

システム情報工学研究科修士論文概要

年 度	平成 27 年度	学位名	修士(工学)
専 攻	知能機能システム	専攻	著者氏名 陳 磊
指導教員氏名 宇津呂武仁			
論文題目 トピックモデルにおける話題同定の高精度化に関する研究			
論文概要 <p>本論文では、検索エンジン・サジェストを情報源として収集されたウェブページの集合に対して、トピックモデルを適用し、生成されたトピック中に対応付けられたサジェストの頻度を用いて主要話題を含むウェブページを選定することにより、トピックモデルにおける話題同定を高精度に行う手法を提案する。</p> <p>本論文では、まず、検索エンジンを用いて、1 つのクエリ・フォーカスに対して、最大約 1,000 語の検索エンジン・サジェストを収集する。次に、クエリ・フォーカスと収集されたサジェストを用いた AND 検索によって、上位 20 件のウェブページを収集する。収集されたウェブページを集約するために、トピックモデルとして潜在的ディリクレ配分法を用いる。収集されたウェブページの集合に対して、トピックモデルを適用することにより、ウェブページを数十個のトピックに集約する。ウェブページを収集する際、検索において用いたサジェストをウェブページのラベルとみなす。これによって、全 1,000 個のサジェストも、同様に数十個のトピックに集約される。次に、1 つのトピックに対して、トピック中の各サジェストの頻度を求める。1 つのトピックにおいて、1つのウェブページのラベルであるサジェストの頻度の最大値を、そのトピックにおいて当該ウェブページに付与されたサジェストの頻度とみなす。そして、そのトピックにおいて、各ウェブページの頻度に対して下限値を設けて、下限値以上の頻度を持つウェブページを、そのトピックの主要話題を含むウェブページとみなして選定する。下限値未満の頻度を持つウェブページを、そのトピックの副次的な話題を含むウェブページとみなして除外する。各トピックに対して、このような処理を行い、各トピック中の話題を厳選し、各トピックの主要話題を含むウェブページの集合を収集することによって、トピックモデルにおける話題同定の高精度化を実現する。以上の提案手法に対して、人手によって主要話題を含むウェブページを選択して参照用集合を作成し、評価実験を行う。評価対象として、「結婚」および「就活」をクエリ・フォーカスとして用いて評価実験を行なった。その結果、提案手法を用いず、トピックモデルを直接適用した場合と比較して、主要話題を含むウェブページ選定の適合率を改善することができた。以上の結果によって、本論文の手法の有効性を示すことができた。</p>			
審査日 平成 28 年 1 月 27 日			
審査員	(大学名 職名)	(学位)	(氏名)
主査	筑波大学 教授	博士(工学)	宇津呂 武仁
副査	筑波大学 准教授	博士(工学)	矢野 博明
副査	筑波大学 准教授	博士(情報科学)、 博士(デザイン学)	星野 准一