

ナレッジグラフを用いた商品名非依存型検索システムの構築と評価

Development and Evaluation of a Knowledge Graph-Based Search System Independent of Product Names

高岡 顕恒[†] 花沢 明俊[‡]
Takaoka Akinobu Hanazawa Akitoshi

1. はじめに

現在、ホームセンターをはじめとする多くの小売店で、顧客が求める商品を探すための独自の検索システムが構築されている。従来、多くの検索システムは商品名や型番、明確なブランド名を入力文章とする検索に最適化されてきたが、ユーザーが必ずしも商品名を把握しているとは限らない。また、ホームセンターなどのマニアックな商品を豊富に販売する小売店では、顧客は商品の用途や特徴は把握していても、商品名は把握していないケースが予想される。従来のキーワード検索やベクトル検索では、複雑な入力文章の意味を正確に理解するには限界があった。ここで、「巨大かつ複雑なデータを意味的に整理・関連づける」技術としてナレッジグラフに注目した。本研究では、ナレッジグラフを用いて、商品名を知らずに用途や特徴だけを表現した検索クエリにも対応できる検索手法の提案を目的とする。ナレッジグラフ上で入力クエリを整理し、類似度スコアリングを行う独自のアルゴリズムを開発した。

2. 方法

提案手法の概要を図 1 に示す。

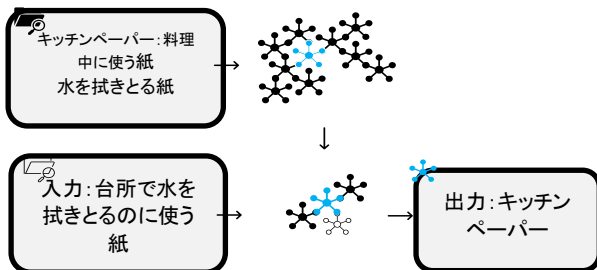


図 1 概要

2.1 商品説明テキストの作成

商品名、商品説明文のテキストセットを用意する。説明文は、商品名を含まず、50 字程度で商品用途や商品特徴を含む。

2.2 グラフデータセットの作成

・事前処理:

SpaCy により、商品説明文から名詞・動詞・形容詞を抽出し、原形化してノード化。

・ノード:

- 商品名ノード (例: キッチンペーパー)
 - 商品説明ノード (例: 水、拭く、紙)
- ※ 固有表現ノード (出現頻度の小さい単語)

・エッジ:

- 直接関連エッジ (α): ある商品名とその商品の説明文中の単語を接続。

- 間接関連エッジ (β):

ある商品と違う商品の説明文中の単語を接続

- ・重み付け:

固有表現を含む場合は補正係数 γ を掛ける。

今回の実験環境では、グリッドサーチで最適化し、 $(\alpha, \beta, \gamma) = (1.6, 2.0, 0.8)$ を採用している。

2.3 検索クエリ文章作成

商品名、検索クエリのセットテキストを用意する。検索クエリは、商品名を含まず、商品用途や商品特徴を含む。

2.4 商品検索

- ・クエリ前処理:

形態素解析で名詞・動詞・形容詞を抽出しノード化。抽象語 (例: する、こと) はリストにより除外。

- ・ノードマッチング:

- 一致ノード: グラフデータセットにあるノードクエリ語がグラフ内に存在すれば、エッジ重みを直接スコアに加算。
- 不一致ノード: グラフデータセットにないノード類似ノード (コサイン類似度が最大) を探索。経由した 2 つのエッジ重みの積をスコア加算 ※閾値 (0.4) 以上の場合のみ、類似ノードとみなす

- ・スコアリングと出力:

各商品について抽出されたノードとのエッジ重みの総加算値を求め、スコア上位の商品を検索結果として提示する。 ※閾値 (0.3) 以下のスコアはノイズとして加算しない。

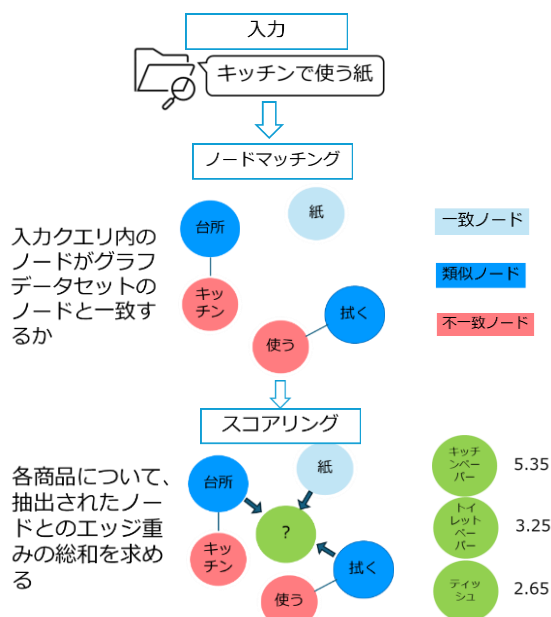


図 2 検索アルゴリズム

2.5 分析手法

適当な任意の工具 100 個について自身で作成したデータセットと、実際に工務店に存在する商品の品揃えから 100 個抽出したデータセットの 2 つの商品データセットについて精度評価を行った。また、既存検索手法 (キーワード、ベクトル) と精度比較を行った。

