

論文内容要旨

報告番号	甲 先 第 167 号	氏 名	于海涛
学位論文題目	Intent Role Oriented Query Parsing and Its Application in Subtopic Mining (意図役割指向クエリ解析及びサブトピックマイニングへの応用)		
<p>内容要旨</p> <p>Due to users' brevity in expressing their intents, many queries are ambiguous and/or underspecified. Understanding the underlying intent or information need encoded within a query has long been regarded as an important problem for effective information retrieval.</p> <p>In this thesis, I propose two intent roles (kernel-object & modifier) and study how to address this problem by devising the right approach with respect to the search query's type. Specifically, I propose to address this problem from the following perspectives: (i) Depending on whether a query can be represented with kernel-object and modifier, we classify queries as role-explicit ones and role-implicit ones. I propose a high-precision approach that can automatically extract role-explicit queries from a given query log. The experimental results show that the proposed approach achieves a satisfactory performance, more than 79% in terms of the word-level and query-level metrics. This approach enables us to easily obtain sufficient role-explicit queries for later training utilization. (ii) Using a set of predictive features, I build different classifiers to identify whether an arbitrary query is a role-explicit one or a role-implicit one. Moreover, a simplified word n-gram role (SWNR) model is proposed to identify the composing kernel-object and modifier for a given role-explicit query. The experimental results show that: (1) The classifiers can achieve more than 90% precision in identifying role-explicit queries. (2) The SWNR model can achieve a satisfactory performance, more than 73% in terms of the word-level and query-level metrics. (iii) I investigate the value of intent role annotation in the scenario of subtopic mining at a fine-grained word-level.</p> <p>An innovative approach for subtopic mining through modifier graph clustering is proposed. The key idea is that: The modifier graph is decomposable into clusters with strong intra-cluster interaction and relatively weak inter-cluster interaction. Each modifier cluster reasonable indicates a possible subtopic. The Experimental results based on a gold standard topic set show the modifier-graph based approach outperforms the baseline method in terms of I-rec@10, D-nDCG@10 and D#-nDCG@10. Besides the satisfactory results, another advantage of this approach is that: It is robust to the sparseness problem from which many previous methods suffer. Moreover, our success in the INTENT task of NTCIR-10 also demonstrates the potential value of intent role oriented query representation and parsing.</p>			

論文審査の結果の要旨

報告番号	<div style="border: 1px solid black; display: inline-block; padding: 2px;">甲先</div> 乙先 第 167 号 工修	氏 名	于 海 濤
審査委員	主査 寺田 賢治 副査 獅々堀 正幹 副査 任 福継		
学位論文題目 Intent Role Oriented Query Parsing and Its Application in Subtopic Mining (意図役割指向クエリ解析及びサブトピックマイニングへの応用)			
<p>審査結果の要旨</p> <p>本論文では、テキストマイニング及び質問回答分野で極めて重要な質問文における曖昧性の解消問題に着目し、意図役割指向クエリ解析及びサブトピックマイニングへの応用に関する研究を行った。一般に、質問文が短くて、曖昧的なものが含まれるが、効率的なテキストマイニングの実現するには、その質問文の曖昧性を解消することが望ましい。この問題に対して、本研究では、意図役割という概念を考察し、意図役割指向クエリ解析手法を提案し、従来の方法のパフォーマンスを改善することが期待されている。</p> <p>最初に、様々な質問を収集し、特徴調査を行った。特に質問文を役割明示したものと暗黙了解できるものと分類し、カーネルオブジェクトと修飾語で表示される手法を提案した。また、与えられた質問から自動的に役割明示的な質問を抽出するアプローチを提案した。提案したアプローチの性能と有効性に関する実験を行ったが、その有効性を確かめることができた。</p> <p>次に、役割明示されていない暗黙質問について、予め解析した特徴と統計的な情報を活かし、単純化された単語 n-グラム役割(SWNR)モデルを導入した。その結果、90%を超えた精度が上げられた。</p> <p>最後に、提案したアプローチを活かし、サブトピックテキストマイニングへの応用に関する検討を行った。特に、代表的な NTCIR-10 の INTENT タスクに参加し、本論文で提案した方法を実証することができた。</p> <p>提案された手法に基づいて、実験システムを構築し、様々な評価実験を行ったが、提案された手法の有効性を確かめることができた。</p> <p>以上本研究は、当該分野の既存の問題を解決した貢献から価値のある研究であり、本論文は学位論文としての水準を満たし、博士(工学)の学位授与に値するものと判定する。</p>			