

ランキングの部分的な改善を目的とした 能動学習手法

柴田 幸輝

本研究では検索エンジン管理者からのフィードバックに基づいて、ランキングアルゴリズムの部分的な改善を可能にする手法を提案する。ここでいう部分的改善とは「特定のクエリで特定の文書の順位を理想の順位に変える行為」と定義される。例えば人物名の検索結果で犯罪者に関係するサイトの順位を下げるといった変更がこの部分的改善に当てはまる。

提案手法において、検索エンジン管理者は、改善が必要なクエリとそのランキングの改善箇所をフィードバックとして与える。システム内では検索エンジン管理者の意図を推定するモデルを用意する。検索エンジン管理者の意図を推定するモデルはクエリが改善の対象であるかそうでないかを2値分類するクエリ識別モデルとユーザが変更したい文書とその順位を推定するためのユーザランキング関数の2つのモデルから構成される。以下の1~3ステップを繰り返して検索エンジン管理者の意図を推定するモデルを学習させる。1. 検索エンジン管理者の意図を推定するモデル側からフィードバックして欲しいクエリを提示、2. 検索エンジン管理者は提示されたクエリに対してフィードバックを与える、3. 検索エンジン管理者の意図を推定するモデルは与えられたフィードバックを学習する。一連のフィードバックの後、全てのクエリ、文書においてユーザの意図を推定するモデルを適用し、部分的な改善を行う。モデル側からフィードバックして欲しいクエリを提示する際のクエリの選び方としては変更の対象となるクエリをより多く選ぶために、クエリ識別モデルが与えられたクエリが改善の対象に当てはまると予測する確率とフィードバックを与える前のランキングとその時点でのユーザランキング関数が出力するランキングのスピアマンの順位相関係数の差の重み付きの和が最大となるようなクエリを選んだ。

提案手法の有効性を検証するためには、特定のクエリで特定の文書の順位のみが変わったデータセットが必要であるが、それに適したデータセットが入手できなかったため、新たにデータセットを作成した。データセットの作成には日本語の質問検食用データセットである OpenLive-Q を利用し、OpenLive-Q 内のクエリを特定の基準に沿ってグルーピングし、グルーピングされたクエリの結果のみ特定の文書の順位を下げることで特定のクエリで特定の文書の順位が変わった理想の順位を再現した。実験ではこのデータセットを用いて、ベースライン手法と提案手法を比較し、提案手法の有効性を検証した。実験を行った結果、提案するクエリの選び方が多く文書の順位を改善した結果となった。

(指導教員 加藤 誠)