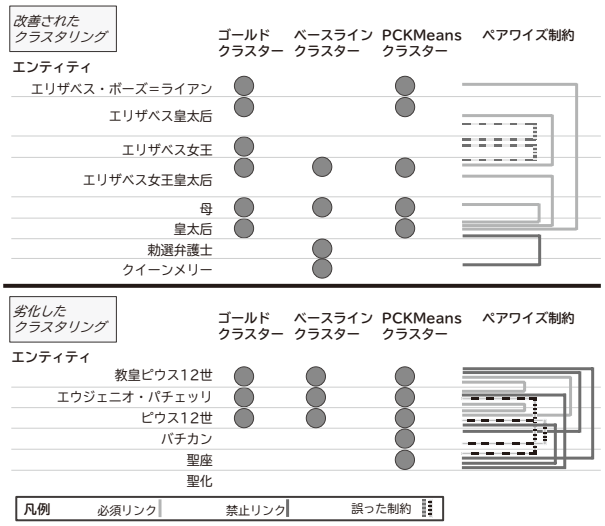
ハンガリアンアルゴリズム[28]を使用して、各クラスタリングをグラウンドトゥルースに対応させた後、各予測クラスターとそれに対応するグラウンドトゥルースクラスターの間のジャッカード類似度を計算します。LLMベースの介入によって得られたクラスターをベースラインのK-Meansクラスターと比較することで、介入により最も大きな改善があるクラスターと、介入により最も大きな劣化が起こるクラスターを特定します[*6]。

改善されたクラスターと（K-Meansベースラインに対して）劣化したクラスターを1つずつ示していますが、これらは均等な割合で発生するわけではありません。**表3**に手法ごとに改善されたクラスターと劣化したクラスターの数を示します。**図5**、**6**、及び**7**では、キーフレーズの拡張、ペアワイズ制約の組み込み、及びLLMによる修正の後のクラスターの例を示し、それらを使用して各アルゴリズムに影響を与える主要な要因に対するインチュイションを提供します。OPIEC59kでは、LLMベースの介入が主にクラスターの改善につながることが明らかです。

キーフレーズクラスタリング：曖昧性解消のために適切な粒度を提供

図5では、LLMが生成したキーフレーズが、エンティティの曖昧性を効果的に解消できることが示されています（例：“Conqueror”（征服者）と“Quest”（クエスト）に対して大きく異なるキーフレーズを生成し、埋め込みベースのベースラインクラスタリングはこれらを誤ってグループ

[*6] 評価中のクラスターの整合性の問題による可能性があるため、どちらかのアルゴリズムからの出力が対応するグラウンドトゥルースクラスターとゼロの重複を持つクラスターは無視します。



ペアワイズ制約を組み込んだ後に、変更されたクラスターを特定し、擬似オラクルによって生成された関連するペアワイズ制約を示します。

図6　ペアワイズ制約の組み込み例

LLMを用いてクラスター割り当てを修正した後に、変更されたクラスターを特定します。

図7　LLMによる修正の後のクラスターの例

表4　OPIEC59kを含むいくつかのデータセットにおけるペアワイズ制約の精度

データセット/指標	OPIEC59k	Tweet	Bank77
データサイズ	2138	4500	3080
ペア拘束条件での総合精度	86.7	96.8	81.7
LLM割当数	109	78	108
再割当の精度	55.0	89.7	41.7

各データセットの上位500ポイントを再ランキングする際、LLMが元のクラスタリングと異なることは稀です。異なる場合、LLMが誤っている可能性が高くなります。

表5　デモまたは指示がない場合のLLMが介入する効果の比較

データセット/手法	OPIEC59k	CLINC	
	平均F1	精度	正規化相互情報
キーフレーズクラスター	80.0	79.4 ±0.0	92.6 ±0.0
指示なしの場合	79.1	78.4 ±0.0	92.7 ±0.0
デモなしの場合	79.8	78.7 ±0.0	91.8 ±0.0
指示あり-ベースモデル	---	74.8 ±0.0	90.7 ±0.0
指示あり-大型モデル	---	77.7 ±0.0	91.5 ±0.0
指示あり-超大型モデル（参考文献35）	---	77.2 ±0.0	91.9 ±0.0
指示あり-超大型モデル（GPT-3.5の補助あり）	---	70.8 ±0.0	88.6 ±0.0

