Together We Stand: Siamese Networks for Similar Question Retrieval

Arpita Das[1] Harish Yenala[1] Manoj Chinnakotla[2,1] Manish Shrivastava[1]

[1]IIIT Hyderabad, Hyderabad, India
{arpita.das,harish.yenala}@research.iiit.ac.in
m.shrivastava@iiit.ac.in

[2]Microsoft, Hyderabad, India
manojc@microsoft.com

Abstract

Community Question Answering (cQA) services like Yahoo! Answers[1], Baidu Zhidao[2], Quora[3], StackOverflow[4] *etc.* provide a platform for interaction with experts and help users to obtain precise and accurate answers to their questions. The time lag between the user posting a question and receiving its answer could be reduced by retrieving similar historic questions from the cQA archives. The main challenge in this task is the "lexico-syntactic" gap between the current and the previous questions. In this paper, we propose a novel approach called *"Siamese Convolutional Neural Network for cQA (SCQA)"* to find the semantic similarity between the current and the archived questions. SCQA consist of twin convolutional neural networks with shared parameters and a contrastive loss function joining them.

SCQA learns the similarity metric for question-question pairs by leveraging the question-answer pairs available in cQA forum archives. The model projects semantically similar question pairs nearer to each other and dissimilar question pairs farther away from each other in the semantic space. Experiments on large scale real-life "Yahoo! Answers" dataset reveals that SCQA outperforms current state-of-the-art approaches based on translation models, topic models and deep neural network based models which use non-shared parameters.

1 Introduction

The cQA forums have emerged as popular and effective means of information exchange on the Web. Users post queries in these forums and receive precise and compact answers in stead of a list of documents. Unlike in Web search, opinion based queries are also answered here by experts and users based on their personal experiences. The question and answers are also enhanced with rich metadata like categories, subcategories, user expert level, user votes to answers *etc.*

One of the serious concerns in cQA is *"question-starvation"* (Li and King, 2010) where a question does not get immediate answer from any user. When this happens, the question may take several hours and sometimes days to get satisfactory answers or may not get answered at all. This delay in response may be avoided by retrieving semantically related questions from the cQA archives. If a similar question is found in the database of previous questions, then the corresponding best answer can be provided without any delay. However, the major challenge associated with retrieval of similar questions is the *lexico-syntactic gap* between them. Two questions may mean the same thing but they may differ lexically and syntactically. For example the queries *"Why are yawns contagious?"* and *"Why do we yawn when we see somebody else yawning?"* convey the same meaning but differ drastically from each other in terms of words and syntax.

Several techniques have been proposed in the literature for similar question retrieval and they could be broadly classified as follows:

1. **Classic Term Weighting Based Approaches:** Classical IR based retrieval

---

[1]https://answers.yahoo.com/
[2]http://zhidao.baidu.com/
[3]http://www.quora.com/
[4]http://stackoverflow.com/

models like $BM25$ (Robertson et al., 1994) and Language modeling for Information Retrieval ($LMIR$) (Zhai and Lafferty, 2004) score the similarity based on the weights of the matching text terms between the questions.

2. **Translation Models:** Learning word or phrase level translation models from question-answer pairs in parallel corpora of same language (Jeon et al., 2005; Xue et al., 2008; Zhou et al., 2011). The similarity function between questions is then defined as the probability of translating a given question into another.

3. **Topic Models:** Learning topic models from question-answer pairs (Ji et al., 2012; Cai et al., 2011; Zhang et al., 2014). Here, the similarity between questions, is defined in the latent topic space discovered by the topic model.

4. **Deep Learning Based Approaches:** Deep Learning based models like (Zhou et al., 2016),(Qiu and Huang, 2015), (Das et al., 2016) use variants of neural network architectures to model question-question pair similarity.

Retrieving semantically similar questions can be thought of as a classification problem with large number of categories. Here, each category contains a set of related questions and the number of questions per category is small. It is possible that given a test question, we find that there are no questions semantically related to it in the archives, it will belong to a entirely new unseen category. Thus, only a subset of the categories is known during the time of training. The intuitive approach to solve this kind of problem would to learn a similarity metric between the question to be classified and the archive of previous questions. Siamese networks have shown promising results in such distance based learning methods (Bromley et al., 1993; Chopra et al., 2005). These networks possess the capability of learning the similarity metric from the available data, without requiring specific information about the categories.

