Semantic similarity in Q&A using Deep learning techniques

Master's Thesis Project report submitted

in partial fulfilment for the award of the degree of

Masters of Technology

in

Computer Science and Engineering

by

Sandesh C

(12CS30041)

Under the supervision of

Professor Pawan Goyal

Department of Computer Science and Engineering

IIT Kharagpur

Spring Semester, 2016-17

April 28, 2017

DECLARATION

I certify that

(a) The work contained in this report has been done by me under the guidance of my supervisor.

(b) The work has not been submitted to any other Institute for any degree or diploma.

(c) I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.

(d) Whenever I have used materials (data, theoretical analysis, figures, and text) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the references. Further, I have taken permission from the copyright owners of the sources, whenever necessary.

Date: April XX, 2017 (Sandesh C)
Place: Kharagpur (12CS30041)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

IIT KHARAGPUR

KHARAGPUR - 721302, INDIA

CERTIFICATE

This is to certify that the project report entitled "**Semantic similarity in Q&A using Deep learning techniques**" submitted by **Sandesh C** (Roll No. 12CS30041) to IIT Kharagpur towards partial fulfilment of requirements for the award of degree of Masters of Technology in Computer Science and Engineering is a record of bona fide work carried out by him under my supervision and guidance during Spring Semester, 2016-17.

Date: April XX, 2017
Place: Kharagpur

Professor Pawan Goyal
Department of Computer Science and Engineering
IIT Kharagpur
Kharagpur - 721302, India

Abstract

Community Question Answering (CQA) forums have since long been plagued with the problem of answer reranking, to automate the process of finding good comments to a question. Here in this work, we take up the problem of Question-Comment similarity with a simple approach where a question-comment pair is represented as concatenation of: distributed paragraph vector representations of question text, comment text and centroidal comment (of that question), along with various syntactic and metadata features. A multilayer perceptron is used to compute the similarity scores for such a question-commnet pair.

Despite it's simplicity the model attains competitive results compared to the best submissions at SemEval '16 Task 3 - Community Question Answering (Subtask A).

Acknowledgements

First and foremost I would like to thank my guide, Prof. Pawan Goyal, for providing an opportunity to work on a challenging and relevant problem, that has blossomed to be a great learning experience. I also thank Prof. Pawan Goyal, for the constant support and guidance over the course of the project.

I would like to thank our faculty advisor, Prof. Rajat Subhra Chakraborty, for his continued efforts to help and enrich the academic experience our batch on vairous occasions. Also to acknowledge, the Department of Computer Science and Engineering, IIT Kharagpur, has been helpful through providing easy access to substantial resources for computation.

Finally I would like to thank my parents, family and friends for their continued belief and support through out the course of the programme.

Contents

List of Tables

Abbreviations

CQA	**C**ommunity **Q**uestion **A**nswering
QL	**Q**atar **L**iving
SGD	**S**tochastic **G**radient **D**escent
PV	**P**aragraph **V**ector
CBOW	**C**ontinious **B**ag-**O**f-**W**ords
DM	**D**istributed **M**emory
DBOW	**D**istributed **B**ag-**O**f-**W**ords
MAP	**M**ean **A**veraged **P**recision
MRR	**M**ean **R**eciprocal **R**ate
AvgRec	**Av**era**g**e **Rec**all
P	**P**recision
R	**R**ecall
Acc	**Acc**uraccy

Introduction

1.1 Introduction

CQA forums such as Stack Overflow[1] and Qatar Living[2], are gaining popularity online. These forums are seldom moderated, quite open, and thus they typically have little restrictions, if any, on who can post and who can answer a question. On the positive side, this means that one can freely ask any question and can then expect some good, honest comments. On the negative side, it takes effort to go through all possible comments and to make sense of them. For example, it is not unusual for a question to have hundreds of comments, which makes it very time-consuming for the user to inspect and to winnow through them all. The present work is intended to automate the process of finding good comments to questions in a community-created discussion forum, by automatically ranking the existing comments.

1.2 SemEval Task – 3

SemEval Tasks[3] (Semantic Evaluation) are an ongoing series of evaluations of computational semantic analysis systems. The **SemEval Task 3** in particular deals with semantic comparison for words and texts in the domain of Community Question

[1] https://stackoverflow.com/
[2] http://www.qatarliving.com/
[3] http://alt.qcri.org/semeval2017/

Answering (CQA). In essence, the main CQA task can be defined as follows: *"given (i) a new question and (ii) a large collection of question-comment threads created by a user community, rank the comments that are most useful for answering the new question"*.

1.2.1 Subtask A – Question-Comment Similarity

In this project we address the CQA task by exploiting the semantic similarity in Q&A using Deep learning techniques. In particular we focus on a single subtask under **SemEval - Task 3**, namely the **Subtask A**.

Subtask A *Given a question from a question-comment thread, rank the comments as per their relevance (similarity) with respect to the question.*

1.3 Thesis Organization

The thesis is further organized as follows: Chapter 2 addresses the recent works pertaining to the task of finding question-comment similarity, and also sheds light on a few such works that use Deep Learning methodologies to solve this problem; Chapter 3 then provides a detailed explanation of the multilayer perceptron based prediction model approach adopted in our work, using distributed document representations and various syntactic, metadata features. Finally, Chapter 4 tabulates the results obtained with our approach on SemEval '16 Task 3 - Subtask A dataset, which we shall see are competitive with the best results published at the same event.

Literature Survey

The tasks falling under the Community Question & Answering section of SemEval goes in the direction of passage reranking, where automatic classifiers are normally applied to pairs of questions and comment passages to derive a relative order between passages. This is in other words the task of Answer re-ranking.

