

自動検索のためのクエリ生成手法の開発と分析

渡邊 涼太

膨大なデータの中から各ユーザに対して適切な検索結果を提示するため、これまでにクエリ生成や検索結果の向上に関する様々な研究が情報検索で行われてきた。しかし、既存の研究で提案されている情報検索技術はあくまでも主体的に情報検索を行う力のあるユーザのみをサポートするものであった。それを踏まえて、本研究では、二つの研究目的を設定した。一つ目は、「知りたいことが発生する→ユーザがクエリを考える→検索する」という流れのうち、「クエリを考え、検索する」手順を省略することのできる、自動検索システムの開発である。二つ目は、この自動検索システムにおいてどのようなクエリ生成手法が有効であるのかを検証することである。この目的を達成するために、ベースレイヤー・文脈レイヤー・ユーザレイヤーの3レイヤーからそれぞれクエリを生成しその検索結果とともに入力文との適合性を判断する、という手順で研究を行った。文脈レイヤーでは入力文に出現する語それぞれの直前10語を文脈情報として利用した。ユーザレイヤーではユーザが入力文中のある語に興味を持っていると仮定し、その語の文脈情報やその語によって検索された文書のリード文から抽出した語をユーザプロフィールとして利用した。使用した入力文は10種類であり、その全てに対してベースレイヤーで1つ、文脈レイヤーでword2vecとdoc2vecの2つ、ユーザレイヤーでword2vecとdoc2vecそれぞれ3つの手法を用いて実験を行った。

主な結果として、自動でクエリを生成する状況においては、ベースレイヤーと文脈レイヤーのみでは十分でなく、ユーザレイヤーで生成されるクエリとその検索結果が有用である場合が多いということがわかった。その一方で、ベースレイヤーで生成されたクエリでも十分な検索結果が得られる場合もあった。また、単語の分散表現と比較して文書の分散表現を用いた場合では検索結果が十分でないことが多くあった。これは、分散表現を取得するための疑似文書の生成過程で適切に語順を保持できなかったことが理由として考えられる。これらから得られる知見としては、自動検索システムのクエリ生成過程においては、ユーザプロフィールが重要になるということ、そして本研究で採用した手法では文書より単語の分散表現の方が適しているということである。今後の課題としては、本研究では用いることができなかった検索結果のランキングや時間経過を考慮したユーザプロフィールの構築などの手法の試行、文脈窓の幅や語ごとの重みなどのパラメータを変化させることによってクエリや検索結果がどのように変わるのかを調査し、その結果を不特定多数の人間に評価してもらうことでより客観的な結論を導き出すことが挙げられる。

(指導教員 上保秀夫)