In this paper, we propose a novel unified model called *Siamese Convolutional Neural Network for cQA*. $SCQA$ architecture contain deep convolutional neural networks as twin networks with a contrastive energy function at the top. These twin networks share the weights with each other (parameter sharing). The energy function used is suitable for discriminative training for Energy-Based models (LeCun and Huang, 2005). $SCQA$ learns the shared model parameters and the similarity metric by minimizing the energy function connecting the twin networks. Parameter sharing guarantees that question and its relevant answer are nearer to each other in the semantic space while the question and any answer irrelevant to it are far away from each other. For example, the representations of *"President of USA"* and *"Barack Obama"* should be nearer to each other than those of *"President of USA"* and *"Tom Cruise lives in USA"*. The learnt similarity metric is used to retrieve semantically similar questions from the archives given a new posted question.

Similar question pairs are required to train SCQA which is usually hard to obtain in large numbers. Hence, $SCQA$ overcomes this limitation by leveraging Question-Answer pairs (Q, A) from the cQA archives. This also has additional advantages such as:

• The knowledge and expertise of the answerers and askers usually differ in a cQA forum. The askers, who are novices or non-experts, usually use less technical terminology whereas the answerers, who are typically experts, are more likely to use terms which are technically appropriate in the given realm of knowledge. Due to this, a model which learns from Question-Answer (Q, A) training data has the advantage of learning mappings from non-technical and simple terms to technical terms used by experts such as $shortsight => myopia$ etc. This advantage will be lost if we learn from (Q, Q) pairs where both the questions are posed by non-experts only.

• Experts usually include additional topics that are correlated to the question topic which the original askers may not even be aware of. For example, for the question *"how can I overcome short sight?"*, an expert may give an answer containing the concepts *"laser surgery"*, *"contact lens"*, *"LASIK surgery"* etc. Due to this, the concept *short sight* gets associated with these expanded concepts as well. Since, the askers are non-experts, such

rich concept associations are hard to learn from (Q, Q) training archives even if they are available in large scale. Thus, leveraging (Q, A) training data leads to learning richer concept/term associations in SCQA.

In summary, the following are our main contributions in this paper:

- We propose a novel model $SCQA$ based on Siamese Convolutional Neural Network which use shared parameters to learn the similarity metric between question-answer pairs in a cQA dataset.

- In SCQA, we overcome the non-availability of training data in the form of question-question pairs by leveraging existing question-answer pairs from the cQA archives which also helps in improving the effectiveness of the model.

- We reduce the computational complexity by directly using character-level representations of question-answer pairs in stead of using sentence modeling based representations which also helps in handling spelling errors and out-of-vocabulary (OOV) words in documents.

The rest of the paper is organized as follows. Section 2 presents the previous approaches to conquer the problem. Section 3 describes the architecture of $SCQA$. Sections 4 and 5 explain the training and testing phase of $SCQA$ respectively. Section 6 introduces a variant of $SCQA$ by adding textual similarity to it. Section 7 describes the experimental set-up, details of the evaluation dataset and evaluation metrics. In Section 8, quantitative and qualitative results are presented. Finally, Section 9 concludes the paper.

2 Related Work

The classical retrieval models $BM25$ (Robertson et al., 1994), $LMIR$ (Zhai and Lafferty, 2004) do not help much to capture semantic relatedness because they mainly consider textual similarity between queries. Researchers have used translation based models to solve the problem of question retrieval. Jeon *et al.* (2005) leveraged the similarity between the archived answers to estimate the translation probabilities. Xue *et al.* (2008) enhanced the performance of word based translation

model by combining query likelihood language model to it. Zhou *et al.* (2011) used phrase based translation model where they considered question answer pairs as parallel corpus. However, Zhang *et al.* (2014) stated that questions and answers cannot be considered parallel because they are heterogeneous in lexical level and in terms of user behaviors. To overcome these vulnerabilities topic modeling was introduced by (Ji et al., 2012; Cai et al., 2011; Zhang et al., 2014). The approach assumes that questions and answers share some common latent topics. These techniques match questions not only on a term level but also on a topic level.

Zhou *et al.* (2015) used a fisher kernel to model the fixed size representation of the variable length questions. The model enhances the embedding of the questions with the metadata "category" involved with them. Zhang *et al.* (2016) learnt representations of words and question categories simultaneously and incorporated the learnt representations into traditional language models.

Following the recent trends, deep learning is also employed to solve this problem. Qiu *et al.* (2015) introduced convolutional neural tensor network (CNTN), which combines sentence modeling and semantic matching. CNTN transforms the word tokens into vectors by a lookup layer, then encode the questions and answers to fixed-length vectors with convolutional and pooling layers, and finally model their interactions with a tensor layer. Das *et al.* (2016) used deep structured topic modeling that combined topic model and paired convolutional networks to retrieve related questions. Zhou *et al.* (2016) used a deep neural network (DNN) to map the question answer pairs to a common semantic space and calculated the relevance of each answer given the query using cosine similarity between their vectors in that semantic space. Finally they fed the learnt semantic vectors into a learning to rank (LTR) framework to learn the relative importance of each feature.

