Competing Topic Naming Conventions in Quora: Predicting Appropriate Topic Merges and Winning Topics from Millions of Topic Pairs

Binny Mathew
binnymathew@iitkgp.ac.in
Indian Institute of Technology, Kharagpur

Suman Kalyan Maity
suman.maity@kellogg.northwestern.edu
Kellogg School of Management, Northwestern University

Pawan Goyal
pawang@cse.iitkgp.ac.in
Indian Institute of Technology, Kharagpur

Animesh Mukherjee
animeshm@cse.iitkgp.ac.in
Indian Institute of Technology, Kharagpur

ABSTRACT

Quora is a popular Q&A site which provides users with the ability to tag questions with multiple relevant topics which helps to attract quality answers. These topics are not predefined but user-defined conventions and it is not so rare to have multiple such conventions present in the Quora ecosystem describing exactly the same concept. In almost all such cases, users (or Quora moderators) manually merge the topic pair into one of the either topics, thus selecting one of the competing conventions. An important application for the site therefore is to identify such competing conventions early enough that should merge in future. In this paper, we propose a *two-step approach that uniquely combines the anomaly detection and the supervised classification frameworks* to predict whether two topics from among millions of topic pairs are indeed competing conventions, and should merge, achieving an F-score of 0.711. We also develop a model to predict the direction of the topic merge, i.e., the winning convention, achieving an F-score of 0.898. Our system is also able to predict ~ 25% of the correct case of merges within the first month of the merge and ~ 40% of the cases within a year. This is an encouraging result since Quora users on average take 936 days to identify such a correct merge.

CCS CONCEPTS

• **Networks** → *Social media networks*; *Online social networks*; • **Human-centered computing** → *Collaborative and social computing*; *Social content sharing*.

KEYWORDS

Quora, Topic Merge, Classification, Online Q&A , Human Evaluation

ACM Reference Format:

Binny Mathew, Suman Kalyan Maity, Pawan Goyal, and Animesh Mukherjee. 2020. Competing Topic Naming Conventions in Quora: Predicting Appropriate Topic Merges and Winning Topics from Millions of Topic Pairs. In *7th ACM IKDD CoDS and 25th COMAD (CoDS COMAD 2020), January 5–7, 2020, Hyderabad, India.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3371158.3371173

CoDS COMAD 2020, January 5–7, 2020, Hyderabad, India

ACM ISBN 978-1-4503-7738-6/20/01…$15.00
https://doi.org/10.1145/3371158.3371173

1 INTRODUCTION

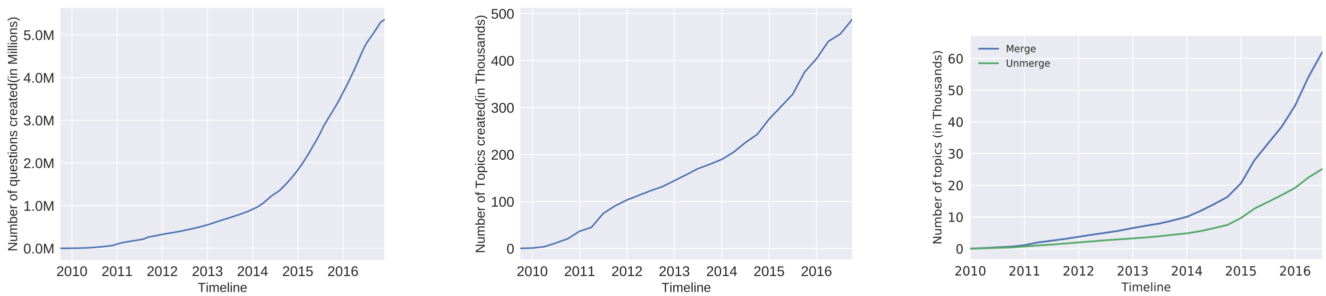
Conventions form an important part of our daily life. They affect our basic perceptions [5] as well as guide everyday behaviors of the general public [3]. With the emergence of social media, conventions started being used in online space. Twitter, for example, follows conventions like "#" used for specifying keywords, "@" for specifying mentions, and “RT” which is an acronym for retweet. None of these were defined by Twitter, but gradually became conventions [23] due to their usability and wide acceptance among the users. Similar competing conventions has been observed in other places as well [14, 45]. In this paper, we investigate the competition among the topic naming conventions prevalent in Quora.

Quora is a popular Q&A site with ~100 million[1] monthly unique visitors. As of April 2017, 13 million questions [48] have been posted on Quora. We have obtained ~ 5.4 million questions which is a significant proportion of all the Quora questions and to the best of our knowledge, is the largest Quora question base. Figure 1a shows the cumulative growth of the number of questions posted on Quora. Each of these questions are assigned topics that allows to categorize these questions. These topics are defined by the user according to his/her convenience. Figure 1b shows the cumulative growth of the number of topics being created on Quora.

Since topics on Quora are user-defined conventions to express certain concepts, topic merging on Quora can be considered as a competition among conventions in which the topic names which are merged can be considered to be conventions and the competition between these conventions determines the winner topic name which can better represent the underlying concept. These conventions might be created at different time periods and a newly created topic name may represent the concept better than the existing topic name. In such a case, the users can replace the old convention with the newly formed convention. We find that in case of Quora, 33% of the times an existing convention gets replaced by a new convention.

The topics on Quora allows the users to group questions of similar interest. If multiple topics are allowed to represent the same concept, it would degrade the quality of the questions and possibly

[1]https://www.quora.com/How-many-people-use-Quora-7/answer/Adam-DAngelo



(a) Cumulative growth of questions posted on Quora

(b) Cumulative growth of topics created on Quora

(c) The Cumulative growth of Topic Merge and Topic Unmerge in Quora

Figure 1: Question and topic growth on Quora.

hamper correct routing of the questions to the correct experts who actually are able to answer the questions. In our dataset, we observe that a Quora user takes on average 936 days to merge duplicate topics from its date of creation. This means that whatever question is posted with the new duplicate topic name, it would most likely miss the actual topic audience. This in turn reduces the quality of the answers received to these questions. In this paper, we propose to characterize the competing topic naming conventions on Quora, and develop various features that can effectively indicate, at an early stage, if a pair of topics that are competing, would merge in future. This would allow the community to merge topics at an early stage of the topic evolution, in turn, promoting early and appropriate knowledge aggregation. Using our model, we are able to predict around 25% of the merge cases in the first month of the topic creation itself. This would allow the duplicate topics to be merged at an early stage itself, thereby, increasing the quality of the answers received.