In recent years, many advanced models have been developed for automating answer selection, producing a large body of work. For instance, Wang et al. (2007) proposed a probabilistic quasi synchronous grammar to learn syntactic transformations from the question to the candidate answers; Heilman and Smith (2010) used an algorithm based on Tree Edit Distance (TED) to learn tree transformations in pairs; Wang and Manning (2010) developed a probabilistic model to learn tree-edit operations on dependency parse trees; and Yao et al. (2013) applied linear chain CRFs with features derived from TED to automatically learn associations between questions and candidate answers. One interesting aspect of the above research is the need for syntactic structures; this is also corroborated in [Severyn and Moschitti (2012); Severyn and Moschitti (2013)]. Note that answer selection can use models for textual entailment, semantic similarity, and for natural language inference in general.

Although recently quite a few work in this domain have started to adopt Deep Learning Techniques to solve the problem of answer re-ranking. For eg. Lin and Wang (2015) treated the answer selection task as a sequence labeling problem and proposed recurrent convolutional neural networks to recognize good comments. In a follow-up work, Zhou et al. (2015) included long-short term memory (LSTM) units in their convolutional neural network to learn the classification sequence for the thread. In parallel, Barrón-Cedeno et al. (2015) exploited the dependencies between the thread comments to tackle the same task. This was done by designing features that look globally at the thread and by applying structured prediction models, such as Conditional Random Fields Lafferty et al. (2001).

This research direction was further extended by Joty et al. (2015), who used the output structure at the thread level in order to make more consistent global decisions. For this purpose, they modeled the relations between pairs of comments at any distance in the thread, and they combined the predictions of local classifiers in a graph-cut and in an ILP frameworks.

Noteably, at SemEval-2015 Task 3, Shafiq Joty and Nakov (2016) proposed two novel joint learning models that are on-line and integrate inference within the learning process. The first one jointly learns two node- and edge-level MaxEnt classifiers with stochastic gradient descent and integrates the inference step with loopy belief propagation. The second model is an instance of fully connected pairwise CRFs (FCCRF). The FCCRF model significantly outperformed all other approaches and yielded the best results on the task (SemEval-2015 Task 3). Crucial elements for its success were the global normalization and an Ising-like edge potential.

Thus influenced by the trend we shall tread in the direction of exploring Deep Learning Techniques to effectively solve the problem of finding Question - Comment similarity; building on the success of previous attempts. Note that we use the terms relevant-comment and answer interchangeably thoughout the document.

Approach

For this task, we adopt a neural approach to open-domain non-factoid QA developed by Bogdanova and Foster (2016), which focused on "answer re-ranking", i.e. given a list of candidate answers to a question, order the answers according to their relevance to the question. The approach is very simple and requires no feature engineering. Question-answer pairs are represented by concatenated distributed representation vectors and a multilayer perceptron is used to compute the score for an answer (the probability of an answer being the best answer to the question). Despite its simplicity, their work achieved state-of-the-art performance on the Yahoo! Answers dataset of manner or How questions introduced by Jansen et al. (2014). This improved performance was attributed to the use of paragraph vector representations instead of averaging over word vectors, and to the use of suitable data for training these representations. This project aims at improving the simplistic model proposed by Bogdanova and Foster (2016) with a few enhancements to achieve state-of-art performance at the **SemEval Task 3 - Subtask A** of finding Question – Comment similarity.

It is for this reason we use Paragraph Vectors (Le and Mikolov (2014)) for quantifying the question-comment text documents. Paragraph Vector is an unsupervised framework that learns continuous distributed vector representations for pieces of texts. The texts can be of variable-length, ranging from sentences to documents.

The name Paragraph Vector is to emphasize the fact that the method can be applied to variable-length pieces of texts, anything from a phrase or sentence to a large document.

3.1 Learning Algorithm

We used a simple feedforward neural network, i.e. a multilayered perceptron, to predict the best answer as performed by Bogdanova and Foster (2016). As shown in Figure 3.1, the first layer of network takes the vector representation for a question-comment pair *(q, c)* as input, which is a concatenation of the distributed representations q and c for the question and the comment respectively. Each representation is a real-valued vector of a fixed dimensionality d, which is a parameter to be tuned. The input layer is concatenated with another d dimensional vector, namely the centroidal comment, which is centroid of the distributed representation of all comments to the question q (subsection 3.3.1). This is further concatenated with another set of features generated from the pair (q, c) as described in section 3.3. The latter two enhancements is the reason our approach shall improve upon the performance achieved by Bogdanova and Foster (2016).

This layer is then followed by one or more hidden layers, the number of layers and units in each of these layers are also parameters to be experimentally tuned. We consider the activation function as well to be a parameter to be tuned by exprimentation. Finally, a softmax layer is used to compute the output probability p, i.e. the probabilities p1 and p2 of the negative (i.e. not best answer) and positive (i.e. best answer) classes respectively. For each question, all its user-generated comments are ranked according to their probability of being the best answer, as predicted by the network.