On a different line of research, several Textual-based Question Answering (QA) systems (Qanda[5], QANUS[6], QSQA[7] *etc.*) are developed that retrieve answers from the Web and other textual sources. Similarly, structured QA systems
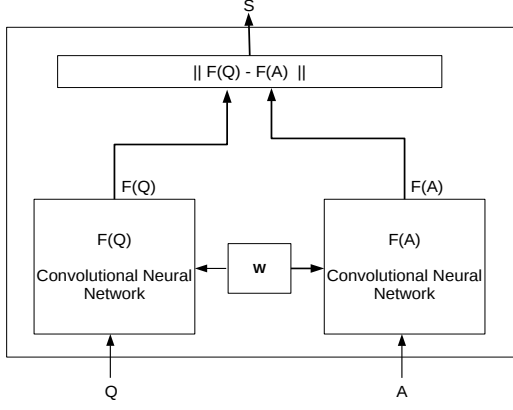
---

[6]http://www.qanus.com/
[7]http://www.dzonesoftware.com/products/open-source-question-answer-software/
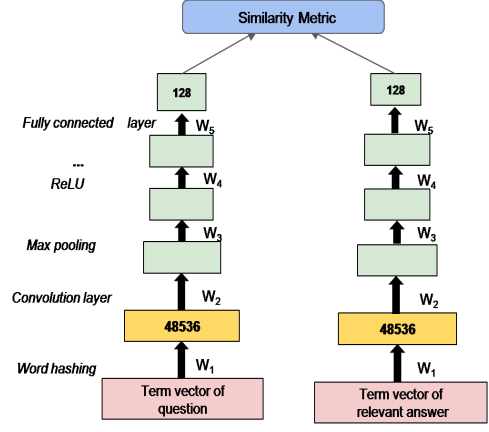
Figure 1: Architecture of Siamese network.

Figure 2: Architecture of SCQA. The network consists of repeating convolution, max pooling and ReLU layers and a fully connected layer. Also the weights W_1 to W_5 are shared between the sub-networks.

(Aqualog[8], NLBean[9] *etc.*) obtain answers from structured information sources with predefined ontologies. QALL-ME Framework (Ferrandez et al., 2011) is a reusable multilingual QA architecture built using structured data modeled by an ontology. The reusable architecture of the system may be utilized later to incorporate multilingual question retrieval in $SCQA$.

2.1 Siamese Neural Network

Siamese Neural Networks (shown in Figure 1) were introduced by Bromley *et al.* (1993) to solve the problem of signature verification. Later, Chopra *et al.* (2005) used the architecture with discriminative loss function for face verification. Recently these networks are used extensively to enhance the quality of visual search (Liu et al., 2008; Ding et al., 2008).

Let, $F(X)$ be the family of functions with set of parameters W. $F(X)$ is assumed to be differentiable with respect to W. Siamese network seeks a value of the parameter W such that the symmetric similarity metric is small if X_1 and X_2 belong to the same category, and large if they belong to different categories. The scalar energy function $S(Q, A)$ that measures the semantic relatedness between question answer pair (Q,A) can be defined as:

$$S(Q, A) = \|F(Q) - F(A)\| \qquad (1)$$

In $SCQA$ question and relevant answer pairs are fed to train the network. The loss function is minimized so that $S(Q, A)$ is small if the answer A is relevant to the question Q and large otherwise.

3 Architecture of SCQA

As shown in Figure 2, $SCQA$ consists of a pair of deep convolutional neural networks (CNN) with convolution, max pooling and rectified linear (ReLU) layers and a fully connected layer at the top. CNN gives a non linear projection of the question and answer term vectors in the semantic space. The semantic vectors yielded are connected to a layer that measures distance or similarity between them. The contrastive loss function combines the distance measure and the label. The gradient of the loss function with respect to the weights and biases shared by the sub-networks, is computed using back-propagation. Stochastic Gradient Descent method is used to update the parameters of the sub-networks.

3.1 Inputs to SCQA

The size of training data used is in millions, thus representing every word with one hot vector would be practically infeasible. Word hashing introduced by Mcnamee *et al.* (2004) involves letter n-gram to reduce the dimensionality of term vectors. For a word, say, "table" represented as (#table#) where # is used as delimiter, letter 3-grams would be #ta, tab, abl, ble, le#. Thus word hashing is character level representation of documents which takes care of OOV words and words with minor spelling errors. It represents a query using a lower dimensional vector with dimension equal to number of unique letter trigrams in the training dataset (48,536 in our case).