1.1 Research objectives and contributions

In this paper, we study the phenomena of topic merging in Quora as competing conventions, and investigate various factors influencing this phenomena. Toward this objective, we make the following contributions:

- We obtain a massive dataset and investigate the detailed characteristics of the merging topics. We also introduce the Quora topic ontology - a well defined hierarchy of topics in Quora with each topic linked to its parent topic and child topics.
- We propose a two-step approach which *uniquely combines the anomaly detection with the supervised classification framework* to automatically predict whether a topic pair is competing and would get merged in future, achieving an F-score of $\sim$ 0.711 for predicting true merges from a million test instances. We propose another classification model to predict the winner of the competition and achieve an F-score of 0.898.
- As a direct application, we also launch an early prediction scheme, whereby, we show that we are able to predict $\sim$ 25% of the correct case of merges within the first month of the merge and $\sim$ 40% of the cases within a year.
- We further perform an experiment with human subjects to identify how well humans can predict whether two topics should merge or not, as well as, the direction of topic merge.

Our study can be summarized by two very unique and interesting observations about the competing topic naming conventions. First, the content of the questions tied to the competing topics and the distance of the topics themselves in the Quora topic ontology together serve as strong indicators of whether the topic pairs are competing and would merge in future. Second, the winning topic among the competing pair is best determined by factors like number of characters/words in the topic name, date of creation of the topic and the number of questions/answers tied to the topic.

2 RELATED WORKS

Social Q&A sites allow users to post questions and get quality answers from the community. Sites like Yahoo! Answers, Baidu Knows, Stack Overflow and Quora have been a subject of active research [2, 7, 27, 29–31, 34, 39–41, 46, 48, 51, 53]. Patil et al. [40] study experts and non- experts in Quora and develop statistical models to automatically detect experts. Maity et al. [31] analyze the dynamics of topical growth over time and propose a regression model to predict the popularity of the topics. The studies done in [2, 27, 39] focus on ranking the users from expertise measures based on users' history and activities.

2.1 Conventions

Several studies have been conducted to study the emergence of conventions. One group of such studies is based on small scale experimental work. In Wilkes-Gibbs and Clark [49] for instance, the authors asked the participants to develop short-hand verbals which allowed them to communicate more efficiently and complete the tasks quickly.

Social conventions and their emergence is a fairly old area of study [15, 16]. According to Edna [16], social conventions correspond to special type of norms related to coordination problems, that is, conventions are "those regularities of behavior which owe either their origin or their durability to their being solutions to recurrent (or continuous) co-ordination problems, and which, with time, turn normative". One of the interesting ways to study the dynamics of conventions using real data is through social media [14, 17, 22, 23, 52]. In Kooti et al. [23] the authors study the emergence of retweet convention in Twitter. They perform in-depth study of the conventions and provide valuable insights. Further, in Kooti et al. In Rotabi et al. [45], the authors study conventions that vary even as the underlying meaning remains constant by tracking the spread of macros and other author-defined conventions in the e-print arXiv over 24 years. They found that the interaction among co-authors over time plays a crucial role in the selection of conventions.

2.2 Blending and Compounding

Lexical blending is a linguistic phenomenon in which a word is formed by fusing two or more words into one another (e.g., biopic = biography + picture, pulsar = pulse + quasar). This linguistic form of word reduction has been studied widely [8, 9, 11–13, 19, 26, 35, 43].

Another closely related phenomena is lexical compounding. Few studies have been conducted on lexical compounding in English [6, 20] and other languages like Italian, French, German, Spanish, Chinese etc. [4, 18, 25, 38, 47]. Hashtag compounding is a linguistic process in which two separate hashtags coalesce to form a single hashtag. Mitra et al. [37] gave an approach to identify word sense changes in text media across timescales. In this work, they talk about word sense merging, in which two senses of the same word merge into a single prevalent sense. Hashtag compounding is a linguistic process in which two separate hashtags coalesce to form a single hashtag. Maity et al. [32] studied various socio-linguistic properties responsible for hashtag compound formation and proposed a model that can identify if a compounded hashtag will become more popular than its individual constituent hashtags.

2.3 The present work

The present work significantly differs from the above lines of research. As outlined in the introduction we model the problem of topic merging as a competition of conventions motivated by a few previous studies [23, 45]. The major difference of our work as compared to the work done in the conventions is in using the topic hierarchy information for the competition of conventions. In case of blending and compounding, the two words join to form a new word, but in our problem, one topic is actually merging into another. So no new topic is generated as opposed to blending and compounding. For the first time, we show how the competition of conventions is driven by the content of the questions corresponding to the two candidate topics and their distance in the *topic ontology*. As an added novelty, we also develop a two-step approach to predict merges as well as an approach to predict the directions of merge.

3 DATASET DESCRIPTION

We obtain the Quora dataset from the authors of [31] and then built up on it to have a massive set of 5.4 million questions and 488,122 topics (along with the topic logs). In addition to the above dataset, we have also used the Quora topic ontology.

Quora topic ontology: Quora has an extensive set of topics that is constantly growing and expanding. To organize the huge set of topics in Quora, the site managers use parent-child relationships between topics. This topic hierarchy forms a Directed Acyclic Graph[2]. Each topic can have multiple parents as well as multiple child topics. The root topic of the topic hierarchy graph is "Major-Topics" and we can navigate the topic graph down from this topic.

We extract the topics' parent-child information from the topic logs. Using the parent-child relation, we construct the Quora topic ontology graph.

3.1 Filters applied

We use the topic log of 488,122 topics to extract the merge and unmerge information. We find 65,231 cases of topic merges and 38,502 topic unmerges.

