Can Readability Enhance Recommendations on Community Question Answering Sites?

Oghenemaro Anuyah, Ion Madrazo Azpiazu, David McNeill, Maria Soledad Pera
People and Information Research Team
Department of Computer Science, Boise State University
Boise, Idaho, USA
{oghenemaroanuyah,ionmadrazo,davidmcneill,solepera}@boisestate.edu

ABSTRACT

We present an initial examination on the impact text complexity has when incorporated into the recommendation process in community question answering sites. We use Read2Vec, a readability assessment tool designed to measure the readability level of short documents, to inform a traditional content-based recommendation strategy. The results highlight the benefits of incorporating readability information in this process.

CCS CONCEPTS

• **Information systems** → *Social recommendation*; *Question answering*;

KEYWORDS

Community question answering, Readability, Recommender

1 INTRODUCTION

Community question answering (CQA) sites allow users to submit questions on various domains so that they can be answered by the community. Sites like Yahoo! Answers, StackExchange, or StackOverflow, are becoming increasingly popular, with thousands of new questions posted daily. One of the main concerns of such sites, however, is the amount of time a user has to wait before his question is answered. For this reason, CQA sites depend upon knowledge already available and refer users to older answers, i.e., answers provided for previously-posted questions and archived on the site, so that users can get a more immediate response to their inquiries. This recommendation process has been extensively studied by researchers using a wide range of content similarity measures that go from the basic bag-of-words model to semantically related models, such as ranksLDA [6].

We argue that the recommendation process within CQA sites need to go beyond content matching and answer-feature analysis and consider that not every user has similar capabilities, in terms of both reading skills and domain expertise. User's reading skills can be measured by readability, which refers to the ease with which a reader can comprehend a given text. This information has been applied in the past with great success for informing tasks such as K-12 book recommendation [5], Twitter hashtag recommendation [1] and review rating prediction [3]. Yet, it has not made its way to CQA recommendations, where we hypothesize it can have a significant impact, given that whether the user understands the answer provided by a recommender can highly condition the value the user gives to the answer.

In this paper, we present an initial analysis that explores the influence of incorporating reading level information into the CQA recommendation process. With this objective in mind, we consider the *answer recommendation* task, where a user generates a query that needs to be matched with an existing question and its corresponding answer. We address this task by ranking question-answer pairs and selecting the top-ranked pair to recommend to the user. For doing so, we build upon a basic content-based recommendation strategy which we enhance using readability estimations. Using a recent Yahoo! Question-Answering dataset, we measure the performance of the basic recommender and the one informed by text complexity and demonstrate that readability has indeed an impact on user satisfaction.

2 READABILITY-BASED RECOMMENDATION

We describe below the strategy we use for conducting our analysis. Given a query q, generated by a user U, we locate each candidate answer C_a—along with the question Q_a associated with C_a—that potentially addresses the needs of U expressed in q. Thereafter, the highest-ranked C_a-Q_a pair is recommended to U.

2.1 Examining Content

To perform content matching, we use an existing WordNet-based semantic similarity algorithm described by Li et al. in [4]. We use this strategy for computing the degree of similarity between q and C_a, denoted $Sim(q, C_\text{a})$, and also the similarity between q and Q_a, denoted as $Sim(q, Q_a)$. We depend upon these similarity scores for ensuring that the recommended C_a-Q_a pair matches U's intent expressed in q. We use a semantic strategy, as opposed to the well known bag-of-words, to better capture sentence resemblance when sentences include similar, yet not exact-matching words, e.g. ice cream and frozen yogurt.

2.2 Estimating Text Complexity

To estimate the reading level of C_a and U (the latter inferred indirectly through q), we first considered traditional readability formulas, such as Flesch Kincaid [2]. However, we observed that these formulas were better suited for scoring long texts. Consequently, we instead use **Read2Vec**, which is a deep neural network-based readability model tailored to estimate complexity of short texts. The deep neural network is composed of two fully connected layers and a recurrent layer. Read2Vec was trained using documents from Wikipedia and Simple Wikipedia, and obtained a statistically significant improvement (72% for Flesch vs. 81% for Read2Vec) when

RecSys 2017 Poster Proceedings, August 27–31, Como, Italy.

predicting the readability level of short texts, compared to traditional formulas including Flesch, SMOG and Dale-Chall [2].

Given that the answer to be recommended to U should match U's reading ability to ensure comprehension, we compute the Euclidean distance between the corresponding estimations, using Equation 1.

$$\mathrm{d}(q, C_a) = \left|R2V(q) - R2V(C_\mathrm{a})\right| \tag{1}$$

where $R2V(q)$ and $R2V(C_\mathrm{a})$ are the readability level of q and C_a, respectively, estimated using Read2Vec.

2.3 Integrating Text Complexity with Content

We use a linear regression model[1] for combining the scores computed for each C_a–Q_a pair. This yields a score, $Rel(C_a, Q_a)$, which we use for ranking purposes i.e., the pair with the highest score is the one recommended to U.

$$Rel(C_a, Q_a) = \beta_0 + \beta_1 Sim(q, C_a) + \beta_2 Sim(q, Q_a) + \beta_3 d(q, C_a) \tag{2}$$

where β_0 is the bias weight, and β_1, β_2 and β_3 are the weights that capture the importance of the data points defined in Sections 2.1 and 2.2. This model was trained using least squares for optimization.