The input to the twin networks of $SCQA$ are word hashed term vectors of the question and

answer pair and a label. The label indicates whether the sample should be placed nearer or farther in the semantic space. For positive samples (which are expected to be nearer in the semantic space), twin networks are fed with word hashed vectors of question and relevant answers which are marked as "best-answer" or "most voted answers" in the cQA dataset of Yahoo! Answers (question-relevant answer pair). For negative samples (which are expected to be far away from each other in the semantic space), twin networks are fed with word hashed vectors of question and answer of any other random question from the dataset (question-irrelevant answer pair).

3.2 Convolution

Each question-answer pair is word hashed into (q_i-a_i) such that $q_i \in \mathbb{R}^{n_t}$ and $a_i \in \mathbb{R}^{n_t}$ where n_t is the total number of unique letter trigrams in the training data. Convolution layer is applied on the word hashed question answer vectors by convolving a filter with weights $\mathbf{c} \in \mathbb{R}^{hxw}$ where h is the filter height and w is the filter width. A filter consisting of a layer of weights is applied to a small patch of word hashed vector to get a single unit as output. The filter is slided across the length of vector such that the resulting connectivity looks like a series of overlapping receptive fields which output of width w.

3.3 Max Pooling

Max pooling performs a kind of non-linear downsampling. It splits the filter outputs into small non-overlapping grids (larger grids result to greater the signal reduction), and take the maximum value in each grid as the value in the output of reduced size. Max pooling layer is applied on top of the output given by convolutional network to extract the crucial local features to form a fixed-length feature vector.

3.4 ReLU

Non-linear function Rectified linear unit (ReLU) is applied element-wise to the output of max pooling layer. ReLU is defined as $f(x) = max(0, x)$. ReLU is preferred because it simplifies backpropagation, makes learning faster and also avoids saturation.

3.5 Fully Connected layer

The terminal layer of the convolutional neural sub-networks is a fully connected layer. It converts the output of the last ReLU layer into a fixed-length semantic vector $\mathbf{s} \in \mathbb{R}^{n_s}$ of the input to the sub-network. We have empirically set the value of n_s to 128 in $SCQA$.

4 Training

We train $SCQA$ for a question while looking for semantic similarity with the answers relevant to it. $SCQA$ is different from the other deep learning counterparts due to its property of parameter sharing. Training the network with a shared set of parameters not only reduces number of parameters (thus, save lot of computations) but also ensures consistency of the representation of questions and answers in semantic space. The shared parameters of the network are learnt with the aim to minimize the semantic distance between the question and the relevant answers and maximize the semantic distance between the question and the irrelevant answers.

Given an input $\{q_i, a_i\}$ where q_i and a_i are the ith question answer pair, and a label y_i with $y_i \in \{1,-1\}$, the loss function is defined as:

$$loss(q_i, a_i) = \begin{cases} 1 - \cos(q_i, a_i), & \text{if } y = 1; \\ \max(0, \cos(q_i, a_i) - m), & \text{if } y = -1; \end{cases}$$

where m is the margin which decides by how much distance dissimilar pairs should be moved away from each other. It generally varies between 0 to 1. The loss function is minimized such that question answer pairs with label 1 (question-relevant answer pair) are projected nearer to each other and that with label -1 (question-irrelevant answer pair) are projected far away from each other in the semantic space. The model is trained by minimizing the overall loss function in a batch. The objective is to minimize :

$$L(\Lambda) = \sum_{(q_i, a_i) \in C \cup C'} loss(q_i, a_i) \quad (2)$$

where C contains batch of question-relevant answer pairs and C' contain batch of question-irrelevant answer pairs. The parameters shared by the convolutional sub-networks are updated using Stochastic Gradient escent (SGD).

5 Testing

While testing, we need to retrieve similar questions given a query. During testing we make pairs of all the questions with the query and feed them

to $SCQA$. The term vectors of the question pairs are word hashed and fed to the twin sub-networks. The trained shared weights of the $SCQA$ projects the question vector in the semantic space. The similarity between the pairs is calculated using the similarity metric learnt during the training. Thus $SCQA$ outputs a value of distance measure (score) for each pair of questions. The threshold is dynamically set to the average similarity score across questions and we output only those which have a similarity greater than the average similarity score.

6 Siamese Neural Network with Textual Similarity

$SCQA$ is trained using question-relevant answer pairs as positive samples and question-irrelevant answer pairs as negative samples. It poorly models the basic text similarity because in the (Q, A) training pairs, the answerers often do not repeat the question words while providing the answer. For example: for the question "Who is the President of the US?", the answerer would just provide "Barrack Obama". Due to this, although the model learns that $president\ the\ US\ =>\ Barrack\ Obama$, the similarity for $president => president$ wouldn't be much and hence needs to be augmented through $BM25$ or some such similarity function.