Given a question-comment pair (q, c), the possible values for the ground-truth label are 1 (best answer) and 0 (not a best answer). The network is trained by minimizing

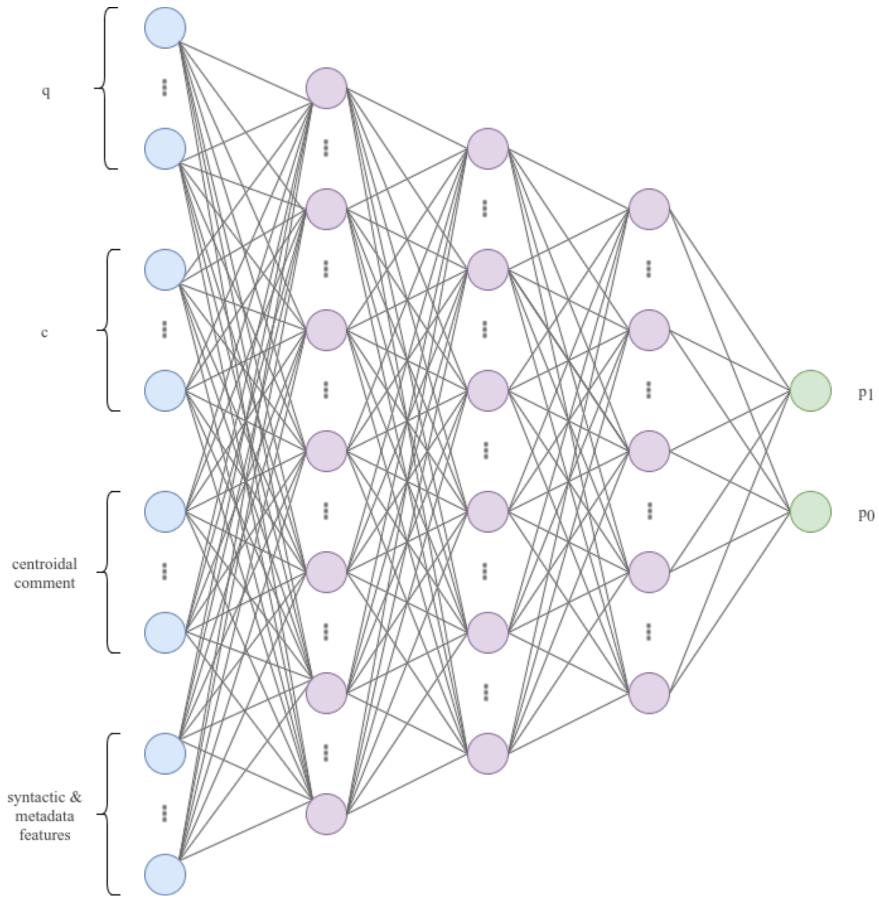

FIGURE 3.1: Architecture of proposed Feedforward Neural Network

the L2-regularized cross-entropy loss function between the ground-truth labels and the network predictions on the training set. We use either stochastic gradient descent (SGD) or Adam solver and early stopping to minimize the loss over the training set.

3.2 Document Representations

This approach requires question-comment pairs to be represented as a fixed-size vector. We experimentally evaluate the Paragraph Vector model (PV) proposed by Le and Mikolov (2014). The PV is an extension of the widely used continuous bag-of-words (CBOW) and skip-gram word embedding models, known as word2vec.

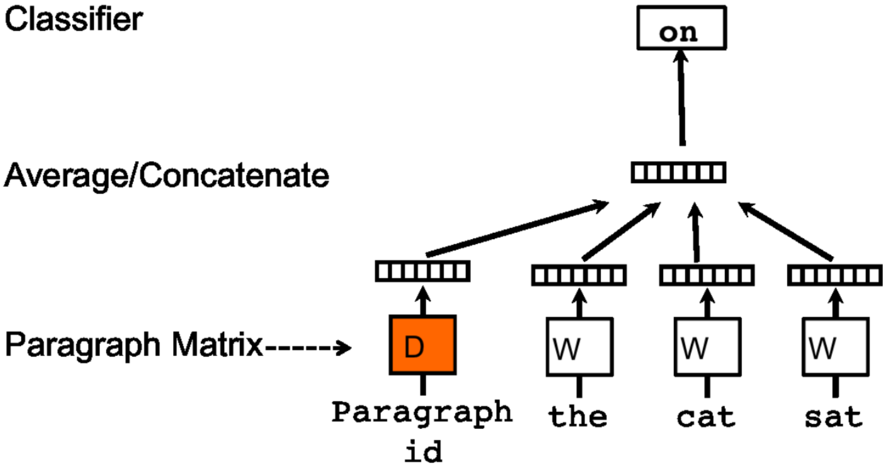

FIGURE 3.2: Distributed Memory (DM) framework for learning paragraph vector. In this model, the concatenation or average of word vectors with a context of few words is used to predict the next word. The paragraph vector represents the missing information from the current context and can act as a memory of the topic of the paragraph.

However, in contrast to CBOW and skip-gram models that only learn word embeddings, the PV is able to learn representations for pieces of text of arbitrary length, e.g. sentences, paragraphs or documents. The types of PV include (1) the distributed memory (DM) model, that predicts the next word in a text window using the concatenation of the word vectors of previous words and the paragraph vector; (2) the distributed bag-of-words (DBOW) model, that – similar to the skip-gram model – predicts words (in a small window) randomly sampled from the paragraph, given the paragraph vector. We experiment with both DM and DBOW models. Figure 3.2 and Figure 3.3 provide an illustration for these paragraph vector models. Also, note that we shall use the terms paragraph vector (PV) and document vector/representation interchangeably.

