

相互情報量を用いた 質問応答システムのためのクエリ拡張

阿部 裕司[†], 古宮 嘉那子[‡], 小谷 善行[‡]

[†]東京農工大学大学院 情報工学専攻 [‡]東京農工大学大学院 工学研究院

研究背景

質問応答

自然言語の質問に対し、適切な回答を返す

クエリ拡張

文書検索の段階で「質問文には含まれないが回答に含まれやすそうな語」を加えて検索

クエリ拡張語の選択手法

質問応答事例をコーパスとみなし

質問語と回答語の共起情報を利用[1]

事例
 質問: K・S氏が逮捕されたのはなぜですか？
 回答: 証券取引法に違反したからです。

共起の度合い = 相互情報量

$$I(w_q, w_a) = P(w_q, w_a) \log \frac{P(w_q, w_a)}{P(w_q)P(w_a)} + P(w_q, \bar{w}_a) \log \frac{P(w_q, \bar{w}_a)}{P(w_q)P(\bar{w}_a)} + P(\bar{w}_q, w_a) \log \frac{P(\bar{w}_q, w_a)}{P(\bar{w}_q)P(w_a)} + P(\bar{w}_q, \bar{w}_a) \log \frac{P(\bar{w}_q, \bar{w}_a)}{P(\bar{w}_q)P(\bar{w}_a)}$$

既存のクエリ拡張の問題点

オープンドメインの質問応答に対しては
話題のずれが大きくなりすぎる

【質問】ソフトバンクとヤフーの関係は？

【クエリ拡張語】ソフトバンク ⇒ ホークス
ヤフー ⇒ メール

提案手法

質問側の2語に対し共起しやすい

回答側の1語を選出するように計算式を変更

$$I(w_{q1}, w_{q2}, w_a) = P(w_{q1}, w_{q2}, w_a) \log \frac{P(w_{q1}, w_{q2}, w_a)}{P(w_{q1}, w_{q2})P(w_a)} + P(w_{q1}, w_{q2}, \bar{w}_a) \log \frac{P(w_{q1}, w_{q2}, \bar{w}_a)}{P(w_{q1}, w_{q2})P(\bar{w}_a)} + P(\bar{w}_{q1}, w_{q2}, w_a) \log \frac{P(\bar{w}_{q1}, w_{q2}, w_a)}{P(\bar{w}_{q1}, w_{q2})P(w_a)} + P(\bar{w}_{q1}, w_{q2}, \bar{w}_a) \log \frac{P(\bar{w}_{q1}, w_{q2}, \bar{w}_a)}{P(\bar{w}_{q1}, w_{q2})P(\bar{w}_a)}$$

【質問】ソフトバンクとヤフーの関係は？

【クエリ拡張語】ソフトバンク, ヤフー ⇒ 子会社

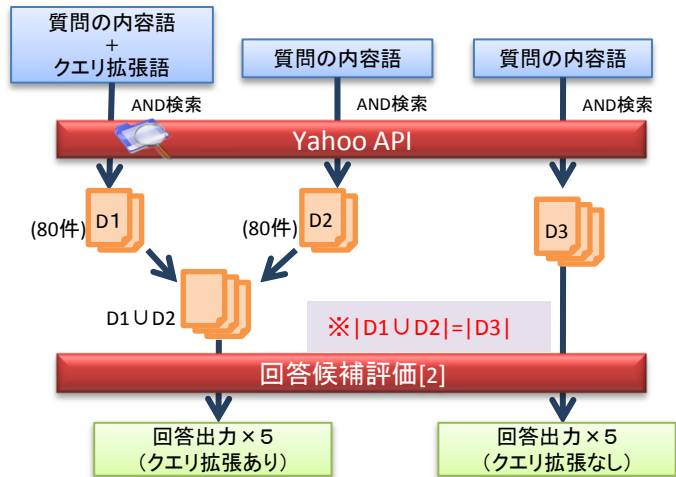
評価実験

既存の質問応答システム[2]に提案手法を導入し
NTCIR8-ACLIA2[3]の100質問に回答

クエリ拡張を利用して検索した文書集合と
利用していない文書集合を

同じ回答候補評価モジュールで評価

質問応答事例としてYahoo!知恵袋データを利用



回答候補評価

$$Score(\text{候補文}) = \frac{1}{\ln(1 + \text{文字数}(\text{候補文}))} \cdot \left\{ \begin{array}{l} \text{「内容の関連度」スコア} \\ \text{「記述の回答らしさ」スコア} \end{array} \right\}^{0.5}$$

クエリ拡張	なし	既存[1](1:1)	提案(2:1)
入力質問数	100	100	100
正解率	0.42	0.40	0.45
MRR	0.262	0.233	0.273

まとめ

質問側の2語に対する回答側の1語の共起情報を利用したクエリ拡張手法を提案

提案手法の導入により質問応答システムの回答性能が向上

参考文献

- [1] Adam Berger, Rich Caruana, David Cohn, Dayne Freitag, Vibhu Mittal. "Bridging the Lexical Chasm: Statistical Approaches to Answer-Finding". In Proceedings of SIGIR '00 Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval (2000).
 [2] 石下 円香, 佐藤 充, 森 辰則. "Web文書を対象とした質問の型に依らない質問応答手法". 人工知能学会論文誌 24(4), 339-350(2009).
 [3] Teruko Mitamura, Hideki Shima Tetsuya Sakai, Noriko Kando, Tatsunori Mori, Koichi Takeda, Chin-Ywe Lin, Ruihua Song, Chuan-Jie Lin and Cheng-Wei Lee: Overview of the NTCIR-8 ACLIA Tasks: Advanced Cross-Lingual Information Access, In Proceedings of 8th NTCIR Workshop Meeting(2010).