Though $SCQA$ can strongly model semantic relations between documents, it needs boosting in the area of textual similarity. The sense of word based similarity is infused to $SCQA$ by using BM25 ranking algorithm. Lucene[10] is used to calculate the BM25 scores for question pairs. The score from similarity metric of $SCQA$ is combined with the BM25 score. A new similarity score is calculated by the weighted combination of the $SCQA$ and $BM25$ score as:

$$score = \alpha * SCQA_{score} + (1 - \alpha) * BM25_{score} \tag{3}$$

where α control the weights given to $SCQA$ and $BM25$ models. It range from 0 to 1. $SCQA$ with this improved similarity metric is called Siamese Convolutional Neural Network for cQA with Textual Similartity (*T-SCQA*). Figure 4 depicts the testing phase of *T-SCQA*. This model will give better performance in datasets with good mix of questions that are lexically and semantically

---

[10] https://lucene.apache.org/

Hyperparameter	Value
Batch Size	100
Depth of CNN	3
Learning rate	0.01
Momentum	0.05
Kernel width of Convolution	10
Kernel width of MaxPooling	100
Length of semantic vector	128

Table 1: Hyperparameters of $SCQA$.

similar. The value of α can be tuned according to the nature of dataset.

7 Experiments

We collected Yahoo! Answers dataset from Yahoo! Labs Webscope[11]. Each question in the dataset contains title, description, best answer, most voted answers and meta-data like categories, sub categories etc. For training dataset, we randomly selected 2 million data and extracted question-relevant answer pairs and question-irrelevant answer pairs from them to train $SCQA$. Similarly, our validation dataset contains 400,000 question answer pairs. The hyperparameters of the network are tuned on the validation dataset. The values of the hyperparameters for which we obtained the best results is shown in Table 1.

We used the annotated survey dataset of 1018 questions released by Zhang *et al.* (2014) as testset for all the models. On this gold data, we evaluated the performance of the models with three evaluation criteria: Mean Average Precision (MAP), Mean Reciprocal Rank (MRR) and Precision at K (P@K).

Each question and answer was pre-processed by lower-casing, stemming, stopword and special character removal.

7.1 Parameter Sharing

In order to find out whether parameter sharing helps in the present task we build two models named Deep Structured Neural Network for Community Question Answering($DSQA$) and Deep Structured Neural Network for Community Question Answering with Textual Similarity *T-DSQA*. $DSQA$ and *T-DSQA* have the same architecture as $SCQA$ and *T-SCQA* with the exception that in
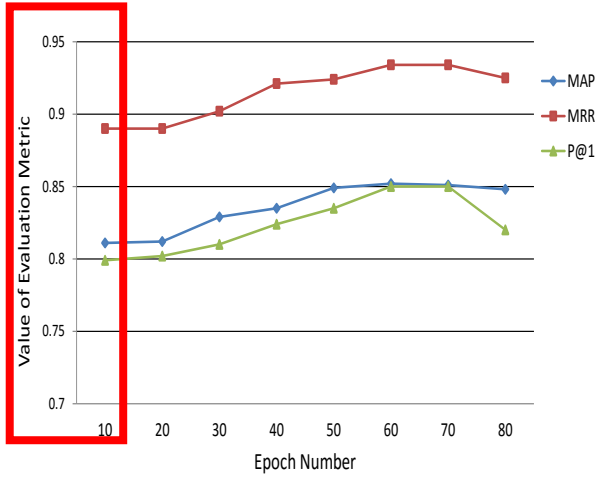
---

[11] http://webscope.sandbox.yahoo.com/catalog.php?datatype=l

Figure 3: Variation of evaluation metrics with the epochs.

Method	MAP	MRR	P@1
LMIR	0.762	0.844	0.717
translation(word)	0.786	0.870	0.807
translation+LM	0.787	0.869	0.804
translation(phrase)	0.789	0.875	0.817
Q-A topic model	0.787	0.879	0.810
Q-A topic model(s)	0.800	0.888	0.820
DSQA	0.755	0.921	0.751
T-DSQA	0.801	0.932	0.822
SCQA	0.811	0.895	0.830
T-SCQA	**0.852***	**0.934***	**0.849***

Table 2: Results on Yahoo! Answers dataset. The best results are obtained by *T-SCQA* (bold faced). The difference between the results marked(*) and other methods are statistically significant with p <0.001.