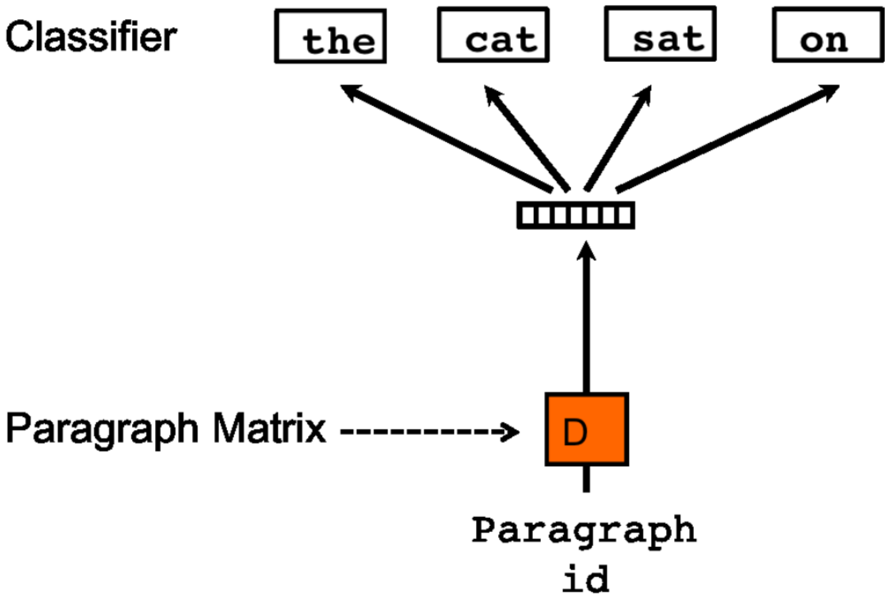

FIGURE 3.3: Distributed Bag of Words (DBOW) version of paragraph vectors. The paragraph vector is trained to predict the words in a small window

3.3 Feature Set

Apart from paragraph vectors of the Question (q) and Comment (c) that the feed-forward network takes as input, we describe below the surplus features incorporated in our model:

3.3.1 Centroidal Comment

Inorder to rank the comments, it is only intuition that we must use the information in other comment texts to accurately provide relative relevance scores, which in turn reflects the rank, for comment texts. It is for this reason we introduced the centroidal comment, denoted as avg_com_q, computed as:

$$avg\_com_q = \frac{\sum\limits_{c \in q} c}{||\sum\limits_{c \in q} c||} \tag{1}$$

3.3.2 Syntactic and Metadata Features

We used several semantic vector similarity and metadata feature groups as mentioned in Mihaylov and Nakov (2016). For the ease of the reader, we shall describe the same feature groups below.

Note that for the similarity measures mentioned below, we used cosine similarity:

$$1 - \frac{u.v}{||u||.||v||} \tag{2}$$

Semantic Word Embeddings. We used semantic word embeddings obtained from Word2Vec models trained on the unannotated data set from QatarLiving. For each piece of text such as comment text, question body and question subject, we constructed the centroid vector from the vectors of all words in that text (excluding stopwords).

$$centroid(w_{1...n}) = \frac{\sum_{i=1}^{n} w_i}{n} \tag{3}$$

We construct centroid vectors (3) from the question text (subject + body) and the comment text to design various features as described below.

Semantic Vector Similarities. We used various similarity features calculated using the centroid word vectors on the question text (subject + body) and on the comment text, as well as on parts thereof:

Question to Answer similarity. We assume that a relevant answer should have a centroid vector that is close to that for the question. We used the question text to comment text vector similarities.

Maximized similarity. We ranked each word in the comment text to the question text centroid vector according to their similarity and we took the average similarity of the top N words. We took the top 1, 2, 3, 4 and 5 words similarities as features.

The assumption here is that if the average similarity for the top N most similar words is high, then the comment might be relevant.

Aligned similarity. For each word in the question text, we chose the most similar word from the comment text and we took the average of all best word pair similarities as suggested in Tran et al. (2015).

Part of speech (POS) based word vector similarities. We performed part of speech tagging using the Stanford tagger Toutanova et al. (2003), and we took similarities between centroid vectors of words with a specific tag from the comment text and the centroid vector of the words with a specific tag from the question text. The assumption is that some parts of speech between the question and the comment might be closer than other parts of speech.

Word clusters (WC) similarity. We clustered the word vectors from the Word2Vec vocabulary in 1,000 clusters using K-Means clustering. We then calculated the cluster similarity between the question body word clusters and the answer text word clusters. For all experiments, we used clusters obtained from the Word2Vec model trained on QatarLiving forums with vector size of 100, window size 10.

LDA topic similarity. We performed topic clustering using Latent Dirichlet Allocation (LDA) as implemented in the gensim toolkit Rehurek and Sojka (2010) on Train1 + Train2 + Dev questions and comments. We built topic models with 100 topics. For each word in the question text and for the comment text, we built a bag-of-topics with corresponding distribution, and calculated similarity. The assumption here is that if the question and the comment share similar topics, they are more likely to be relevant to each other

Paragraph Vector similarities. The similarity among the distributed vector representations of question text (q), answer text (a) and the centroidal answer (avg_com_q), taken two at a time are also included.

Metadata. In addition to the semantic features described above, we also used some common sense metadata features:

Answer contains a question mark. If the comment has an question mark, it may be another question, which might indicate a bad answer.

Answer length. Assumption here is that longer answers could bring useful details.

Question length. If the question is longer, it may be more clear, which may help users give a more relevant answer.

Question to comment length. If the question is long and the answer is short, it may be less relevant.

The answer's author is the same as the corresponding question's author. If the answer is posted by the same user who posted the question and it is relevant, why has he/she asked the question in the first place?

Answer rank in the thread. Earlier answers could be posted by users who visit the forum more often, and they may have read more similar questions and answers. Moreover, discussion in the forum tends to diverge from the question over time.

Question category. We took the category of the question as a sparse binary feature vector (a feature with a value of 1 appears if question is in the category). The assumption here is that the question-comment relevance might depend on the category of the question.

Comments by the same User. The number of comments by the author of a given comment to the same question and the order of the comments (first, second, ...) is also included as a feature. If the author produced an incomplete answer in the first attempt, he/she might be obliged to produce another comment subsequently.