同じプロンプトを与えた場合でも、GPT-3.5をベースとしたキーフレーズクラスタリングは、異なるサイズの指示でファインチューニングしたエンコーダよりも優れた性能を発揮します。

化してしまう）。劣化した例では、これらのキーフレーズがテキストの文脈よりも各エンティティの表層形式に過度に焦点を当てる可能性があることも認められます。これは、複雑なドキュメントにおいてキーフレーズを活用する際に、モデリングとプロンプト技法をより精密にする余地があることを示唆しています。

PCKMeans：不正確で矛盾した制約が大きな影響を持ち得る

図6に示すように、改善されたケースでは、LLMはいくつかのポイント間の関係（例：“Mother”（母）と“Queen Mother”（皇太后））を、埋め込みに対するK-Meansクラスタリングではグループ化されなかったものを正確に特定します。劣化したケースでは、LLMが矛盾する制約を生成し、誤検出につながるケースが見られます。LLMは“Eugenio Pacelli”（エウジェニオ・パチェッリ）と“Pius XII”（ピウス12世）がリンクされなければならず、かつ“Pius XII”と“Holy See”（聖座）がリンクできないと正確に予測しますが、“Eugenio Pacelli”と“Holy See”の間にリンクがあると誤って予測します。これら矛盾した制約がある結果として、PCKMeansアルゴリズムは更なるポイントを誤ってクラスターにグループ化します。**表4**では、OPIEC59kを含むいくつかのデータセットにおけるペアワイズ制約の精度を示します。

LLMによる修正：最終的で硬直的な制約により過度な修正に至る可能性がある

図7の劣化したクラスターでは、LLMは、特定の式典で授与された特定の賞でなくアカデミー賞全体に焦点を当てるべきですが、このクラスターの粒度を理解できていません。LLMによるOPIEC59kの修正が全体的に効果を上げているにもかかわらず（表3）、この例によって、各ポイントについてLLMから確定的な判断を取り出すというこの方法の欠点が際立っています。

確定的であるというこの性質は、LLMによる修正の効果に影響を与えます（**表5**）。表1と表2では、この手法は、データセットに対しすべての指標についてわずかな利得を一貫してもたらしており、改善幅は0.1から5.2の絶対ポイントの範囲内です。表4では、LLMに最も不確実性の高いクラスター割り当ての上位500を再考するよう指示すると、LLMはごく少数のケースでのみポイントを再割り当てします。LLMのペアワイズオラクルは通常正確ですが、元のクラスタリングの信頼度が既に低いポイントに対しては極端に不正確です。

5.3 アブレーションスタディ：なぜLLMはテキスト拡張で優れているのか？

表1及び表2は、キーフレーズクラスタリングが最も強力なアプローチであり、5つのデータセットのうち3つで最良の結果（他の2つのデータセットでは2番目に強力な手法である擬似オラクルPCKMeansに匹敵するパフォーマンス）を達成することを示しています。これは、LLMがクラスタリングを容易にするためにテキストの内容を拡張するのに有用であることを示唆しています。

それでは、LLMがこの機能において有用である原因は何でしょう？ タスク固有のモデリング指示を指定できる能力、デモンストレーションを通じて暗黙的に類似性関数を指定できる能力、または小規模なニューラルエンコーダには欠けている知識を持っていること、どれがその理由でしょうか。