8 Results

We did a comparative study of the results of the previous methods with respect to $SCQA$ and *T-SCQA*. The baseline performance is shown by query likelihood language model (LM). For the translation based methods $translation(word)$, $translation+LM$ and $translation(phrase)$ we implemented the papers by Jeon *et al.* (2005), Xue *et al.* (2008), Zhou *et al.* (2011) respectively. The first paper deals with word based translation, the second enhanced the first by adding language model to it and the last paper implements phrase based translation method to bridge lexical gap. As seen from Table 2, the translation based methods outperforms the baseline significantly. The models are trained using GIZA++[12] tool with the question and best answer pair as the parallel corpus. For the topic based *Q-A topic model* and *Q-A topic model(s)*, we implemented the models $QATM$-PR (Question-Answer Topic Model) Ji *et al.*(2012) and $TBLM_{SQATM-V}$ (Supervised Question-Answer Topic Model with user votes as supervision) Zhang *et al.* (2014) respectively. Again it is visible from the Table 2 that topic based approaches show slight improvement over translation based methods but they show significant improvement over baseline. The mod-

---

[12]http://www.statmt.org/moses/giza/GIZA++.html

els $DSQA$ and *T-DSQA* were built using convolutional neural sub-networks joined by a distance measure at the top. There is no sharing of parameters involved between the sub-networks of these models. It is clear from the comparison of results between *T-DSQA* and *T-SCQA* that parameter sharing definitely helps in the task of similar question retrieval in cQA forums. *T-SCQA* outperforms all the previous approaches significantly.

8.1 Quantitative Analysis

$SCQA$ and *T-SCQA* learns the semantic relationship between the question and their best and most voted answers. It is observed that by varying the weights of $SCQA$ and $BM25$ scores, the value of MAP changes significantly (Figure 5). The weight is tuned in the validation dataset. We trained our model for several epochs and observed how the results varied with the epochs. We found that the evaluation metrics changed with increasing the number of epochs but became saturated after epoch 60. The comparison of evaluation metrics with epochs can be visualised in Figure 3.

The comparisons $SCQA$ and *T-SCQA* with the previously proposed models is shown in Table 2. For baseline we considered the traditional language model $LMIR$. The results in the table are consistent with the literature which says translation based models outperform the baseline methods and topic based approaches outperform the translational methods. Also, it is observed that deep learning based solution with parameter sharing is more helpful for this task than without parameter sharing. Note, that the results of previous models stated in Table 2 differ from the original
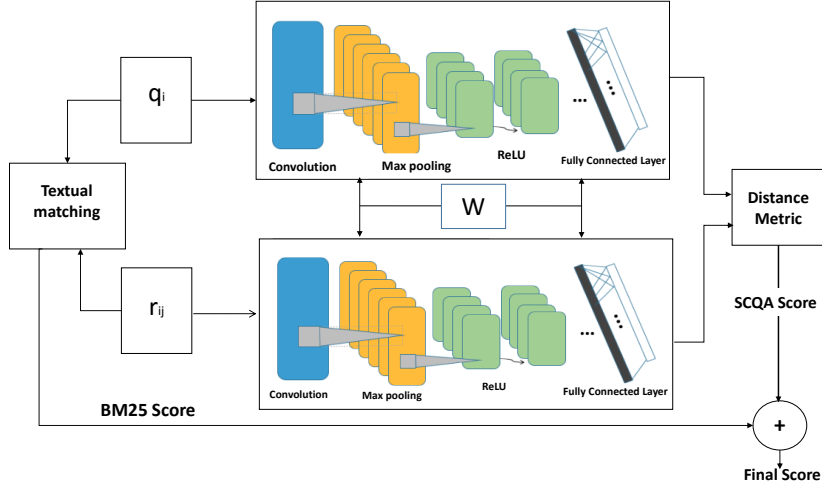
Figure 4: Testing phase of *T-SCQA*. Here the q_i is the ith query and r_{ij} is the jth question retrieved by q_i. The twin CNN networks share the parameters ($\mathbf{W}$) with each other. The connecting distance metric layer outputs the $SCQA$ score and the textual matching module outputs the $BM25$ score. The weighted combination of these scores give the final score. r_{ij} is stated similar to the query q_i if the final score of the pair exceeds an appropriate threshold.
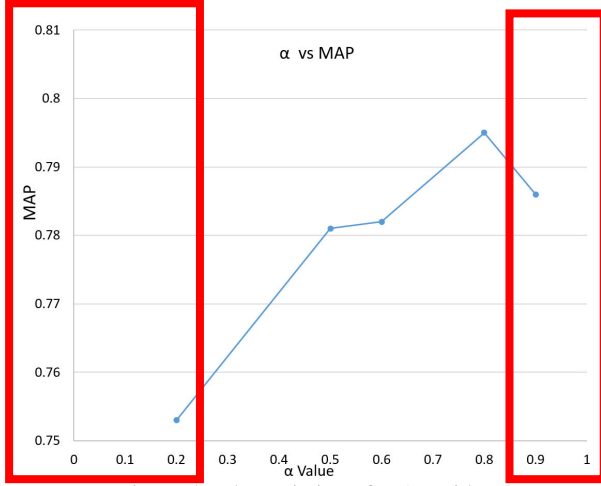

Figure 5: The variation of MAP with α.

papers since we tried to re-implement those models with our training data (to the best of our capability). Though we use the test data released by Zhang *et al.* (2014) we do not report their results in Table 2 due to the difference in training data used to train the models.