Time difference between Question and Comment posting. Immediate comments could reflect incomplete answers to longer questions, while comments posted after substantial time might reflect well-thought answers.

Experiments

4.1 Data

Though Bogdanova and Foster (2016) experiments with the *Yahoo! Answers dataset*[4], we have used the data provided as a part of the popular **SemEval Task 3** for **Subtask A** (Nakov et al. (2016)). Table 1 contains the statistics about the forementioned dataset. This dataset contains about 42K (q, c) pairs to learn from; spreading over about 5.4K questions. We shall refer to this data as the CQA-QL corpus in future. Further we also use a large unannotated dataset, released by the same source, from Qatar Living with 189,941 questions and 1,894,456 comments, which is used for unsupervised learning/training domain-specific word/document embeddings.

Category	**Train (Part-I)**	**Train (Part-II)**	**Train+Dev+Test (from SemEval 2015)**	**Dev**	**Test**	**Total**
Questions	**1,411**	**379**	**2,480+291+319**	**244**	**327**	**5,451**
Comments	**14,110**	**3,790**	**14,893+1,529+1,876**	**2,440**	**3,270**	**41,908**
-Good	*5,287*	*1,364*	*7,418+813+946*	*818*	*1,329*	*17,975*
-Bad	*6,362*	*1,777*	*5,971+544+774*	*1,209*	*1,485*	*18,122*
-Potentially	*2,461*	*649*	*1,504+172+156*	*413*	*456*	*5,811*

TABLE 1: Statistics on English CQA-QL corpus from SemEval-2017 Task 3 (Subtask A)

[4] http://webscope.sandbox.yahoo.com/

4.2 Experimental Setup

We use the gensim[5] implementation of DM and DBOW paragraph vector models. The data for training the unsupervised *doc2vec* model (PV model) is the forementioned large unannotated dataset from Qatar Living forums. Each paragraph (q or c) was converted to lowercase, tokenized by space character and cleaned of stop words before training *doc2vec* models. The parameters of training these models being the window size (maximum distance between the predicted word and context words used for prediction within a document) and number of epochs of training, were cross-validated to give low errors on the training dataset. We further use normalized versions of the document vector representations thus generated, to be fed as inputs to the feedforward neural network described in section 3.1.

For the implementation of the feedforward neural network as described in section 3.1, we shall use the popular python library *scikit-learn*[6]'s *MLPClassifier*[7].

4.3 Results

4.3.1 Document Vector Representations

For training each question/comment text was treated as a document/paragraph and assigned a label, which can be used as a key to retrieve the document vector. Furthermore post training the doc2vec model is able to infer a document vector for any new question/comment text whose vocabulary is from the original corpus. The errors post training is computed as averaged squared error over all question/comment text, by computing squared error between the document vector learnt by the model corresponding to the text's label and the document vector inferred from the

[5]https://radimrehurek.com/gensim/models/doc2vec.html
[6]http://scikit-learn.org/stable/index.html
[7]http://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html

question/comment text. The squared errors are computed for both normalized and unnormalized document vectors. For comparison purposes normalized and unnormalized squared error between any two random document vector is tabulated beside these errors (averaged over as many iterations as the number of question/comment text). Experiments show that 100-dimensional PV trained over the ∼2.3M samples from the unannotated QL corpus, gives sufficiently low errors post normalization. Further more, PV-DBOW prove to outperform the PV-DM representations as seen in Table 2. It contains few of the best results has rows sorted by the value of *column 'Ratio'*, as it is the indicator of how good the representation is. The complete list of experiments is tabulated under Appendix A.

Category	**Window Size**	**Epochs**	**Squared Error**	**Normalized Sq. Error (A)**	**Sq. Error (Random)**	**Norm. Sq. Error (Random) (B)**	**Ratio (B/A)**
PV-DBOW	10	5	10.79	0.12	0.56	0.80	6.74
PV-DBOW	10	10	13.16	0.12	0.61	0.82	6.61
PV-DM	10	5	0.66	0.21	0.98	0.99	4.67
PV-DM	15	10	0.93	0.22	0.98	0.98	4.47

Table 2: Training document vector representations PV-DM and PV-DBOW – Best results

4.3.2 SemEval Task 3 – Subtask A

The training data comprises of 38,638 comments spanning over 5,124 questions. The neural net input is a tuple of the form $(q, c, avg\_ans_q, ft_{(q,c)})$, where,

avg_com_q is (normalized) average over the PV of all comments to question q

$ft_{(q,c)}$ is feature vector corresp. to the pair (q,c) as described in section 3.3

SemEval Task 3 has as an official evaluation measure used to rank the participating systems, the metric of Mean Average Precision (*MAP*), calculated for the ten comments a participating system has ranked highest. Further metrics such as Mean Reciprocal Rank (*MRR*) and Average Recall (*AvgRec*) for top-10 results; Precision (*P*), Recall (*R*), *F₁* (with respect to the Good/Relevant class) and Accuracy (*Acc*) are also reported.

4.3.2.1 Preliminary experiments with (q, c) inputs

Intially experiments were conducted with only (q, c) pair as input to the neural nets. The nets were trained using multiple solvers, activation functions, hidden layer configurations. The best performance for each parameter configuration is as tabulated in Table 3, while the complete results are tabulated in Appendix B.