この質問に答えるためにアブレーションスタディを行います。OPIEC59kとCLINCについては、キーフレーズクラスタリング法の使用を考慮しますが、プロンプトから指示またはデモンストレーション例のどちらかを省略します。CLINCについては、小さなエンコーダに短い指示を与えることができるInstructorモデルの特徴に基づくK-Meansクラスタリングと比較します。

指示とデモンストレーションは相補的な利得をもたらす

経験的には、LLMにプロンプトとして指示またはデモンストレーションのどちらかを与えることで、クラスターの品質を向上させることが可能になりますが、両者を与えることにより、最も一貫性のあるポジティブな効果が得られます。質的には、指示を与えながらデモンストレーションを省略すると、一貫性の低い大規模なキーフレーズのセットが生成される一方で、指示なしでデモンストレーションを与えると、より焦点の合ったキーフレーズのグループが生成されますが、所望の側面（例：トピックvsインテント）が反映されないことがあります。

指示によるファインチューニングされたエンコーダは十分な知識を提供できない

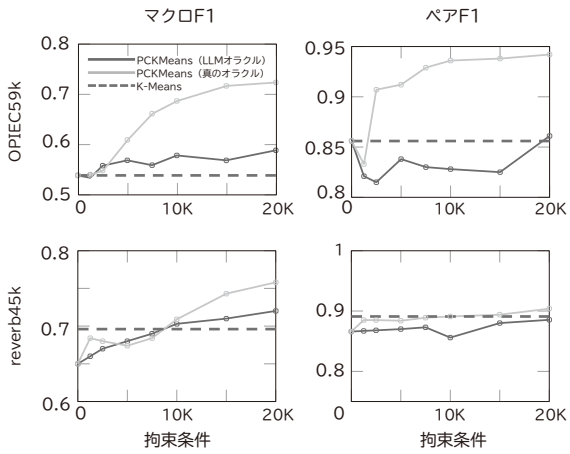
なぜGPT-3.5を使用したキーフレーズクラスタリングが、指示のみ（デモンストレーションなし）の設定において、Instructor（指示によりファインチューニングされたエンコーダ）を使用する場合よりも優れているのでしょうか。控えめなスケーリングカーブは、単にスケールが唯一の要因ではないことを示唆しています。つまり、GPT-3.5はGPT-3（175B）と同等かそれ以上のパラメータを含んでいる可能性が高く、一方でホンチャン ス氏らの「One Embedder, Any Task: Instruction-Finetuned Text Embeddings」[35] のInstructor-base/large/XLはそれぞれ110M、335M、及び1.5Bのパラメータを含んでいます。

使用されるプロンプトには次の2種類があることに注目してください。GPT-3.5に使用したプロンプトは非常に詳細ですが、Instructorに対するプロンプトは元の設計に従って簡潔なものになっています（例：“Represent utterances for intent classification”（インテント分類のための表現を示してください））。更に、Instructor-XLにGPT-3.5用のプロンプトを与えて実験しました（表5の一番下の行）。GPT-3.5に与えるプロンプトに対してInstructor-XLのパフォーマンスが悪いことが分かります。指示によりファインチューニングされた現行のエンコーダは、Few-shotクラスタリングを容易にする詳細でタスク固有のプロンプトをサポートするには不十分であることが推測されます。

表6　異なるLLMを用いたクラスタリングアプローチの擬似ラベリングコストの比較

手法	データサイズ	コスト（米ドル）		
		PCKMeans	修正	キーフレーズ
OPIEC59k	2138	$42.03	$12.73	$2.24
ReVerb45k	12295	$33.81	$10.24	$10.66
Bank77	3080	$10.25	$3.38	$1.23
CLINC	4500	$9.77	$2.80	$0.95
Tweet	2472	$11.28	$3.72	$0.99