Merge: We apply several filters to the topic merge pairs. Initially, we remove 'trivial' merges, i.e., those merges which are minor lexical variants (e.g., plural forms). To take care of such cases, we apply the standard Jaro-Winkler similarity on the topic names and all topic pairs with similarity more than 0.8 were removed. After this, we remove all the merge pairs which were abbreviations like 'ICC' which essentially merged into 'International Cricket Council'. One should note that if a topic gets merged into another topic, all the questions of the former (source) topic change their tags to the latter (destination) one, so we would not be able to get the questions of the two distinct topics that were being merged. To tackle this problem, we separately re-crawl all the questions for the source and the destination topics by making a call to utility *https://www.quora.com/topic/[source/destination topic name]/all_questions*[3]. Then we obtain the set of questions that were posted before the merge for both the source and the destination topic. We then filter out those topics that did not have any questions tagged with them. After applying all the above filters, we obtain 2829 merge pairs.

Non-merge: In order to build a strong competitive negative set to contrast the behavior of the actual merge cases, we define the 'non-merge' class. In this class we consider topic unmerge pairs (which should be typically hard to distinguish from the true merges) and the topic pairs in the one-hop neighborhood of the actual merging pairs on the Quora ontology graph (i.e., conceptually close topics to the merging pair but did not themselves undergo any merge). We apply the same filters that we had applied to the merge pairs. Finally, we obtain 11,648 topic neighbor pairs and 2,421 topic unmerge pairs.

4 PROPERTIES OF MERGE & NON-MERGE TOPICS

In this section, we first look into the various reason for topic merging to occur in Quora and then study various characteristic properties that distinguish the actual topic merges from the non-merge cases.

4.1 Characteristics of Topic Merging

We carry out a detailed analysis on the 2,829 topic merge pairs and 14,069 (11, 648+2, 421) topic non-merge pairs.

[2] https://www.quora.com/Are-Quora-topic-hierarchies-a-directed-acyclic-graph

[3] Example: If the source topic name is 'ICC', we can make a call to *https://www.quora.com/topic/ICC/all_questions* to get all the questions under the topic 'ICC'.

4.1.1 *Question content features.*

We derive various properties from the question texts and observe the differences between the topic merge and non-merge pairs.

n-gram overlap of question texts: Question text overlaps between two topics are potential indicators of topic merging. To find out if there are question text overlaps, we extract 1, 2, 3 and 4-grams from the questions belonging to the topic pairs. For each n-gram, we compute the unweighted[4] and weighted version of the overlap coefficient.

In the weighted version of the overlap, instead of taking each common element once, we consider the minimum frequency of the common element in both the sets. The weighted version of the overlap coefficient is given in Equation 1

$$\sum_{K \epsilon X \cap Y} \frac{min(K\_freq(X), K\_freq(Y))}{min(\mid X \mid, \mid Y \mid)} \tag{1}$$

where K represents the set of question text n-gram that are common in the two topics X and Y. $K\_freq(X)$ computes the frequency of occurrence of K in Topic X and $\mid X \mid$ represents the number of n-gram elements formed using the questions in Topic X. Figure 2a shows the proportion of topics with bigram overlap (unweighted) distribution of question words. We can observe that non-merge topic pairs have lower overlap coefficient as compared to merge topic pairs.

Topic name in question text: If two topics are describing the same concept, then there is a high probability that this topic name will be present in the question text of the other topic. Based on this hypothesis, we extracted 1, 2, 3 and 4-grams of the topic name and the question words of the other topic. Figure 2b shows the unigram distribution of the word overlap of one topic's (topic 1) name and the words present in the other topic's (topic 2) questions. From the figure, we can clearly observe that this hypothesis is true for the merge pairs.

tf-idf similarity: Question similarity between topic pairs is also a factor influencing topics to merge. To obtain question similarity between the topics, we first generate the tf-idf vectors for the questions for both the topics. We then calculate the cosine similarity between them. Figure 2e shows the distribution. From the figure, it is evident that non-merge topics have lower similarity values as compared to the merge pairs.

Co-occurring topic overlap: Each topic is associated with a set of questions and each such question is associated with other topics. These topics are called the co-occurring topics. For each topic pair, we calculate the weighted and unweighted overlap coefficient of the co-occurring topics. Figure 2c shows the distribution of the unweighted overlap coefficients of the co-occurring topics. We can observe that the non-merge topics tend to have lower overlap coefficient as compared to the merge topics.

Word2Vec similarity: We use Word2Vec [36] to generate word embeddings for the words in the question text corresponding to a topic using gensim [42]. We set minimum word count to 5 and the size of feature vector to 150. We then generate a word vector for a document (i.e., all questions associated to a topic) by averaging all the vectors of the question words in that topic. After this, we perform cosine similarity to determine the semantic similarity of the two topics. It is evident from Figure 2j that the merge pairs have a higher similarity between their word vectors.

Doc2Vec similarity: We use Doc2Vec [24] to generate vectors for each document using gensim [42]. A document here consists of all questions associated to a particular topic. We set minimum word count to 5 and the size of feature vector to 150. The cosine similarity is calculated between the documents corresponding to a topic pair. Once again, the merge pairs seem to have higher overall similarity between their corresponding document vectors (see Figure 2k).

Part-of-speech: We use the Stanford POS tagger [33] to find the part of speech tag for each question word associated to a topic. For each topic we build a POS tag vector containing the count of the different POS tags in the associated questions. We then compute the similarity between two topics as the cosine similarity between the POS tag vectors of the two topics. This similarity is much higher for the merge pairs (see Figure 2l).

4.1.2 *Quora topic ontology.*

Quora topic ontology provides parent-child relationships between topics. We derived various features that make use of the ontology to show the contrast between the merge and the non-merge pairs.

Co-occurring topics' parent and child overlap: Apart from overlap between the co-occurring topics of the topic pair, we also calculate the parent and child topic overlap (weighted and unweighted) of the co-occurring topics. Figure 2d shows the distribution of the co-occurring topics' parent-child overlap (weighted) and it is evident from the plot that the merge pairs tend to have higher overlap as compared to the non-merge topic pairs.

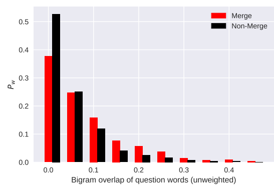
Path length between topics: In our merge dataset, only 17.4% of the pairs are part of the topic hierarchy. Rest of the pairs have at least one topic which is not part of the ontology graph. Therefore, calculating the path length between the topics directly is not very useful. Instead, we use the average minimum path length between the co-occurring topics. We first find the five most frequent co-occurring topics of each pair and then we calculate the average of the minimum path length between all pairs of these co-occurring topics. The distribution of the five least common topics in the merge and non merge pairs is presented in Figure 2g. We can observe that merge pairs tend to have lower average distance as compared to non-merge pairs.

