

構造化データを用いたクエリ拡張手法に関する研究

石塚 悠真

情報検索において、検索者が入力した語と文書集合中で用いられている語彙との差異に起因する語彙ミスマッチは、検索性能低下の主要因の一つである。この問題に対し、疑似適合フィードバックや語彙埋め込みに基づくクエリ拡張手法が広く用いられてきたが、拡張語選定の根拠が不明瞭であり、検索意図と乖離した語が付加される場合があるという課題が指摘されている。

本研究では、構造化知識ベースである Wikidata が保持する明示的なプロパティ構造に着目し、拡張の根拠を説明可能な形でクエリ拡張を行う手法を提案する。入力クエリから代表表現を抽出し、対応するエンティティを起点として、構造的モチーフ条件に基づき意味的に関連するエンティティを拡張語候補として導出する。また、代表表現の選択が拡張結果に与える影響を考慮し、構造的成立性に基づいて代表表現を選択する手法についても検討した。

MS MARCO テストクエリを用いた評価実験の結果、固定代表表現に基づく手法ではほぼ 100%、構造的成立性に基づく手法では約 80%のクエリに対して拡張語候補の抽出が可能であり、本手法が一定の適用可能性を有することが確認された。一方で、BM25 を用いた検索性能評価では、拡張後クエリの $nDCG@10$ は固定代表表現手法で 0.1562、構造的成立性に基づく手法で 0.1609 となり、いずれも元クエリの評価値である 0.2076 を下回る結果となった。これらの結果から、構造的に妥当な拡張語であっても、検索意図を十分に考慮しない場合には検索性能の向上につながらないことが示唆された。

今後の課題として、検索意図を考慮した拡張語の選別や重み付け手法との併用により、構造的拡張の説明可能性を維持しつつ、検索性能の向上を図ることが挙げられる。また、本手法を他の検索タスクへ拡張し、適用範囲を検証することも今後の展望である。

(指導教員 高久 雅生)