

不完全な評価セットに対する検索システムの性能評価指標の開発

○大平 倫宏^{*1)}、富山 真一^{*2)}

■キーワード 検索システム、評価指標、ランキング、機械学習

1. 人手をかけずに検索システムの評価が可能
2. コストが低いため中小企業に適している
3. 新規に検索対象が追加された場合に対応可能

■研究の目的

検索エンジンの性能を評価する際には、検索語と対応する検索結果との関連性について、あらかじめ決めておく必要がある。これを人手で行うのはコストがかかるため、一定数の既知の関連性(不完全な評価セット)だけを用いて評価が行えることが望ましい。本研究では、機械学習等の手法を用いて、不完全な評価セットを基に検索システムの性能を評価するための方法及び指標を開発した。

■研究内容

(1) 機械学習法によって未知の関連性を推定する方法

図1のように、関連性が既知である評価セットを基に、機械学習を利用して、未知の検索対象の関連性を推定する方法の開発を行った。機械学習法としては、サポートベクターマシンを用いた。NTCIR-6 CLIR^[1]の読売新聞、毎日新聞のデータを利用して実験を行った。得られた推定結果から、従来使用されている検索ランキングの評価指標であるnDCG^[2]を用いて、評価セットが完全に評価されている場合のnDCGとの比を求めたのが図2である。

(2) 未知の関連性の影響を考慮した指標

関連性の推定によらずに、関連性が未知の場合の影響を考慮した指標が存在する^[2]。本研究では、従来の方では考慮されていなかった検索意図を考慮した指標を開発した。その性能評価指標に対して、評価セットが完全に評価されている場合の指標値との相対誤差を求めたのが、図3である。対象データとしては、NTCIR-9 INTENT^[1]を利用している。

(3) まとめ

(1)の方法は、図2のように、評価の定まっている文書の数が増えるほど正確となり、適合度が100%に収束してゆく。

(2)の指標については、図3のように、提案法は未知の文書数が増加しても安定しており、検索意図を考慮しない従来の方であるinf-nDCG^[2]と比べて、同程度の結果となっている。



図1. 既知の関連性を利用して未知の関連性を推定

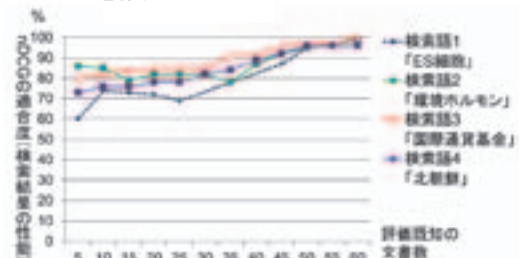


図2. 利用文書数に対する(1)の指標の変化

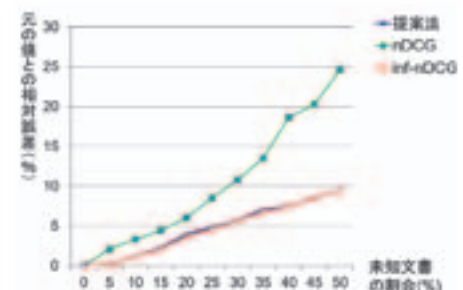


図3. 未知文書数に対する(2)の指標の変化

■研究の新規性・優位性

- ・ 人手をかけずに検索システムを評価可能となり、中小企業でも高性能な検索エンジンの開発が可能となる。
- ・ (2)の指標については、従来の方よりも正確に検索意図を反映することができる。

■産業への展開・提案

- ①商品レコメンデーションエンジンへの応用
- ②医療画像診断システムへの応用
- ③海外特許検索システムへの応用

参考文献

[1] NTCIR, <http://research.nii.ac.jp/ntcir/index-ja.html>

[2] Emine Yilmaz, Evangelos Kanoulas, Javed A. Aslam, Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, July 20-24 (2008)

*1) 生活技術開発セクター、*2) 情報技術グループ