Adamic/Adar similarity between co-occurring topics: Adamic / Adar similarity [1] is defined as inverted sum of degrees of the common neighbors for two topics in the Quora topic ontology graph.

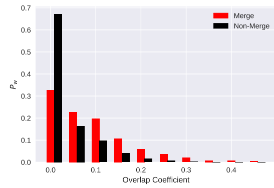
We compute the Adamic/Adar similarity value between the co-occurring topic pairs. Figure 2h shows the distribution of the average Adamic/Adar value for the ordered pairs of five most frequently co-occurring topics for each candidate topic pair. As usual, the merge pairs have higher Adamic/Adar values.

Similarity measures on the ontology: We also calculate semantic similarity between topics to distinguish merge and non-merge topic pairs. Semantic similarity of a set of topics is a metric in which the idea of distance between the two topics estimates the strength of the semantic relationship between the topics. We use the Quora topic ontology to find the similarity between two topics. Specifically, we use the Lin similarity [28], Resnik similarity [44], Wu-Palmer similarity [50] and JCN similarity [21] for this task. Figure 2i shows

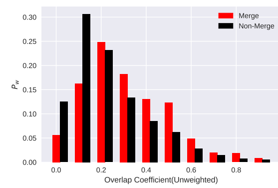
[4]https://en.wikipedia.org/wiki/Overlap_coefficient



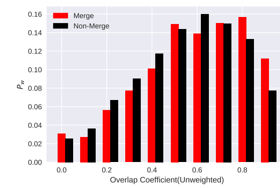
(a) n-gram overlap of question words.



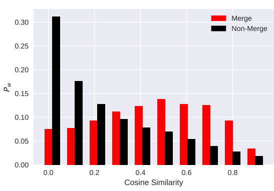
(b) Topic name in question text.



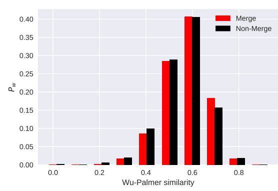
(c) Co-occurring topic overlap.



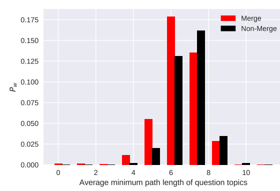
(d) Parent overlap of the co-occurring topics.



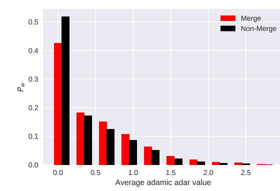
(e) Tf-idf similarity value.



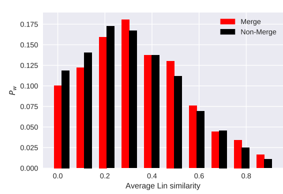
(f) Normalized Wu-Palmer similarity between question words corresponding to topic pairs.



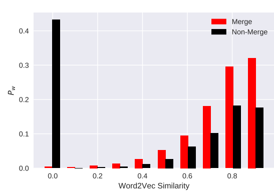
(g) Average minimum path length of five least frequently co-occurring topic.



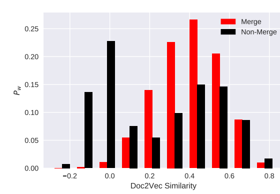
(h) Adamic/Adar score for the co-occurring topics.



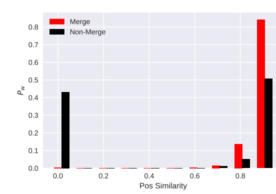
(i) Lin similarity of top co-occurring question topics.



(j) Word2Vec similarity.



(k) Doc2Vec similarity.



(l) Part of Speech similarity.

Figure 2: Distribution of various properties of the merge and non-merge pairs. The Y-axis represents the fraction of topics (P_w) and the X-axis represents the different properties of the topic pairs.

the distribution of average Lin similarity of five most frequently co-occurring topics for each merge and non-merge pair. The result shows that the merge pairs are semantically closer than the non-merge pairs.

4.1.3 External similarity features. ***Wordnet similarity***: We use several similarity measures defined over the Wordnet to calculate the similarity. For each topic pair, we calculate the similarity between the question words corresponding to the two topics. Figure 2f shows the similarity distribution for normalized Wu-Palmer. We can see that non-merge pairs have lower similarity as compared to merge pairs showing that the merge pairs have questions that are indeed semantically more close.

5 PREDICTION FRAMEWORK

A Quora user on average takes 936 days to find a duplicate topic and merge it with the original. One of the main reasons for this delay in merging is the huge number of topics in the Quora. It would require a lot of effort to manually inspect each pair of topics and determine, if they should merge or not. We propose to build a system that would help the user in this task. As the number of topic pairs possible could be millions, we cannot directly use a supervised classification task. Instead, we propose a two-step approach combining the *anomaly detection* and the *supervised classification framework* to automatically predict whether two given topics would merge or not in future. The positive instances for our approach are those topic pairs which should merge as they represent the same concept. The negative class, on the other hand, are the set of topic pairs which do not represent the same concept. The entire flow of the approach is illustrated in Figure 3 and described below.

5.1 Merge prediction

Positive training and test instances for the supervised classification task: We divide the merge instances into 7:3 ratio chronologically. This means that 70% (1981) of the merge cases that occur before *Feb 21, 2016* are considered as positive training instances and the rest 30% (849) of the merges which occur after this time point are considered as positive test instances.
Negative training instances for the supervised classification task: For the negative training instances of the supervised classification we consider the unmerge pairs and the neighborhood pairs (i.e., pairs in the non-merge class).