In the test dataset released by Zhang *et al.* (2014), there are fair amount of questions that possess similarity in the word level hence *T-SCQA* performed better than $SCQA$ for this dataset. *T-SCQA* gives the best performance in all evaluation measures. The results of *T-SCQA* in Table 2 uses the trained model at epoch 60 with the value of α as 0.8.

8.2 Qualitative Analysis

In Table 3 few examples are shown to depict how results of *T-SCQA* reflect strong semantic information when compared to other baseline methods. For Q1 we compare performance of *LMIR* and *T-SCQA*. *LMIR* outputs the question by considering word based similarity. It focuses on matching the words "how", "become", "naturally" *etc*, hence it outputs "How can I be naturally funny?" which is irrelevant to the query. On the other hand, $T-SCQA$ retrieves the questions that are semantically relevant to the query. For Q2 we compare the performance of *T-SCQA* with phrase based translation model (Zhou et al., 2011). The outputs of *translation(phrase)* model shows that the translation of "nursery" and "pre-school" to "daycare", "going to university" to "qualifications" are highly probable. The questions retrieved are semantically related, however asking craft ideas for pre-school kids for the event of mother's day is irrelevant in this context. The results of our model solely focuses on the qualifications, degrees and skills one needs to work in a nursery. For Q3 we compare the performance of *T-SCQA* with supervised topic model (Zhang et al., 2014). The questions retrieved by both the models revolve around the topic "effect of smoking on children". While the topic model retrieve questions which deal with smoking by mother and its effect on child, *T-SCQA* retrieve questions which deals not only with the affects of a mother smoking but also the effect of passive smoking on the child. For Q4 we com-

	Query	Comment
Q1: How can I become naturally happy?		$LMIR$ performs word based matching using "how","become", "naturally" *etc.*
$LMIR$	1.How can I be naturally happy? 2.How can I become naturally funny?	
T-$SCQA$	1.Are some of us naturally born happy or do we learn how to become happy? 2.How can I become prettier and feel happier with myself?	
Q2: Do you need to go to university to work in a nursery or pre-school?		For *translation* (phrase) university->degree nursery->daycare are highly probable translations but craft ideas for daycare is irrelevant.
translation (*phrase*)	1.What degree do you need to work in a nursery? 2. I work at a daycare with pre-school kids(3-5). Any ideas on crafts for mother's day?	
T-$SCQA$	1.Will my B.A hons in childhood studies put me in as an unqualified nursery nurse? 2.What skills are needed to work in a nursery, or learned from working in a nursery?	
Q3: Does smoking affect an unborn child?		Both models retrieve questions on topic "effect of smoking on children" but *T-SCQA* could retrieve based on passive smoking through father.
Q-A topic model(s)	1.How do smoking cigarettes and drinking affect an unborn child? 2.How badly will smoking affect an unborn child?	
T-$SCQA$	1.How does cigarette smoking and alcohol consumption by mothers affect unborn child? 2.Does smoking by a father affect the unborn child? If there is no passive smoking, then is it still harmful?	
Q4: How do I put a video on YouTube?		*T-DSQA* could not decipher "put". It relates "put" to download and transfer of videos while *T-SCQA* relates it to uploading videos.
T-DSQA	1.How can I download video from YouTube and put them on my Ipod? 2.I really want to put videos from YouTube to my Ipod..how?	
T-$SCQA$	1.How do I post a video on YouTube? 2.How can I make a channel on YouTube and upload videos on it? plz help me...	

Table 3: This table compares the qualitative performance of *T-SCQA* with *LMIR*, phrase based translation model *translation(phrase)*, supervised topic model *Q-A topic model(s)* and deep semantic model without parameter sharing *T-DSQA*. For queries Q1-4 *T-SCQA* show better performance than the previous models .

pare the performance of *T-SCQA* with *T-DSQA*. *T-DSQA* retrieves the questions that are related to downloading and transferring YouTube videos to other devices. Thus, *T-DSQA* cannot clearly clarify the meaning of "put" in Q4. However, the retrieved questions of *T-SCQA* are more aligned towards the ways to record videos and upload them in YouTube. The questions retrieved by *T-SCQA* are semantically more relevant to the query Q4.