2023年6月にOpenAIのgpt-3.5-turbo-0301 APIを使用しました。



OPIEC59kにおいて、ペアワイズ制約K-Meansに対する擬似オラクルのフィードバックをより多く収集することで、他の指標を低下させることなくマクロF1指標を改善します。真のオラクル制約を持つ同じアルゴリズムと比較すると、このアルゴリズムのノイズの多いオラクルに対する感度が分かります。

図8　ペアワイズ制約K-Meansのマクロ F1 の向上の実現

5.4 LLMを擬似オラクルとして使用すると費用対効果が高い

LLMをクラスタリングプロセスのガイドとして使用することで、クラスター品質を向上できることが分かりました。しかし、LLMは高額であり、市販のLLM APIをクラスタリング中に使用するとクラスタリングプロセスに追加のコストがかかります。

本研究の3つのアプローチを使用した場合の、LLMのフィードバック収集の擬似ラベリングのコストを**表6**に示します。本研究で提案する3つのアプローチの中で、LLMを使用してペアワイズ制約に擬似ラベリングを行う（LLMが20,000組のポイントを分類する必要がある）ことが最もLLM APIコストが高い方法です。PCKMeansとLLMによる修正ではどちらもデータセットごとに同じ回数だけLLMにクエリを行いますが、キーフレーズ修正のコストはデータセットのサイズに応じて線形にスケーリングされ、非常に大規模なコーパスのクラスタリングは実現不可能になります。

このコストを正当化するほどの性能向上があるのでしょうか。 LLMの代わりに人間のエキスパートをクラスタリングプロセスのガイドに雇用した場合、同等のコストでより良い結果を得ることができるでしょうか。本研究の実験では、擬似ラベリングのペアワイズ制約が最も高いAPIコストを必要とするため、このアプローチをケーススタディとして取り上げます。擬似オラクルのフィードバックが十分に与えられた場合、**図8**で示されているように、ペアワイズ制約K-Meansは、ペアワイズまたはマイクロF1を大幅に低下させることなく、マクロF1の向上を実現できます（これはクラスター純度の向上を示唆しています）。

このコストは妥当でしょうか。 OPIEC59kのOpenAI APIにかかる42ドルの費用（表6）に対して、時給11ドル[36]であると仮定すると、ラベリングの作業スタッフを約3.8時間雇うことができます。アノテーターは1分間に約3組のペアをラベリングできることが分かっています。したがって、作業スタッフの賃金が42ドルの場合は、GPT-3.5が20,000のラベリングをするのと同じコストで、人間により700以下のラベルが生成されることになります。

図8のフィードバック曲線によると、この価格帯でGPT-3.5は真のオラクルがあるペアワイズ制約よりも格段に効果的であることが分かります。真のオラクルによってラベリングされた少なくとも2,500組のペアが提供されない限り、ペアワイズ制約K-Means法はエンティティの標準化に対して何の価値も提供できません。これは、実験のパフォーマンスを最大化することが目的の場合、LLMでクエリを行う方がラベリングする人を雇うよりも費用対効果が高い可能性があることを示唆しています。

6. むすび

LLMをシンプルな方法で使用すると、さまざまなテキストクラスタリングタスクにおいて、クラスター品質を一貫して向上させることができるという結果が得られました。文書表現を豊かにする手段としてLLMが常に有用であることが判明したことで、本研究でのシンプルな概念実証が、

LLMを介する文書拡張に向けたより精緻なアプローチを促がすことにつながるものと考えます。

7. 謝辞

本研究はNEC Laboratories Europeからのフェローシップにより実施されました。Wiern Ben Rim、Saujas Vaduguru、Jill Fain Lehmanのご指導に謝意を表します。また、Chenyang Zhaoには、本研究への貴重なフィードバックを提供していただいたことを感謝いたします。