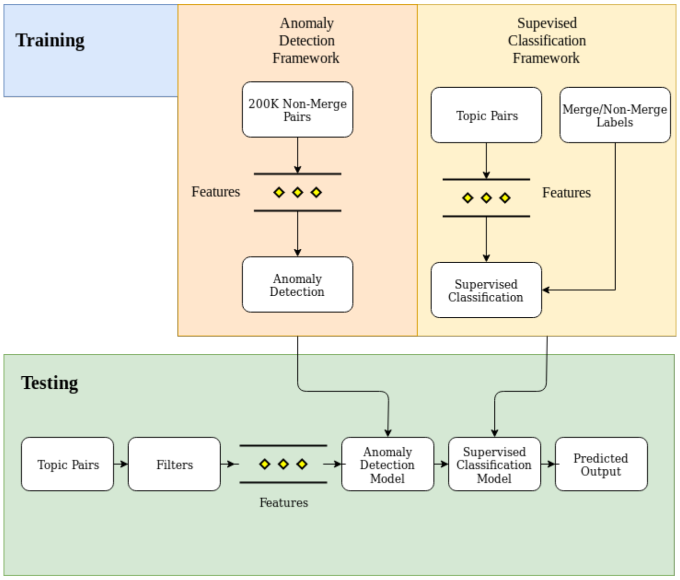


Figure 3: Flowchart of the two step classification approach. The anomaly detection framework acts as a filter which removes substantial negative test instances while retaining a good recall score on the positive test set. These are then passed to the supervised classifier for prediction.

Negative test instances for the supervised classification task: In the application scenario, we need to predict the pair of merged topics from among all the possible topic pairs. For the negative test set, therefore, we want to use all possible pairs of topics but with a set of appropriate filters. We describe these filters below.
Filters: As a first filter, we remove all those topics which had less than 20 questions in them. This brought down the number of topics to 45,743 (i.e., close to 2 billion topic pairs). Next, we apply the same set of filters that we applied to the merge topic pairs (removal of abbreviations and high Jaro-Winkler similarity topic pairs). This reduced the number of pairs to one billion. Further, we calculate the overlap coefficient of the co-occurring topics of each pair. We use a threshold of 0.25 to obtain around 20 million negative test pairs. We find that this filter allows 89% of the positive merge pairs. Note that we use all these filters to make the negative class more non-trivial. Since it was computationally almost impossible to generate all the possible features for the 20 million negative instances, we randomly select one million negative instances for the testing purpose.
Reducing the test instances for the supervised classification task: It is here that we use anomaly detection algorithms to obtain a reduced set of test instances that we can feed to the supervised classifier. In particular we generate 200K non-merge pairs (fully distinct from those used as negative examples for training the supervised classifier) to train the anomaly detection algorithms. Using this, we identify those test instances from the 1M (negative test) + 849 (positive test) topic pairs that are 'outliers', i.e., closely resemble the properties of the merge class and are therefore, strong candidates of merging. We retain these instances in the final test set for the supervised classifier. The key objective is to maintain a high recall while allowing for some false positives.
Models used for anomaly detection: We use standard anomaly detection methods such as One-Class SVM (OCSVM), Isolation Forest (IF), Local Outlier Factor (LOF), Elliptic Envelope (EE) for our experiments. We set contamination to 0.2 in all the experiments. For OCSVM, we set the kernel as 'rbf'. For IF, the number of estimators is 100. For LOF, we set the number of neighbors as 20.
Models used for supervised classification: The classification methods used include Random Forest (RF), Naive Bayes (NB), Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree (DT), XGBoost (XGB). Since our training data is imbalanced we use a cost-sensitive classification approach. In case of Random Forest (RF) we set the number of tree as 10. For SVM, we user Linear Kernal and C as 1.0. For LR, we use 'l2' penalty with C as 1.0 and set 'liblinear' as the optimization algorithm. For XGB, we set the learning rate as 0.1, booster as gbtree and use 100 boosted trees to fit.
The overall two-step method: As discussed above we first use an anomaly detection algorithm followed by a supervised classification approach. The anomaly detection framework acts as a filter which removes substantial negative test instances while retaining a good recall score on the positive test set. These are then passed to the supervised classifier for prediction. The whole process is explained in Figure 3.
Competing baselines: Examples of automatic merge prediction of linguistic entities is rare in the literature. We use two baselines to compare the performance of our model. We found that Maity et al. [32] has resemblance with our work in this respect and therefore, used it as the first baseline. We adapt the feature set which are appropriate for the topic merge and compare the performance. For the second baseline, we use Universal Sentence Encoder [10] to generate 512 dimensional vector representation of the topics. For each topic, we first concatenate all the questions and pass it to the encoder to generate the 512 dimensional vector representation for a topic. In order to determine if a topic pair is a merge or not, we simply calculate the cosine similarity between their vector representations. In order to distinguish between the merge and non-merge topic pairs, we define a threshold T for the similarity, above which we would call a topic pair as merge and below which it will be a non-merge. We use the training dataset to set the value of T to be the value of cosine similarity which maximizes the F-score. We found that a threshold value of 0.9 performs the best.

We have several interesting observations. First, we find that the precision of all the anomaly detection algorithms are very low. Second, the classification methods seems to perform better than the anomaly detection framework especially in precision and F-score. Third, we find that while the universal sentence encoder seems to be a stronger baseline than Maity et al. [32], the two-step approach performs the best and outperforms both the baselines by a huge margin. The best performance is obtained when we use isolation forest for anomaly detection followed by logistic regression for the classification. Our system could be used to improve the existing workflow at Quora by suggesting possible topics which represent the same concept. This would allow the community to detect and merge topics in an early stage of the topic evolution, in turn, enabling timely and appropriate knowledge aggregation.

Type	Algorithm	Precision	Recall	F-Score
Baseline	Maity et al. [32]	0.192	0.167	0.179
	Universal Sentence Encoder [10]	0.552	0.290	0.380
Outlier Detection	Local Outlier Factor	0.012	0.814	0.024
	Isolation Forest	**0.041**	**0.821**	**0.078**
	One Class Support Vector Machine	0.002	0.498	0.004
	Elliptic Envelope	0.013	0.917	0.026
Supervised Classification	Random Forest	0.328	0.491	0.393
	Naive Bayes	0.374	0.075	0.125
	Support Vector Machine	0.069	0.667	0.125
	Logistic Regression	**0.325**	**0.585**	**0.418**
	Decision Tree	0.003	0.901	0.005
	XGBoost	0.066	0.522	0.117
Our 2-step method	Elliptic Envelope + Logistic Regression	0.773	0.527	0.627
	Isolation Forest + Random Forest	0.706	0.624	0.662
	Isolation Forest + Logistic Regression	**0.866**	**0.603**	**0.711**
	Isolation Forest + Support Vector Machine	0.833	0.491	0.618
	Local Outlier Factor + Logistic Regression	0.612	0.597	0.604

Table 1: Results of topic merge prediction.