9   Conclusions

In this paper, we proposed $SCQA$ for similar question retrieval which tries to bridge the lexico-syntactic gap between the question posed by the user and the archived questions. $SCQA$ employs twin convolutional neural networks with shared parameters to learn the semantic similarity be-

tween the question and answer pairs. Interpolating $BM25$ scores into the model *T-SCQA* results in improved matching performance for both textual and semantic matching. Experiments on large scale real-life "Yahoo! Answers" dataset revealed that T-SCQA outperforms current state-of-the-art approaches based on translation models, topic models and deep neural network based models which use non-shared parameters.

As part of future work, we would like to enhance SCQA with the meta-data information like categories, user votes, ratings, user reputation of the questions and answer pairs. Also, we would like to experiment with other deep neural architectures such as Recurrent Neural Networks, Long Short Term Memory Networks, *etc.* to form the sub-networks.

References

Jane Bromley, James W Bentz, Léon Bottou, Isabelle Guyon, Yann LeCun, Cliff Moore, Eduard Säckinger, and Roopak Shah. 1993. Signature verification using a siamese time delay neural network. *IJPRAI*.

Li Cai, Guangyou Zhou, Kang Liu, and Jun Zhao. 2011. Learning the latent topics for question retrieval in community QA. *IJCNLP*.

Sumit Chopra, Raia Hadsell, and Yann LeCun. 2005. Learning a similarity metric discriminatively, with application to face verification. *CVPR*.

Arpita Das, Manish Shrivastava, and Manoj Chinnakotla. 2016. Mirror on the wall: Finding similar questions with deep structured topic modeling. Springer.

Shilin Ding, Gao Cong, Chin-Yew Lin, and Xiaoyan Zhu. 2008. Using conditional random fields to extract contexts and answers of questions from online forums. *ACL*.

Oscar Ferrandez, Christian Spurk, Milen Kouylekov, Iustin Dornescu, Sergio Ferrandez, Matteo Negri, Ruben Izquierdo, David Tomas, Constantin Orasan, Guenter Neumann, et al. 2011. The qall-me framework: A specifiable-domain multilingual question answering architecture. *Web semantics*.

Jiwoon Jeon, W. Bruce Croft, and Joon Ho Lee. 2005. Finding similar questions in large question and answer archives. *CIKM*.

Zongcheng Ji, Fei Xu, Bin Wang, and Ben He. 2012. Question-answer topic model for question retrieval in community question answering. *CIKM*.

Yann LeCun and Fu Jie Huang. 2005. Loss functions for discriminative training of energy-based models. *AISTATS*.

Baichuan Li and Irwin King. 2010. Routing questions to appropriate answerers in community question answering services. *CIKM*.

Yuanjie Liu, Shasha Li, Yunbo Cao, Chin-Yew Lin, Dingyi Han, and Yong Yu. 2008. Understanding and summarizing answers in community-based question answering services. *ICCL*.

Paul Mcnamee and James Mayfield. 2004. Character n-gram tokenization for european language text retrieval. *Information retrieval*.

Xipeng Qiu and Xuanjing Huang. 2015. Convolutional neural tensor network architecture for community-based question answering. *IJCAI*.

Stephen E Robertson, Steve Walker, Susan Jones, Micheline Hancock-Beaulieu, Mike Gatford, et al. 1994. Okapi at trec-3. *NIST Special Publication*.

Xiaobing Xue, Jiwoon Jeon, and W. Bruce Croft. 2008. Retrieval models for question and answer archives. *SIGIR*.

Chengxiang Zhai and John Lafferty. 2004. A study of smoothing methods for language models applied to information retrieval. *ACM Trans. Inf. Syst.*

Kai Zhang, Wei Wu, Haocheng Wu, Zhoujun Li, and Ming Zhou. 2014. Question retrieval with high quality answers in community question answering. *CIKM*.

Kai Zhang, Wei Wu, Fang Wang, Ming Zhou, and Zhoujun Li. 2016. Learning distributed representations of data in community question answering for question retrieval. *ICWSDM*.

Guangyou Zhou, Li Cai, Jun Zhao, and Kang Liu. 2011. Phrase-based translation model for question retrieval in community question answer archives. *ACL:HLT*.

Guangyou Zhou, Tingting He, Jun Zhao, and Po Hu. 2015. Learning continuous word embedding with metadata for question retrieval in community question answering. *ACL*.

Guangyou Zhou, Yin Zhou, Tingting He, and Wensheng Wu. 2016. Learning semantic representation with neural networks for community question answering retrieval. *Knowledge-Based Systems*.