参考文献

1) Rich Caruana: Clustering: probably approx-imately useless?, CIKM '13: Proceedings of the 22nd ACM international conference on Information & Knowledge Management, pp.1259–1260, 2013 https://dl.acm.org/doi/10.1145/2505515.2514692
2) Juhee Bae, Tove Helldin, Maria Riveiro, Sławomir Nowaczyk, Mohamed-Rafik Bouguelia and Göran Falkman: Interactive Clustering: A Comprehensive Review, ACM Computing Surveys, Volume 53 Issue 1, pp.1-39, 2020 https://dl.acm.org/doi/abs/10.1145/3340960
3) Sugato Basu, Arindam Banerjee and Raymond J. Mooney: Semi-supervised Clustering by Seeding, The 19th International Conference on Machine Learning (ICML-2002), pp.19-26, 2002 https://www.cs.utexas.edu/~ml/papers/semi-icml-02.pdf
4) Sugato Basu, Arindam Banerjee and Raymond J. Mooney: Active Semi-Supervision for Pairwise Constrained Clustering, The SIAM International Conference on Data Mining, (SDM-2004), pp.333-344, 2004 https://www.cs.utexas.edu/~ml/papers/semi-sdm-04.pdf
5) Hongjing Zhang, Sugato Basu and Ian Davidson: A Framework for Deep Constrained Clustering - Algorithms and Advances, ECML/PKDD, 2019 https://arxiv.org/abs/1901.10061
6) Sajib Dasgupta and Vincent Ng: Which Clustering Do You Want? Inducing Your Ideal Clustering with Minimal Feedback, Journal of Artificial Intelligence Research, Volume 39, pp.581-632, 2010 https://arxiv.org/abs/1401.5389
7) Pranjal Awasthi, Maria - Florina Balcan and Konstantin Voevodski: Local algorithms for interactive clustering, Journal of Machine Learning Research 18, 2017 https://www.jmlr.org/papers/volume18/15-085/15-085.pdf
8) Anni Coden, Marina Danilevsky, Daniel F. Gruhl, Linda Kato and Meena Nagarajan: A method to accelerate human in the loop clustering, SDM 2017, 2017 https://research.ibm.com/publications/a-method-to-accelerate-human-in-the-loop-clustering
9) Jinlan Fu, See-Kiong Ng, Zhengbao Jiang and Pengfei Liu: GPTScore: Evaluate as You Desire, 2023 https://arxiv.org/abs/2302.04166
10) John J.Horton: Large Language Models as Simulated Economic Agents: What Can We Learn from Homo Silicus?, Working Paper 31122, National Bureau of Economic Research, 2023 https://arxiv.org/abs/2301.07543
11) Joon Sung Park, Joseph C.O' Brien, Carrie J.Cai, Meredith Ringel Morris, Percy Liang and Michael S.Bernstein: Generative Agents: Interactive