3. 結果

• クエリ例と結果

クエリ例: ” レザークラフトをする時に穴を補強する金具をつけるやつ”

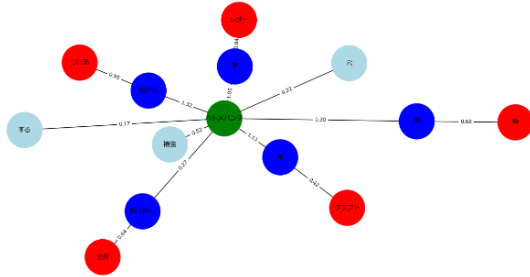


図 3 クエリ例と結果

• 自作の工具データセット

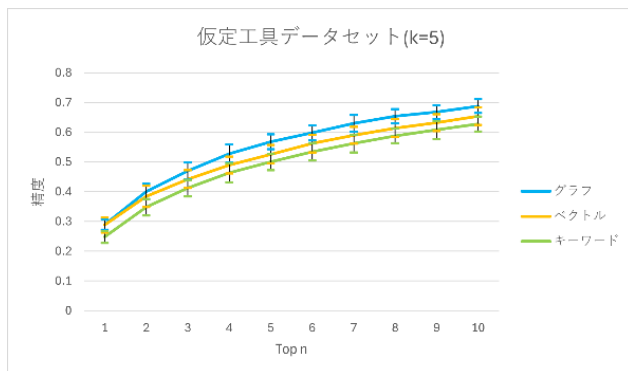


図 4 手法の比較 (自作工具データセット)

• 実際の工具データセット

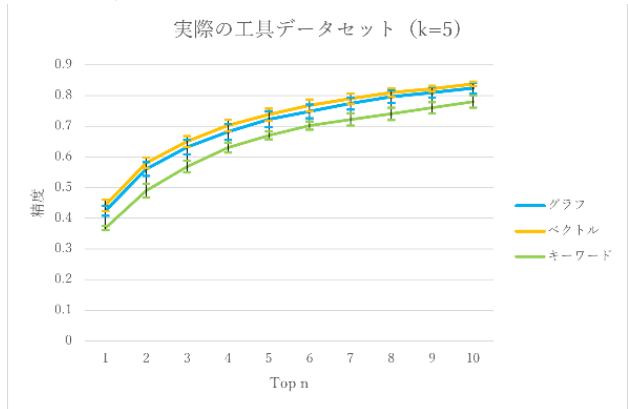


図 5 手法の比較 (実際の工具データセット)

4. 考察

本手法では、既存のベクトル・キーワード情報をもとにグラフベースの検索システムを構築した。図 4、5 の結果から、検索精度は既存手法に匹敵し、データセットによっては上回り、有意差があることが確認できた。2 つのデータセットについて平均近傍距離を調べると、自作の工具データセットは実際の工具データセットと比較して、似たような商品の集中が多いことがわかった。自作の工具データセットでは、典型的な工具のみを扱っているため、商品の埋め込み空間が近くなった。その似たような埋め込みの差違をベクトル検索では判断できないことも多く、検索精度が大きく低下した。グラフ検索も精度は低下したが、必要な情報をうまく取得できたため、ベクトル手法を上回ったと考えられる。この結果から、検索システムにおいても、多くの情報から必要な部分を整理できるグラフ構造は有用性があると考えられる。

5. 結論

本研究では、曖昧な検索クエリに対応した検索システムの構築を目標に、ナレッジグラフと自然言語処理モデルを用いて商品名非依存型の検索システムを構築し、性能の検証を行った。実験の結果、100 商品のうち 10 個提案してそのなかにも含まれる確率が、2 つのデータセットでそれぞれ、82.4%と 68.8%であった。既存手法と比較すると、1 つのデータセットでは有意差は無く、一方のデータセットでは既存手法を上回った。

今後の展望としては、他の文埋め込みモデルでの検証やより階層化したグラフ構造を用いた内部構造の更なる改善があげられる。

謝辞

本論文の執筆にあたり、株式会社グッデイ様には、研究環境やデータの提供をはじめ、多大なるご協力を賜りました。心より感謝申し上げます。

参考文献

- [1] Matthew Honnibal, Ines Montani, “spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing”, To appear, Vol.2, No.1 (2017).
- [2] Tom B. Brown, Benjamin Mann, Nick Ryder, “Language Models are Few-Shot Learners”, Advances in Neural Information Processing Systems, Vol.33, No.1 (2020).
- [3] Aric Hagberg, Daniel Schult, Pieter Swart, “Exploring Network Structure, Dynamics, and Function using NetworkX”, Proceedings of the 7th Python in Science Conference (SciPy2008) <https://networkx.org>
- [4] 株式会社グッデイ, “商品検索サイト”, グッデイ公式オンライン https://itemsearch.gooday.co.jp/index?gl=1*17fsiwi*_gcl_au*Nj_Q3NjU1NTkxLjE3NDk2MTk0MzQ (2025年10月2日アクセス) .
- [5] モノタロウ, “モノタロウ 商品検索システムの精度が向上”, モノタロウ, https://corp.monotaro.com/ir/upload_file/m009-m009_01/Improvedaccuracyofproductsearchsystem.pdf (アクセス日: 2025年10月2日) .

† 九州工業大学 宇宙システム工学科 花沢研究室
Hanazawa Lab, Department of Space Systems Engineering,
Kyushu Institute of Technology