Category	Solver	Activation	MAP	AvgRec	MRR	P	R	F_1	Acc
PV-DBOW	Adam	logistic	0.7049	0.8292	77.62	0.6601	0.5508	0.6005	0.7021
PV-DBOW	SGD	relu	0.7019	0.8251	77.16	0.6327	0.5937	0.6126	0.6948
PV-DBOW	SGD	logistic	0.7018	0.8242	77.32	0.6412	0.5877	0.6133	0.6988
PV-DBOW	SGD	tanh	0.7009	0.8245	76.62	0.6339	0.5914	0.6119	0.6951
PV-DBOW	Adam	relu	0.6993	0.8106	76.94	0.5998	0.5741	0.5867	0.6713
PV-DBOW	Adam	tanh	0.698	0.8231	76.35	0.6386	0.5546	0.5936	0.6914
PV-DM	SGD	relu	0.6578	0.7855	74.58	0.5793	0.5357	0.5567	0.6532

TABLE 3: Preliminary experiments using only (q, c) inputs – Best results.

PV-DBOW clearly outperforms PV-DM representations in these preliminary runs. Building on this, further experiments where conducted using only the PV-DBOW representations.

4.3.2.2 Improvement with inclusion of Centroidal comment

As described in subsection 3.3.1, additionally, to capture the relative goodness of an comment with respect to other comments to the same question, avg_com_q (normalized post averaging over the PV of all comments to question q) was fed as an input to the neural net. The best results for these experiments are tabulated in Table 4. Complete results are tabulated under Appendix C.

Category	Solver	Activation	MAP	AvgRec	MRR	P	R	F_1	Acc
PV-DBOW	SGD	relu	0.7306	0.8416	79.61	0.6607	0.5786	0.6169	0.708
PV-DBOW	SGD	tanh	0.7188	0.8343	79.11	0.6684	0.5658	0.6129	0.7095
PV-DBOW	SGD	logistic	0.7179	0.8346	79.13	0.6652	0.553	0.6039	0.7052
PV-DBOW	Adam	logistic	0.7177	0.8342	78.96	0.6635	0.5726	0.6147	0.7083
PV-DBOW	Adam	tanh	0.7169	0.8339	78.64	0.655	0.5771	0.6136	0.7046
PV-DBOW	Adam	relu	0.7161	0.8263	79.45	0.6298	0.6005	0.6148	0.6942

TABLE 4: Experiments using (q, c, avg_com_q) inputs – Best results.

Clearly there is a significant improvement in MAP scores after inclusion of the centroidal comment for each question as an input feature. Further experiments thus is done inclusive of avg_com_q in the input tuple.

4.3.2.3 Further improvement with Syntactic and Metadata Features

Category	**Solver**	**Activation**	**MAP**	**AvgRec**	**MRR**	**P**	**R**	$\mathbf{F_1}$	**Acc**

TABLE 5: Experiments using $(q, c, avg\_com_q, ft_{(q,c)})$ inputs – Best results.

Conclusions

Appendix A

Training PV-DM and PV-DBOW

Category	Window Size	Epochs	Squared Error	Normalized Sq. Error (A)	Sq. Error (Random)	Norm. Sq. Error (Random) (B)	Ratio (B/A)
PV-DBOW	10	5	10.791	0.118	0.560	0.799	6.738
PV-DBOW	10	10	13.159	0.124	0.606	0.821	6.614
PV-DBOW	15	5	10.790	0.127	0.564	0.796	6.263
PV-DBOW	15	10	13.077	0.132	0.607	0.818	6.193
PV-DBOW	20	5	10.749	0.134	0.569	0.794	5.910
PV-DBOW	20	10	12.932	0.140	0.611	0.816	5.822
PV-DM	10	5	0.664	0.211	0.985	0.987	4.671
PV-DM	15	10	0.929	0.219	0.981	0.983	4.472
PV-DM	15	5	0.670	0.229	0.984	0.984	4.291
PV-DM	20	10	0.830	0.229	0.982	0.982	4.284
PV-DM	25	10	0.780	0.235	0.983	0.982	4.168
PV-DM	20	5	0.706	0.239	0.984	0.981	4.092
PV-DM	15	20	1.597	0.241	0.970	0.976	4.034
PV-DM	25	5	0.735	0.248	0.984	0.980	3.951
PV-DM	20	20	1.434	0.247	0.974	0.977	3.947
PV-DM	25	20	1.339	0.251	0.976	0.978	3.890
PV-DM	15	30	2.062	0.261	0.954	0.963	3.687
PV-DM	20	30	1.867	0.266	0.962	0.968	3.637
PV-DM	25	30	1.741	0.270	0.966	0.971	3.595
PV-DM	10	1	0.964	0.333	0.988	0.979	2.936
PV-DM	15	1	0.975	0.357	0.988	0.971	2.717
PV-DM	20	1	0.979	0.374	0.988	0.969	2.592
PV-DM	25	1	0.981	0.384	0.988	0.967	2.515

Table 6: Training document vector representations PV-DM and PV-DBOW – All results