Simulacra of Human Behavior, 2023
https://arxiv.org/abs/2304.03442
12) Dejiao Zhang, Feng Nan, Xiaokai Wei, Shang-Wen Li, Henghui Zhu, Kathleen McKeown, Ramesh Nallapati, Andrew O. Arnold and Bing Xiang: Supporting Clustering with Contrastive Learning, 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 5419-5430, 2021
https://arxiv.org/abs/2103.12953
13) Yuwei Zhang, Zihan Wang and Jingbo Shang: ClusterLLM: Large Language Models as a Guide for Text Clustering, 2023
https://arxiv.org/abs/2305.14871
14) Maarten De Raedt, Fréderic Godin, Thomas Demeester and Chris Develder: IDAS: Intent Discovery with Abstractive Summarization, 2023
https://arxiv.org/abs/2305.19783
15) Kiri L. Wagstaff and Claire Cardie: Clustering with Instance-level Constraints, The Seventeenth International Conference on Machine Learning, pp.1103-1110, 2000
https://www.wkiri.com/research/papers/wagstaff-constraints-00.pdf
16) Shikhar Vashishth, Prince Jain and Partha Talukdar: CESI: Canonicalizing Open Knowledge Bases Using Embeddings and Side Information, The 2018 World Wide Web Conference, pp.1317-1327, 2018
https://arxiv.org/abs/1902.00172
17) David Arthur and Sergei Vassilvitskii: k-means++: The Advantages of Careful Seeding, ACM-SIAM Symposium on Discrete Algorithms, 2007
https://theory.stanford.edu/~sergei/papers/kMeansPP-soda.pdf
18) Razvan C. Bunescu and Marius Pasca: Using Encyclopedic Knowledge for Named Entity Disambiguation, 11th Conference of the European Chapter of the Association for Computational Linguistics, pp.9-16, 2006
https://aclanthology.org/E06-1002/
19) David N. Milne and Ian H. Witten: Learning to link with wikipedia, I CIKM '08: Proceedings of the 17th ACM conference on Information and knowledge management, pp.509-518, 2008
https://dl.acm.org/doi/10.1145/1458082.1458150
20) Wei Shen, Yang Yang and Yinan Liu: Multi-View Clustering for Open Knowledge Base Canonicalization, 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD ' 22, pp.1578-1588, 2022
https://arxiv.org/abs/2206.11130
21) Kiril Gashteovski, Rainer Gemulla and Luciano Del Corro: Minie: Minimizing Facts in Open Information Extraction, 2017 Conference on Empirical Methods in Natural Language Processing, pp.2630-2640, 2017
https://aclanthology.org/D17-1278/
22) Kiril Gashteovski, Sebastian Wanner, Sven Hertling, Samuel Broscheit and Rainer Gemulla, OPIEC: An Open Information Extraction Corpus, Conference on Automatic Knowledge Base Construction (AKBC), 2019
https://arxiv.org/abs/1904.12324
23) Anthony Fader, Stephen Soderland and Oren Etzioni: Identifying relations for open information extraction, EMNLP '11: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp.1535-1545, 2011
https://dl.acm.org/doi/10.5555/2145432.2145596
24) Charu C. Aggarwal and ChengXiang Zhai: A Survey of Text Clustering Algorithms, Mining Text Data, pp.77-128, 2012
https://link.springer.com/chapter/10.1007/978-1-4614-3223-4_4
25) Iñigo Casanueva, Tadas Temc inas, Daniela Gerz, Matthew Henderson and Ivan Vulic´ : Efficient Intent Detection with Dual Sentence Encoders, 2nd Workshop on Natural Language Processing for Conversational AI, pp.38–45, 2020
https://aclanthology.org/2020.nlp4convai-1.5/
26) Stefan Larson, Anish Mahendran, Joseph J. Peper, Christopher Clarke, Andrew Lee, Parker Hill, Jonathan K. Kummerfeld, Kevin Leach, Michael A. Laurenzano, Lingjia Tang and Jason Mars: An Evaluation Dataset for Intent Classification and Out-of-Scope Prediction, 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp.1311–1316, 2019
https://aclanthology.org/D19-1131/
27) Jianhua Yin and Jianyong Wang: A model- based approach for text clustering with outlier detection, 2016 IEEE 32nd International Conference on Data Engineering (ICDE), pp.625-636, 2016
28) Harold W. Kuhn: The hungarian method for the assignment problem, Naval Research Logistics (NRL), Volume2, Issue1-2, pp.83-97, 1955
https://doi.org/10.1002/nav.3800020109
29) S. Lloyd: Least squares quantization in PCM, IEEE Transactions on Information Theory, Volume 28 Issue 2, pp.129–137, 1982
https://ieeexplore.ieee.org/document/1056489
30) Jacob Devlin, Ming-Wei Chang, Kenton Lee and Kristina Toutanova: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pp.4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics, 2019
https://arxiv.org/abs/1810.04805
31) Antoine Bordes, Nicolas Usunier, Alberto Garcia-Durán, Jason Weston and Oksana Yakhnenko:

Translating Embeddings for Modeling Multi-relational Data, 26th International Conference on Neural Information Processing Systems - Volume 2, NIPS' 13, pp.2787–2795, 2013
https://proceedings.neurips.cc/paper_files/paper/2013/file/1cecc7a77928ca8133fa24680a88d2f9-Paper.pdf
32) Victor Sanh, Lysandre Debut, Julien Chaumond and Thomas Wolf: DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter, 2019
https://arxiv.org/abs/1910.01108
33) Nils Reimers and Iryna Gurevych: Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks, 2019 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2019
https://arxiv.org/abs/1908.10084
34) Junyuan Xie, Ross B. Girshick and Ali Farhadi: Unsupervised Deep Embedding for Clustering Analysis, International Conference on Machine Learning, 2015
https://arxiv.org/abs/1511.06335
35) Hongjin Su, Weijia Shi, Jungo Kasai, Yizhong Wang, Yushi Hu, Mari Ostendorf, Wen-tau Yih, Noah A. Smith, Luke Zettlemoyer and Tao Yu: One Embedder, Any Task: Instruction-Finetuned Text Embeddings, 2022
https://arxiv.org/abs/2212.09741
36) Kotaro Hara, Abigail Adams, Kristy Milland, Saiph Savage, Chris Callison-Burch and Jeffrey P. Bigham: A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk, 2018 CHI Conference on Human Factors in Computing Systems, 2017
https://arxiv.org/abs/1712.05796

執筆者プロフィール

VISWANATHAN Vijay
Carnegie Mellon University

GASHTEOVSKI Kiril
NEC Laboratories Europe and Center for Advanced Interdisciplinary Research, Ss. Cyril and Methodius Uni. of Skopje

LAWRENCE Carolin
NEC Laboratories Europe
Manager

WU Tongshuang
Carnegie Mellon University

NEUBIG Graham
Carnegie Mellon University

NEC 技報のご案内

NEC技報の論文をご覧いただきありがとうございます。
ご興味がありましたら、関連する他の論文もご一読ください。

NEC技報WEBサイトはこちら

NEC技報（日本語）　NEC Technical Journal（英語）

Vol.75 No.2　ビジネスの常識を変える生成AI特集
～社会実装に向けた取り組みと、それを支える生成AI技術～

ビジネスの常識を変える生成AI特集によせて
生成AI技術への取り組み ～基盤から応用、ルール作りまで～

◇ 特集論文

急速に広まる生成AIの市場適用
NECにおける生成AIの取り組みについて
電子カルテと医療文書の作成支援による医師業務効率化
映像×LLMによる報告書作成業務の自動化
映像分析と生成AIによるリアル世界の行動理解
サイバー脅威インテリジェンス生成自動化
生成AIの社内活用を進めるNEC Generative AI Service（NGS）
ソフトウェア・システム開発への生成AIの活用
LLMとMIで革新する素材開発プラットフォーム
LLMと画像分析を活用した被災状況の把握

生成AIの可能性を高める基盤技術
日本語性能に優れたNECの大規模言語モデル（LLM）
NECの生成AIを支える国内企業で最大規模のAIスーパーコンピュータ
より安全な大規模言語モデル（LLM）を目指して
データを秘匿したまま連携できる連合学習技術とLLMへの適用可能性
大規模言語モデル（LLM）によるFew-shotクラスタリングの可能性
オープンドメイン常識推論のための知識拡張型プロンプト学習
AI連携とオーケストレーションのための基盤ビジョンLLM
クエリを考慮した新規手法により関連する企業データを減らしLLM APIの使用コストを最適化

AI技術が社会へ浸透するために
AI標準化・ルールメイクに関する動向とNECの取り組み
人権尊重に向けたNECのAIガバナンスの取り組み
RCModelを用いたAIリスクマネジメントのための人材育成事例

◇ NEC Information
2023年度C&C賞表彰式典開催

Vol.75 No.2
（2024年3月）

特集TOP