5.2 Direction of topic merge

When a user decides to merge two topics, he/she has to make a choice about the direction of merge. The direction decides the outcome of the competition between conventions. When two topics get merged, one of them ceases to exist and would no longer be accessible to the users directly. Let us assume that 'Topic A' gets absorbed into 'Topic B' to form the new 'Topic B', i.e., the winning convention. In the following we discuss the factors instrumental in influencing the direction of topic merge.

Number of characters in the topic name: Figure 4a shows the distribution of the number of characters in 'Topic A' and 'Topic B'. It is evident from the figure that 'Topic A' tends to have less number of characters (average 16.5 characters) as compared to 'Topic B' (average 23.6 characters).

Number of words in the topic name: The average number of words in 'Topic A' is 2.7 whereas 'Topic B' has an average of 3.6 words in the topic name. Figure 4b shows the distribution of the number of words in 'Topic A' and 'Topic B' which shows similar trend as that of Figure 4a.

Topic creation date: We compare the date of creation of topic to check if an old topic is being merged into a new one or vice-versa. We find that only 33% of the times, an older topic gets merged into a newer one.

Number of questions before merge: We calculate the number of questions in both the topics before the merge. Figure 4c shows the distribution of the number of questions in 'Topic A' and 'Topic B' before they were merged. We can observe that the number of questions in 'Topic A' is much lower than that in 'Topic B'.

Number of answers before merge: Figure 4d shows the distribution of the number of answers in 'Topic A' and 'Topic B'. We can clearly observe that 'Topic A' has lesser number of answers as compared to 'Topic B'.

Predicting the direction of topic merge: We use the above discriminating factors in a standard classification framework to predict the direction of topic merge. We use Support Vector Machine (SVM), Random Forest and Naive Bayes classifiers available in Weka Toolkit for classification. We perform a 10-fold cross-validation and report the results in Table 2. We achieve 88.857% accuracy with high precision and recall. From the table, we can observe that SVM achieves a slightly better performance than the other classifiers.

Classifier	Accuracy	Precision	Recall	F-Score
Support Vector Machine	**88.857**	**0.828**	**0.981**	**0.898**
Random Forest	88.379	0.863	0.913	0.887
Naive Bayes	87.546	0.823	0.956	0.885

Table 2: Results for predicting the direction of topic merge. For the two-step approach the best five results are only shown.

5.3 Early prediction of topic merge

In this section, we discuss the performance of our model to detect duplicate topics which should be merged, at an early stage itself. Note that this could be a direct potential application of our work. We use the best performing method (IF + LR) from the earlier section for our experiments. When a topic is created it takes quite some time for the Quora community to identify if it is a duplicate or not. We find that as per our test dataset, on average it takes 990 days for the user (936 days if we consider the full dataset) to identify a topic as duplicate and merge it. In order to identify how early our model can detect the duplicate topics, we use the following experimental setting.

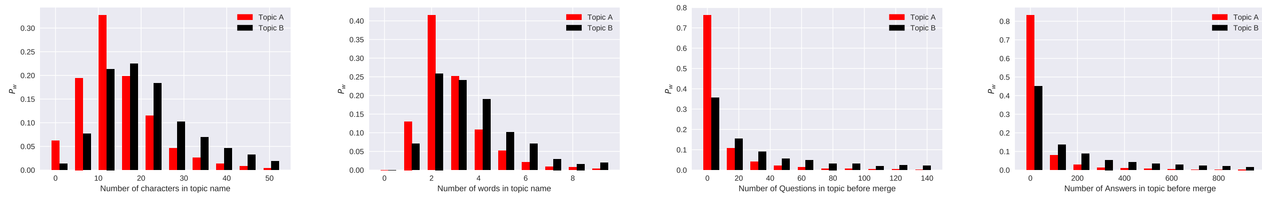
Let (A, B) be the topic pair, which is known to merge in the future in our test set and let B be the topic that has been created later. We compute the recall of our approach in early detection of this pair. We, therefore, restrict the data to these topics, which are used for computing the features. We take a snapshot every month till both the topics merge. Thus, the first snapshot will have only those questions which have been asked till the first month from the creation of topic B, and so on.

Figure 5 shows the recall of the model on a monthly basis. While the recall keeps improving over time, the result is encouraging since our model is able to detect $\sim$ 25% of the merges in the first month itself, $\sim$ 40% in the first year and $\sim$ 50% in the first 30 months. Thus, the model can potentially be deployed to early detect the topic merges. We stress this early prediction scheme can be directly plugged into Quora enabling efficient merging by both users as well as moderators.

6 HUMAN JUDGMENT EXPERIMENT

We conduct human judgment experiment to evaluate how our model performs compared to the case where human subjects (regular Quora users) are tasked to predict the merges. We randomly select 180 positive and 180 negative merge cases. The 180 negative cases consists of 90 unmerge and 90 neighborhood pairs. Each pair of topics is annotated by three different users. To understand if humans are able to predict whether two topics should merge or not as well as the direction of merge, we conduct an online survey[5]. The survey was participated by 45 people (researchers, students and professors) all of whom use Quora. Each participant was asked 24 questions consisting of two parts. Each question had the two

[5]http://tinyurl.com/lgbha5f



(a) Number of characters. **(b) Number of words.** **(c) Number of questions** **(d) Number of answers.**

Figure 4: Distribution of various properties influencing the direction of merge (merging of 'Topic A' and 'Topic B' to 'Topic B').

Evaluation Methods	Merge	Neighbor	Unmerge
Majority Voting	0.75	0.31	0.35
Micro-Average	0.62	0.22	0.29
Macro-Average	0.62	0.22	0.29

Table 3: Evaluation results for Question 1.

Evaluation Methods	Question 1	Question 2
Majority Voting	0.54	0.40
Micro-Average	0.33	0.49
Macro-Average	0.33	0.45

Table 4: Overall human judgment results.