3 INITIAL ANALYSIS

For analysis purposes, we use the L16 Yahoo! Answers Query to Questions **dataset** [7], which consists of 438 unique queries. Each query is associated with related question-answer pairs, as well as a user rating that reflects query-answer satisfaction on a [1-3] range, where 1 indicates "highly satisfied", i.e., the answer addresses the information needs of the corresponding query. This yields 1,571 instances, 15% of which we use for training purposes, and the remaining 1,326 instances we use for testing.

In addition to our *Similarity+Readability* recommendation strategy (presented in Section 2), we consider two **baselines**: *Random*, which recommends question-answer pairs for each test query in an arbitrary manner; and *Similarity*, which recommends question-answer pairs for each test query based on the content similarity between the answer and the query, computed as in Section 2.1.

An initial experiment revealed that regardless of the **metric**, i.e., Mean Reciprocal Rank (MRR) and Normalized Discounted Cumulative Gain (NDCG), the strategies exhibit similar behavior, thus we report our results using MRR.

As shown in Figure 1, recommendations generated using the semantic similarity strategy discussed in Section 2.1 yield a higher MRR than the one computed for the random strategy. This is anticipated, as *Similarity* explicitly captures the query-question and query-answer closeness. More importantly, as depicted in Figure 1, integrating readability with a content-based approach for suggesting question-answer pairs in the CQA domain is effective, in terms of enhancing the overall recommendation process[2]. In fact, as per its reported MRR, *Similarity+Readability* positions suitable question-answer pairs high in the recommendation list, which is a non-trivial task, given that for the majority of the test queries (i.e., 83 %), there are between 5 and 23 candidate question-answer pairs.

[1]We empirically verified that among well-known learning models, the one based on linear regression was the best suited to our task. We attribute this to its simplicity, which can better generalize over few training instances than most sophisticated models.
[2]The weights learned by the model: $\{\beta_0, \beta_1, \beta_2, \beta_3\} = \{2.26, 0.58, 0.20, 0.12\}$.

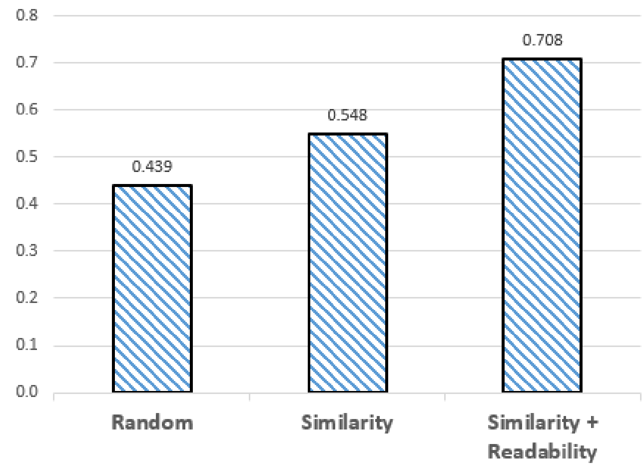

Figure 1: Performance assessment based MRR using the Yahoo! Answers Query to Questions.

4 CONCLUSIONS AND FUTURE WORK

In this study, we analyzed the importance of incorporating readability level information into the recommendation process when it comes to the community based question answering domain. We treat the reading level as a personalization value and compare the readability level on an answer with respect to the reading abilities of a user, inferred through his query. We demonstrated that reading level can be an influential factor in terms of deciding the answer quality and can be used to improve user satisfaction in a recommendation process.

In the future, we plan to conduct a deeper study using other community question answering sites such as Quora or StackExchange. We also plan to analyze queries for additional factors, such as relative content-area expertise, to better predict a user's familiarity with content-specific vocabulary used on archived answers to be recommended. We suspect that readability and domain-knowledge expertise will be highly influential when the recommendation occurs on CQA sites like StackExchange, given the educational orientation of questions posted on the site.

ACKNOWLEDGMENTS

This work has been partially funded by NSF Award 1565937.

REFERENCES

[1] I. M. Azpiazu and M. S. Pera. Is readability a valuable signal for hashtag recommendations? 2016.
[2] R. G. Benjamin. Reconstructing readability: Recent developments and recommendations in the analysis of text difficulty. *Educational Psychology Review*, 24(1):63–88, 2012.
[3] B. Fang, Q. Ye, D. Kucukusta, and R. Law. Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tourism Management*, 52:498–506, 2016.
[4] Y. Li, D. McLean, Z. A. Bandar, J. D. O'shea, and K. Crockett. Sentence similarity based on semantic nets and corpus statistics. *IEEE TKDE*, 18(8):1138–1150, 2006.
[5] M. S. Pera and Y.-K. Ng. Automating readers' advisory to make book recommendations for k-12 readers. In *ACM RecSyS*, pages 9–16, 2014.
[6] J. San Pedro and A. Karatzoglou. Question recommendation for collaborative question answering systems with rankslda. In *ACM RecSys*, pages 193–200, 2014.
[7] Webscope. L16 yahoo! answers dataset . yahoo! answers, 2016. [Online; accessed 17-June-2017].