Appendix B

Preliminary experiments

Category	Solver	Activation	Hidden Layer	MAP	AvgRec	MRR	P	R	F_1	Acc	Runtime(sec)
PV-DBOW	Adam	logistic	[500, 250, -]	0.7049	0.8292	77.62	0.6601	0.5508	0.6005	0.7021	286.79
PV-DBOW	SGD	relu	[100, -, -]	0.7019	0.8251	77.16	0.6327	0.5937	0.6126	0.6948	129.22
PV-DBOW	SGD	logistic	[50, -, -]	0.7018	0.8242	77.32	0.6412	0.5877	0.6133	0.6988	157.93
PV-DBOW	SGD	logistic	[100, -, -]	0.7013	0.8242	77.23	0.6353	0.5884	0.6109	0.6954	185.49
PV-DBOW	SGD	tanh	[500, 100, 100]	0.7009	0.8245	76.62	0.6339	0.5914	0.6119	0.6951	165.03
PV-DBOW	Adam	logistic	[50, 50, 50]	0.7008	0.8252	76.8	0.6043	0.6584	0.6302	0.6859	97.65
PV-DBOW	SGD	relu	[50, -, -]	0.7007	0.824	77.01	0.6374	0.5899	0.6127	0.6969	118.02
PV-DBOW	SGD	tanh	[250, 100, -]	0.7007	0.8237	76.86	0.631	0.5854	0.6073	0.6924	155.15
PV-DBOW	Adam	logistic	[100, 50, 50]	0.7007	0.8251	76.89	0.6325	0.5997	0.6157	0.6957	102.81
PV-DBOW	SGD	logistic	[250, -, -]	0.7006	0.8236	77.36	0.6396	0.5862	0.6117	0.6976	291.97
PV-DBOW	Adam	logistic	[100, 100, 100]	0.7006	0.825	76.9	0.6404	0.5749	0.6059	0.696	103.54
PV-DBOW	SGD	logistic	[500, -, -]	0.7005	0.824	77.2	0.6369	0.5847	0.6097	0.6957	404.83
PV-DBOW	SGD	tanh	[250, 50, 50]	0.7004	0.8241	76.89	0.6339	0.5914	0.6119	0.6951	139.8
PV-DBOW	Adam	logistic	[250, 250, 250]	0.7003	0.8248	76.97	0.6217	0.6245	0.6231	0.693	114.48
PV-DBOW	Adam	logistic	[250, 250, 50]	0.7	0.8248	76.94	0.6205	0.62	0.6202	0.6914	114.36
PV-DBOW	Adam	logistic	[500, 100, -]	0.6999	0.8244	76.92	0.6281	0.6087	0.6183	0.6945	109.9
PV-DBOW	Adam	logistic	[250, 100, -]	0.6998	0.824	76.7	0.6164	0.6275	0.6219	0.6899	104.77
PV-DBOW	Adam	logistic	[250, 50, -]	0.6998	0.8239	76.66	0.6194	0.6381	0.6286	0.6936	108.28
PV-DBOW	Adam	logistic	[500, 50, -]	0.6997	0.8243	76.75	0.6346	0.605	0.6194	0.6979	121.15
PV-DBOW	Adam	logistic	[250, 250, 100]	0.6997	0.8243	76.84	0.6343	0.5899	0.6113	0.6951	120.99
PV-DBOW	SGD	tanh	[250, 100, 50]	0.6996	0.8235	76.66	0.6334	0.5877	0.6097	0.6942	138.22
PV-DBOW	SGD	tanh	[100, 100, 50]	0.6996	0.8237	76.99	0.637	0.5877	0.6114	0.6963	127.14
PV-DBOW	SGD	tanh	[100, -, -]	0.6996	0.823	76.67	0.634	0.5839	0.6079	0.6939	118.76
PV-DBOW	Adam	logistic	[500, 250, 50]	0.6995	0.8238	76.69	0.6447	0.5749	0.6078	0.6985	123.39
PV-DBOW	SGD	tanh	[500, 500, 250]	0.6994	0.8241	76.76	0.6339	0.5914	0.6119	0.6951	279.06
PV-DBOW	SGD	tanh	[500, 500, 100]	0.6994	0.8235	76.89	0.6347	0.5884	0.6107	0.6951	243.03
PV-DBOW	SGD	tanh	[500, 500, -]	0.6994	0.8238	76.43	0.6347	0.5884	0.6107	0.6951	259.35
PV-DBOW	Adam	logistic	[100, 100, 50]	0.6994	0.824	76.51	0.6406	0.5794	0.6085	0.6969	106.14
PV-DBOW	Adam	logistic	[100, 50, -]	0.6994	0.8243	76.74	0.6279	0.6095	0.6186	0.6945	100.68
PV-DBOW	Adam	relu	[500, 50, -]	0.6993	0.8106	76.94	0.5998	0.5741	0.5867	0.6713	139.2
PV-DBOW	SGD	tanh	[500, 500, 50]	0.6992	0.8236	76.5	0.6344	0.5914	0.6121	0.6954	235.19
PV-DBOW	SGD	tanh	[250, 50, -]	0.6992	0.8225	76.76	0.633	0.5892	0.6103	0.6942	140.22
PV-DBOW	SGD	tanh	[100, 100, -]	0.6992	0.8241	76.74	0.6349	0.5862	0.6095	0.6948	129.93
PV-DBOW	Adam	logistic	[500, 500, 500]	0.6992	0.8246	76.77	0.6098	0.6584	0.6331	0.6899	168.88
PV-DBOW	Adam	logistic	[250, 100, 50]	0.6992	0.8238	76.74	0.6224	0.6275	0.625	0.6939	115.71
PV-DBOW	Adam	logistic	[50, 50, -]	0.6992	0.8241	76.87	0.6333	0.6042	0.6184	0.6969	96.84
PV-DBOW	SGD	tanh	[50, 50, -]	0.699	0.8236	76.77	0.6373	0.5884	0.6119	0.6966	114.33
PV-DBOW	SGD	tanh	[50, -, -]	0.699	0.8235	76.95	0.6407	0.5877	0.613	0.6985	114.53
PV-DBOW	SGD	relu	[500, -, -]	0.6989	0.824	76.58	0.6377	0.5907	0.6133	0.6972	195.01
PV-DBOW	SGD	tanh	[500, 100, -]	0.6989	0.8231	76.79	0.6338	0.5862	0.6091	0.6942	173.54

TABLE 7: Preliminary experiments using only (q, c) inputs – All results.