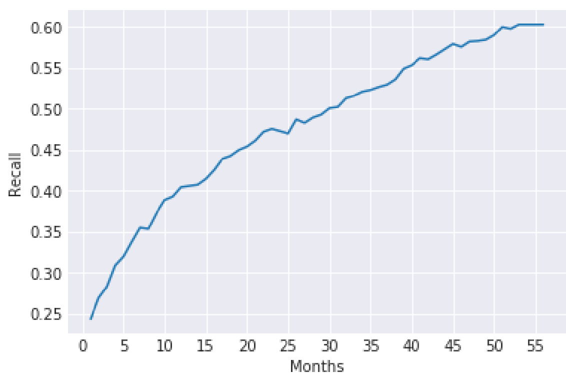


Figure 5: System recall as a function of time (in months) from the creation of the duplicate topic.

topics under consideration along with 3 example questions from each of the topics. The first question asked the participants whether the two topics should merge into a single topic (yes/no type). The second question asked for the direction of merge. A 'No' to the first question would enable subjects to skip the second. Apart from these, the participants were asked two additional questions to understand their familiarity / expertise. First, to understand how active they are on Quora, they were asked, "*How much time (in hours) do you spend on Quora?*". Second, for each topic pair they annotate, the participants were asked to inform their familiarity with the topic. We presented a scale of 1-5 for the participants to choose from, with the value 1 implying, "not familiar with the topic" and a value of 5 implying, "very familiar with the topic".

In Table 3, we report the human evaluation results for each of the three classes separately (Question 1). We observe that humans performed decently with majority voting accuracy of 0.75 for identification of merge classes. For the non-merge cases, humans performed very poorly (worst performance for the neighborhood class). These results indicate that the task of merging topics is difficult for Quora users as they seem to be confused by the non-merge cases in which they performed very poorly. Since the number of participants answering Question 1 for all the questions is 3, the micro-averaging and macro-averaging yields the same result. In Table 4, we show the overall human evaluation results for merge/non-merge identification (Question 1) and identification of direction of merge (Question 2). Overall, we observe that the regular Quora users find it difficult to judge if two given topics should merge or not, and in deciding the direction of merge. To compare our model's performance, we train the best model on the 2649 (i.e, 2829 - 180) merge cases, 2331 (i.e, 2421 - 90) unmerge cases and 11558 (i.e, 11648 - 90) neighbor instances, and test on the Human judged pairs. We observe that our system achieves an accuracy of 84% as compared to the 54% by human judgement.

7 CONCLUSION

In this paper, we studied the phenomena of competing conventions in Quora topics and proposed a model to predict whether two given topics should merge, as well as the direction of the topic merge. We make use of the Quora topic ontology and question content features to construct the model. Our two-step approach achieves an F1-score of ~ 0.71. We observe that the content features (especially similarity of vector embeddings like word2vec, Doc2vec) and topic ontology together perform significantly well for determining topic merges. In evaluation, the humans were able to predict if two topics should merge or not with an accuracy of 0.54 (compared to ~ 0.84 by our framework for the same data). We also propose an early prediction scheme and show that in $\sim 25\%$ cases the actual merge pairs can be predicted within a month and in $\sim 40\%$ cases within the first year. Users on average require 936 days (as per our dataset) to manually identify such merges. This, we believe, is an effective application that can be directly plugged into the Quora system to enable users/moderators to merge actual topic pairs early in time.

REFERENCES

[1] Lada A Adamic and Eytan Adar. 2003. Friends and neighbors on the web. *Social networks* 25, 3 (2003), 211–230.

[2] Lada A Adamic, Jun Zhang, Eytan Bakshy, and Mark S Ackerman. 2008. Knowledge sharing and yahoo answers: everyone knows something. In *WWW*. ACM, 665–674.

[3] Icek Ajzen and Martin Fishbein. 1970. The prediction of behavior from attitudinal and normative variables. *Journal of experimental social Psychology* 6, 4 (1970), 466–487.

[4] Giorgio Francesco Arcodia. 2007. Chinese: a language of compound words? *Selected proceedings of the 5th décembrettes: Morphology in Toulouse* (2007), 79–90.

[5] Solomon E Asch. 1955. Opinions and social pressure. *Readings about the social animal* 193 (1955), 17–26.

[6] William Badecker. 2001. Lexical composition and the production of compounds: Evidence from errors in naming. *Language and cognitive processes* 16, 4 (2001), 337–366.

[7] Leanne Bowler, Jocelyn Monahan, Wei Jeng, Jung Sun Oh, and Daqing He. 2015. The Quality and helpfulness of answers to eating disorder questions in yahoo! answers: Teens speak out. *Proceedings of the Association for Information Science and Technology* 52, 1 (2015).

[8] Rita Brdar-Szabó and Mario Brdar. 2008. On the marginality of lexical blending. *Jezikoslovlje* 9, 1-2 (2008), 171–194.

[9] Allyson K Carter and Cynthia G Clopper. 2002. Prosodic effects on word reduction. *Language and speech* 45, 4 (2002), 321–353.

[10] Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, et al. 2018. Universal sentence encoder. *arXiv preprint arXiv:1803.11175* (2018).

[11] Patrick Connolly. 2013. The innovation and adoption of English lexical blends. *JournaLIPP* 2 (2013), 1–14.

[12] Paul Cook. 2012. Using social media to find english lexical blends. In *Proc. of EURALEX*. 846–854.

[13] Paul Cook and Suzanne Stevenson. 2010. Automatically identifying the source words of lexical blends in English. *Computational Linguistics* 36, 1 (2010), 129–149.

[14] Michele Coscia. 2013. Competition and Success in the Meme Pool: A Case Study on Quickmeme. com.. In *ICWSM*.

[15] Lewis David. 1969. Convention: a philosophical study.

[16] Ullmann-Margalit Edna. 1977. The emergence of norms.

[17] Doron Friedman, Anthony Steed, and Mel Slater. 2007. Spatial social behavior in second life. In *Intelligent virtual agents*. 252–263.

[18] Livio Gaeta and Davide Ricca. 2009. Composita solvantur: Compounds as lexical units or morphological objects. *Rivista di Linguistica* 21, 1 (2009), 35–70.

[19] M Gareth Gaskell and William D Marslen-Wilson. 1999. Ambiguity, competition, and blending in spoken word recognition. *Cognitive Science* 23, 4 (1999), 439–462.

