Best-Answer Selection Using a Machine Learning Tool at NTCIR-8 CQA Pilot Task

Kazuko Kuriyama
Shirayuri College
1-25 Midorigaoka, Chofu-shi
Tokyo 182-8525, Japan
kuriyama@shirayuri.ac.jp

ABSTRACT

In this paper, I report the evaluation results of my best-answer selection system at NTCIR-8 CQA Pilot Task. My goal was to assess what features were useful to select best-answers using a machine learning method. I submitted three runs in the main task. They were lists of best-answer candidates which were selected by a machine learning tool using different sets of features and sorted in order of scores. I used readability of questions and answerers' attributes as features in machine learning as well as basic features, such as length of questions.

Keywords

NTCIR, best-answer selection, SVM^{light}, machine learning

1. INTRODUCTION

Question Community Question Answering (CQA) sites have been growing as one of promising services on the web. The participants of CQA sites post questions and answers to share information with other participants. Questioning and answering are voluntary without monetary rewards except for comments from questioners to answerers and some kinds of evaluations given by participants' voting.

Sets of questions and their answers in CQA sites are thought as databases of knowledge. They are a valuable resource for the people who search for information. However, qualities of them varied widely and there are problems for evaluating and extracting useful knowledge from the thread and thrum.

Many reseachers have attempted to predict users' satisfaction and quality of answers and to find what features and aspects have close relations with best-answers selection by using clustering, machine learning and other methods[1, 2, 6, 3].

I also investigated best-answer selection criteria and some features to used for estimating answers through analysis of question types and readability of questions and answers[7, 8, 9].

In this paper, I focus on some features to be used in machine learning and describe evaluation results. This is the first time that NTCIR workshop proposed a task for CQA research. The organizers of the NTCIR-8 CQA pilot task described the details of the task in the overviews[4, 10], so I leave them out of this paper.

2. SUBMITTED RUNS

I used a machine learning tool SVM^{light}[5] to score best-answer candidates. SVM^{light} is an implementation of Vapnik's Support Vector Machine[12].

The features used in machine learning were (1)readability of answer, (2)length of answer, (3)number of answers posted the answerer, (4)number of best-answers the answerer posted, (5)ratio of best-answers to answers the answerer posted, (6)ratio of answers to questions the answerer posted, (7)readability of question, (8)length of question, (9)number of questions posted the questioner, (10)ratio of questions to answers the questioner posted. The readability scores were computed by a readability analyzer of Japanese texts *Obi2*[11]. Training data was 500 questions and their answers with three different sets of features, which were the set A: {(1),(3),(4),(5)}, the set B: {(1),(2),(3),(4),(5),(6)}, and the set C: all features. I submitted three runs for the main task, LILY-1, LILY-2, LILY-3, which were ranked using the sets A, B and C respectively.

Table 1 shows the evaluation results of my runs using *cqa_eval* and *ir4qa_eval2* NTCIR distributes. The table includes the organizers' run BASELINE-1 which ranks answers at random to compare with mine. All three runs underperformed BASELINE-1.

3. CONCLUSIONS

I participated in NTCIR-8 CQA task by taking a machine learning approach to best-answer selection. Poor performance was observed using features I selected. I would like to proceed with failure analysis and carry out an additional experiment using other features.

4. ACKNOWLEDGMENTS

'Yahoo! Chiebukuro Data' provided to National Institute of Informatics by Yahoo Japan Corporation was used in the implementation of this research.

5. REFERENCES

[1] E. Agichtein et al. Finding high-quality content in social media. In *Procddeings of WSDM '08*, pages 183–194. ACM, 2008.

[2] J. Bian et al. Finding the right facts in the crowd: Factoid question answering over social media. In *Proceedings of WWW '08*, pages 467–476. ACM, 2008.

[3] Y. Cao, H. Duan, C.-Y. Lin, Y. Yu, and H.-W. Hon. Recommending questions using the mdl-based tree cut model. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, pages 81–90. ACM, 2008.

Table 1: Evaluation Results

RunID	BA-Hit@1	GA-Hit@1	GA-nG@1	GA-nDCG	GA-Q	UFA-Hit@1	UFBA-Hit@1
LILY-1	0.1767	0.9887	0.6883	0.9096	0.8927	0.7733	0.7847
LILY-2	0.1767	0.9887	0.6883	0.9191	0.9081	0.7733	0.7847
LILY-3	0.1767	0.9887	0.6883	0.9142	0.9002	0.7733	0.7847
BASELINE-1	0.2713	0.9920	0.7751	0.9311	0.9169	0.8533	0.8607

[4] D. Ishikawa, T. Sakai, and N. Kando. Overview of the ntcir-8 community qa pilot task (part i): The test collection and the task. *Proceedings of the 8th NTCIR Workshop*, 2010.

[5] T. Joachims. *Learning to Classify Text Using Support Vector Machines*. Kluwer, 2002.

[6] P. Jurczyk and E. Agichtein. Discovering authorities in question answer communities by using link analysis. In *Proceedings of CIKM '07*, pages 919–922. ACM, 2007.

[7] K. Kuriyama and N. Kando. Analysis of questions and answers in q&a site. *IPSJ SIG Technical Report*, 2009-FI-95(19):1–8, 2009. (In Japanese).

[8] K. Kuriyama and N. Kando. Analysis of questions and answers in q&a site (2): Based on document structures and attributes. *IPSJ SIG Technical Report*, 2009-FI-96(3):1–8, 2009. (In Japanese).

[9] K. Kuriyama and N. Kando. Analysis of questions and answers in q&a site (3): Predicting best-answers based on users' histories. *IPSJ SIG Technical Report*, 2010-FI-97(7):1–8, 2010. (In Japanese).

[10] T. Sakai, D. Ishikawa, and N. Kando. Overview of the ntcir-8 community qa pilot task (part ii): System evaluation. *Proceedings of the 8th NTCIR Workshop*, 2010.

[11] S. Sato, S. Matsuyoshi, and Y. Kondoh. Automatic assessment of japanese text readability based on a textbook corpus. *Proceedings of LREC-08*, 2008.

[12] V. N. Vapnik. *The Nature of Statistical Learning Theory*. Springer, 1995.