

ドメイン特徴を利用した ランキング学習モデルのドメイン適応

伊藤 拓誠

情報検索分野におけるランキングタスクは、与えられたクエリに対して関連する文書を適合度の高い順に並べ替えることを目的とするタスクである。このようなタスクを行うモデルをランキングモデルと呼び、性能の良いランキングモデルを構築する手法にランキング学習がある。ランキング学習は機械学習を活用した手法で、クエリと文書、それらの適合性判定データからなる学習データを大量に用いてランキングモデルの重みを学習する。ランキング学習では、学習データが不足するドメインのモデルを上手く構築することができず、学習データの中でも適合性判定データは収集にコストがかかり不足しやすいという課題がある。この課題に対して、学習データの乏しいターゲットドメインに学習データの豊富なソースドメインから獲得した知識を転移することで学習データの不足を補う転移学習を応用した研究が行われている。しかしながら、データやモデルの知識をターゲットドメインに適応させるためにターゲットドメインの適合性判定データが少なからず必要であり、ターゲットドメインの適合性判定データがない場合に性能の良いランキングモデルを構築する方法は依然として確立していない。本研究では、ランキング学習でターゲットドメインの適合性判定データが利用できない設定において、ランキング学習モデルのドメイン適応を行う手法を提案する。我々は、ドメイン特徴からそのドメインにおけるランキングモデルの最適な重みを予測する重み予測モデルを構築することで、上記の提案を実現する。この重み予測モデルを利用することで、新たに適合性判定データのないドメインが与えられたとしても、ドメイン特徴からそのドメインに適したランキングモデルの重みを予測し、性能の良いランキングモデルを構築することができる。また、重み予測モデルの学習には複数のドメインが必要となるが、単純にドメインの異なる複数のデータセットを集める方法では数に限界が生じる。そこで本研究では、多種多様なデータを含む1つの大きなデータセットをドメインが異なるように分割して複数のドメインを用意する方法を採用する。実験では、分割元のデータセットにAOLIAを利用し、クエリ・文書ペアの特徴に基づきk-means法で分割して500個のドメインを作成した。その後、それぞれのドメインでドメイン特徴の抽出とランキングモデルの重みの最適化を行い、それらを入出力とする重み予測モデルを学習した。提案手法のランキングタスクにおける評価をした結果、提案手法で構築したモデルはドメインの差異を考慮せずすべてのドメインのデータを利用して学習した汎用的なモデルよりも性能が向上することが明らかになった。また、ドメイン特徴を変更した際の性能の変化を評価した結果、クエリ・文書特徴のみをドメイン特徴として利用した際に提案手法の性能が最も良くなることが判明した。

(指導教員 加藤 誠)