[20] H. J. Giegerich. 2009. *Compounding and Lexicalism.* Oxford University Press, 178–200.

[21] Jay J Jiang and David W Conrath. 1997. Semantic similarity based on corpus statistics and lexical taxonomy. *arXiv preprint cmp-lg/9709008* (1997).

[22] Farshad Kooti, Winter A Mason, Krishna P Gummadi, and Meeyoung Cha. 2012. Predicting emerging social conventions in online social networks. In *CIKM*. ACM, 445–454.

[23] Farshad Kooti, Haeryun Yang, Meeyoung Cha, P Krishna Gummadi, and Winter A Mason. 2012. The Emergence of Conventions in Online Social Networks.. In *ICWSM*.

[24] Quoc Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *ICML*. 1188–1196.

[25] Chih-yao Lee, Chia-hao Chang, Wei-chieh Hsu, and Shu-kai Hsieh. 2010. Qualia Modification in Noun-Noun Compounds: A Cross-Language Survey.. In *ROCLING*.

[26] Arnaud Léturgie. 2012. Are dictionaries of lexical blends efficient learners dictionaries. In *Proc. of EURALEX*. 619–625.

[27] Baichuan Li and Irwin King. 2010. Routing questions to appropriate answerers in community question answering services. In *CIKM*. 1585–1588.

[28] Dekang Lin et al. 1998. An information-theoretic definition of similarity.. In *ICML*, Vol. 98. 296–304.

[29] Suman Kalyan Maity, Aman Kharb, and Animesh Mukherjee. 2017. Language use matters: Analysis of the linguistic structure of question texts can characterize answerability in quora. In *Eleventh International AAAI Conference on Web and Social Media*.

[30] Suman Kalyan Maity, Aman Kharb, and Animesh Mukherjee. 2018. Analyzing the linguistic structure of question texts to characterize answerability in quora. *IEEE Transactions on Computational Social Systems* 5, 3 (2018), 816–828.

[31] Suman Kalyan Maity, Jot Sarup Singh Sahni, and Animesh Mukherjee. 2015. Analysis and Prediction of Question Topic Popularity in Community Q&A Sites: A Case Study of Quora.. In *ICWSM*. 238–247.

[32] Suman Kalyan Maity, Ritvik Saraf, and Animesh Mukherjee. 2016. #Bieber + #Blast = #BieberBlast: Early Prediction of Popular Hashtag Compounds. In *CSCW*. 50–63.

[33] Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. 2014. The stanford corenlp natural language processing toolkit.. In *ACL (System Demonstrations)*. 55–60.

[34] Binny Mathew, Ritam Dutt, Suman Kalyan Maity, Pawan Goyal, and Animesh Mukherjee. 2018. Deep Dive into Anonymity: A Large Scale Analysis of Quora Questions. *arXiv preprint arXiv:1811.07223* (2018).

[35] David A. Medler and Darren Piercey. 2004. Processing Ambiguous Words: Are Blends Necessary for Lexical Decision?

[36] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).

[37] Sunny Mitra, Ritwik Mitra, Suman Kalyan Maity, Martin Riedl, Chris Biemann, Pawan Goyal, and Animesh Mukherjee. 2015. An automatic approach to identify word sense changes in text media across timescales. *Natural Language Engineering* 21, 05 (2015), 773–798.

[38] J. L. Packard. 2010. *The morphology of Chinese: A linguistic and cognitive approach. Cambridge.* UK: Cambridge University Press.

[39] Aditya Pal, Shuo Chang, and Joseph A Konstan. 2012. Evolution of Experts in Question Answering Communities.. In *ICWSM*.

[40] Sumanth Patil and Kyumin Lee. 2016. Detecting experts on Quora: by their activity, quality of answers, linguistic characteristics and temporal behaviors. *Social Network Analysis and Mining* 6, 1 (2016), 5.

[41] Sharoda A Paul, Lichan Hong, and Ed H Chi. 2012. Who is authoritative? understanding reputation mechanisms in quora. *arXiv preprint arXiv:1204.3724* (2012).

[42] Radim Řehůřek and Petr Sojka. [n.d.]. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. 45–50.

[43] Vincent Renner, François Maniez, and Pierre Arnaud. 2013. *Cross-disciplinary perspectives on lexical blending*. Vol. 252. Walter De Gruyter.

[44] Philip Resnik et al. 1999. Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *J. Artif. Intell. Res.(JAIR)* 11 (1999), 95–130.

[45] Rahmtin Rotabi, Cristian Danescu-Niculescu-Mizil, and Jon Kleinberg. 2017. Competition and selection among conventions. In *WWW*. 1361–1370.

[46] H. Shen, Z. Li, J. Liu, and J. E. Grant. 2015. Knowledge Sharing in the Online Social Network of Yahoo! Answers and Its Implications. *IEEE Trans. Comput.* 64, 6 (June 2015), 1715–1728. https://doi.org/10.1109/TC.2014.2322598

[47] Florence Villoing. 2012. French compounds. (2012).

[48] Gang Wang, Konark Gill, Manish Mohanlal, Haitao Zheng, and Ben Y Zhao. 2013. Wisdom in the social crowd: an analysis of quora. In *WWW*. ACM, 1341–1352.

[49] Deanna Wilkes-Gibbs and Herbert H Clark. 1992. Coordinating beliefs in conversation. *Journal of memory and language* 31, 2 (1992), 183–194.

[50] Zhibiao Wu and Martha Palmer. 1994. Verbs semantics and lexical selection. In *ACL*. 133–138.

[51] Jiang Yang and Xiao Wei. 2009. Seeking and Offering Expertise Across Categories: A Sustainable Mechanism Works for Baidu Knows.. In *ICWSM*. Citeseer.

[52] Nick Yee, Jeremy N Bailenson, Mark Urbanek, Francis Chang, and Dan Merget. 2007. The unbearable likeness of being digital: The persistence of nonverbal social norms in online virtual environments. *CyberPsychology & Behavior* 10, 1 (2007), 115–121.

[53] Kai Zhang, Wei Wu, Haocheng Wu, Zhoujun Li, and Ming Zhou. 2014. Question Retrieval with High Quality Answers in Community Question Answering. In *CIKM*. 371–380.