

統計情報を用いた人の心を捉える質問応答文の分析

谷岡 広樹* 中谷 良平† 内田 佳孝†

* 徳島大学 情報センター

† 株式会社ワークスアプリケーションズ

tanioka.hiroki@tokushima-u.ac.jp,
{nakatani_r, uchida_yo}@worksap.co.jp

1 はじめに

質問応答システムは、質問者の質問に対して適切な回答を応答することが求められる。しかしながら、実際のサービスで利用することを想定した質問応答システムにおいては、応答不可能な質問に対する適切に対応することが求められる [7] ことや、過去の質問応答の中から質問意図や興味関心を持つであろう質問応答文を提示する [1] ことが求められる。NTCIR-14 Open Live Test for Question Retrieval (OpenLiveQ-2) タスクでは、システムが返した順位付きの質問応答文の集合に対して、オフラインテストおよびオンラインテストによって評価する。本稿では、Yahoo! Japan から提供された質問応答文の統計情報と文字情報を用いて、人の心を捉える質問応答文の特徴について検討、評価を行ったので報告する。

2 関連研究と提案手法

昨年度のタスク全体の報告書 [6, 2] によると、第1フェーズで実施されるオフラインテストと第2フェーズで行われるオンラインテストでは、性能評価の結果が逆転する場面があったことがわかる。また、OKSAT [3] の報告によると、第1フェーズでは、提示された訓練データを用いて質問応答文に順位づけし、フィードバックされた評価指標を元に分析する方法が用いられている。本システム AITOK は、質問応答文がどの程度人の心を捉えるかは、トレーニング用データとそのクリックスルーデータから判断できると仮定し、統計情報 (表 1) や文字情報 (表 2) を用いて、人の心を捉えるスコア (Catchy スコア) を定義する。次に、第1フェーズでフィードバックされる評価指標 (Q-measure) の増減に基づいて精度の向上を図り、第2フェーズで実際のサービスに対する効果をみる。

表 1: 統計情報のフィールド一覧

Field Name	Description
Title	Title of the question
Snippet	Snippet of the question in a search result
Body	Body of the question

表 2: 文字情報のフィールド一覧

Field Name	Description
Page view	Page view of the question
Number of answers	Number of answers for the question
Update	Last update time of the question
Clickthrough rate	Clickthrough rate

3 アルゴリズム

Catchy スコアの計算には、文字情報からは CaboCha [5] で形態素と係り受け、word2vec [4] で分散表現を求め、統計情報からは閲覧頻度 (Page view)、回答数 (Number of answers)、最終更新日時 (Update)、クリック率 (Clickthrough rate) を利用する。

3.1 Catchy スコア

本タスクでは、クエリごとにクエリ文字列が与えられている。クエリ文字列および文字情報の各フィールドの文字列を形態素解析し、分かち書きされた形態素を 1-gram または 2-gram の Bag-of-Words として用いる。また、係り受けのある形態素を連結することでフレーズを生成する。さらに、形態素の 200 次元の分散表現ベクトルと比較する。

Chatchy スコアには、統計情報のフィールドの数値の中でも、クリック率と閲覧頻度の高い質問応答文に含まれる形態素やフレーズを多く含む質問応答文のスコアが大きくなるようにした。これに、回答数、最終更新日時なども組み合わせることで、第1フェーズの評価指標がどのように変化するかをみる戦略である。さらに、第2フェーズの結果と第1フェーズの結果を比較し、その評価方法の評価結果への影響について考察する。

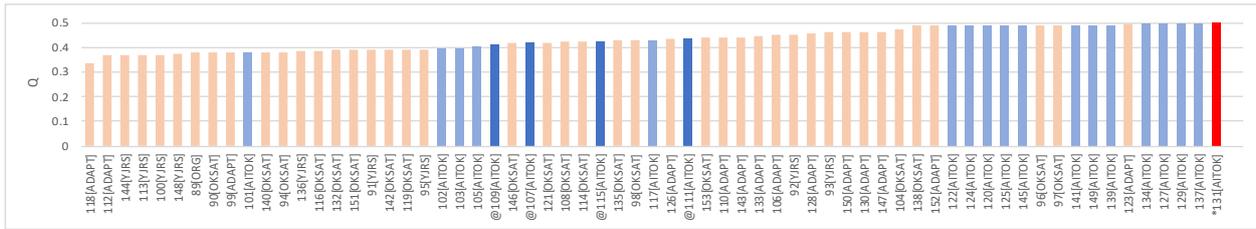


図 1: オフラインテストの結果 (Q-measure)

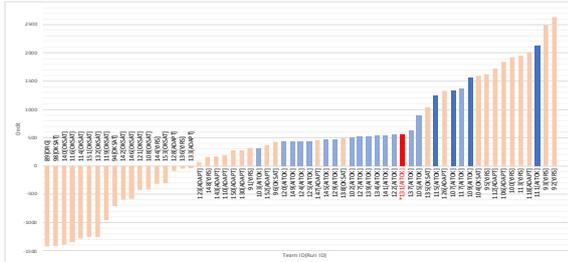


図 2: Top 60 を比較したオンラインテスト

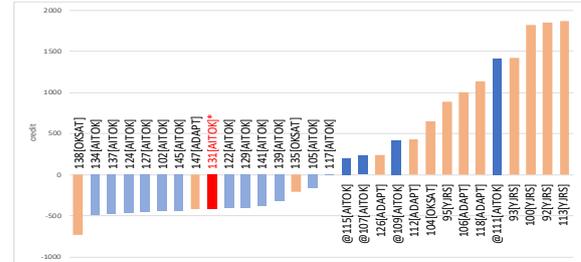


図 3: Top 30 を比較したオンラインテスト

4 実験結果

図 1 はオフラインテストのうち Q-measure の結果を示す。この中で、Run ID:131 が上位となった。その他の精度指標には、nDCG@5, nDCG@10, nDCG@20, nDCG@50, ERR@5, ERR@10, ERR@20, ERR@50 があるが、すべての指標において上位となった。Run ID:131 は、閲覧数と回答数を重視した。一方、オンラインテストの結果を図 2 および 図 3 で示す。この結果では、オフラインテストで上位であった Run ID:131 は Run ID:111 と順位が逆転された。Run ID:111 は、最終更新時間を加味している。

5 おわりに

本稿では、NTCIR-14 OpenLive-Q タスクでの分析結果を、速報性を重視してその概要と結果のみをまとめた。オフラインテストおよびオンラインテストでの評価結果、オフラインテストでは、主に閲覧数に基づいて順位づけすることで高い精度が得られることがわかった。一方、オンラインテストでは、最終更新時間を考慮することが精度向上の要因であることが伺える。この違いは、テスト環境およびテスト方法に原因があるものと考えられる。このことについては、今後さらなる検証が必要であろう。

参考文献

- [1] Makoto P. Kato, Tomohiro Manabe, Sumio Fujita, Akiomi Nishida, and Takehiro Yamamoto. Challenges of Multileaved Comparison in Practice: Lessons from NTCIR-13 OpenLiveQ Task. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM '18*, pp. 1515–1518, New York, NY, USA, 2018. ACM.
- [2] Makoto P. Kato, Takehiro Yamamoto, Tomohiro Manabe, Akiomi Nishida, and Sumio Fujita. Overview of the NTCIR-13 OpenLiveQ Task. In *NTCIR-13 Conference*, 2017.
- [3] Takashi Sato. OKSAT at NTCIR-13 OpenLiveQ Task. In *NTCIR-13 Conference*, 2017.
- [4] word2vec : Tool for computing continuous distributed representations of words. <https://code.google.com/p/word2vec>. [Online; accessed 11-September-2018].
- [5] CaboCha/南瓜:Yet Another Japanese Dependency Structure Analyzer. <http://www.chasen.org/~EIJtaku/software/cabocha/>. [Online; accessed 16-September-2018].
- [6] 加藤誠, 山本岳洋, 真鍋知博, 藤田澄男. コミュニティ Q&A サイトにおける質問検索システムの大規模オンライン評価. In *DEIM, G2-2*, 2018.
- [7] 谷岡広樹, 中谷良平, 与儀涼子, 内田佳孝. 質問応答システムの基盤化と応答可能性の判定. 言語処理学会 第 24 回年次大会. 言語処理学会, 2018.