Category	Solver	Activation	Hidden Layer	MAP	AvgRec	MRR	P	R	F_1	Acc	Runtime(sec)
PV-DBOW	SGD	tanh	[250, -, -]	0.6987	0.8238	76.94	0.6321	0.5869	0.6087	0.6933	147.34
PV-DBOW	SGD	tanh	[100, 100, 100]	0.6987	0.8235	76.59	0.6365	0.5877	0.6111	0.696	131.55
PV-DBOW	SGD	tanh	[500, 100, 50]	0.6986	0.8233	76.5	0.6365	0.5929	0.6139	0.6969	177.37
PV-DBOW	Adam	logistic	[100, 100, -]	0.6986	0.8235	76.28	0.6192	0.6253	0.6222	0.6914	101.25
PV-DBOW	SGD	relu	[250, -, -]	0.6985	0.824	76.55	0.6369	0.5899	0.6125	0.6966	150.09
PV-DBOW	Adam	logistic	[100, -, -]	0.6984	0.8232	76.54	0.643	0.5666	0.6024	0.696	99.73
PV-DBOW	SGD	tanh	[500, 250, 100]	0.6983	0.8233	76.2	0.6356	0.5907	0.6123	0.696	210.44
PV-DBOW	SGD	tanh	[500, 250, 50]	0.6983	0.8229	76.52	0.635	0.5929	0.6132	0.696	192.64
PV-DBOW	SGD	tanh	[500, 250, -]	0.6983	0.8235	76.68	0.6339	0.5877	0.6099	0.6945	199.07
PV-DBOW	Adam	logistic	[500, -, -]	0.6983	0.823	76.51	0.5983	0.6892	0.6406	0.6856	114.23
PV-DBOW	Adam	logistic	[50, -, -]	0.6983	0.8233	76.55	0.6385	0.5847	0.6104	0.6966	93.79
PV-DBOW	SGD	tanh	[500, 250, 250]	0.6982	0.8234	76.61	0.6341	0.5869	0.6096	0.6945	216.67
PV-DBOW	SGD	tanh	[500, 50, 50]	0.6982	0.8232	76.63	0.6309	0.5892	0.6093	0.693	163.63
PV-DBOW	SGD	tanh	[500, -, -]	0.6982	0.8235	76.68	0.6339	0.5824	0.6071	0.6936	185.16
PV-DBOW	Adam	logistic	[500, 500, 50]	0.6982	0.8211	76.28	0.6311	0.5869	0.6082	0.6927	210.76
PV-DBOW	Adam	logistic	[250, 250, -]	0.6981	0.8229	76.51	0.6149	0.6441	0.6292	0.6914	111.81
PV-DBOW	Adam	tanh	[100, -, -]	0.698	0.8231	76.35	0.6386	0.5546	0.5936	0.6914	94.23
PV-DBOW	Adam	tanh	[50, -, -]	0.698	0.8232	76.64	0.6264	0.617	0.6217	0.6948	94.3
PV-DBOW	SGD	tanh	[500, 50, -]	0.6979	0.8229	76.37	0.6338	0.5899	0.6111	0.6948	179.26
PV-DBOW	SGD	tanh	[100, 50, 50]	0.6979	0.8239	76.52	0.6339	0.5877	0.6099	0.6945	120.28
PV-DBOW	SGD	tanh	[250, 250, 50]	0.6975	0.8222	76.47	0.6316	0.5869	0.6084	0.693	153.34
PV-DBOW	SGD	tanh	[50, 50, 50]	0.6975	0.8224	76.66	0.6369	0.5899	0.6125	0.6966	114.63
PV-DBOW	SGD	tanh	[250, 250, 100]	0.6974	0.8232	76.6	0.6323	0.5899	0.6104	0.6939	159.4
PV-DBOW	SGD	tanh	[250, 250, -]	0.6974	0.8224	76.32	0.633	0.5854	0.6083	0.6936	158.46
PV-DBOW	SGD	tanh	[250, 100, 100]	0.6974	0.8228	76.39	0.6322	0.5884	0.6095	0.6936	141.7
PV-DBOW	Adam	tanh	[500, -, -]	0.6974	0.8228	76.48	0.6274	0.6245	0.6259	0.6966	116.24
PV-DBOW	SGD	tanh	[100, 50, -]	0.6973	0.8231	76.44	0.6356	0.5892	0.6115	0.6957	122.56
PV-DBOW	SGD	tanh	[500, 500, 500]	0.6971	0.8221	76.48	0.6331	0.5907	0.6111	0.6945	340.49
PV-DBOW	Adam	logistic	[250, -, -]	0.6966	0.822	76.19	0.647	0.553	0.5963	0.6957	106.11
PV-DBOW	Adam	tanh	[250, -, -]	0.6964	0.8218	76.23	0.6331	0.5869	0.6091	0.6939	110.01
PV-DBOW	SGD	tanh	[250, 250, 250]	0.6962	0.8214	76.16	0.6335	0.5892	0.6105	0.6945	179.46
PV-DBOW	Adam	logistic	[500, 250, 250]	0.6959	0.8205	76.23	0.6296	0.6087	0.619	0.6954	201.92
PV-DBOW	Adam	relu	[500, 500, 500]	0.6953	0.8134	75.64	0.5863	0.6185	0.602	0.6676	220.9
PV-DBOW	Adam	logistic	[500, 250, 100]	0.6952	0.8195	75.94	0.6439	0.5538	0.5955	0.6942	187
PV-DBOW	Adam	logistic	[500, 500, 100]	0.6948	0.8196	76.16	0.6254	0.6193	0.6223	0.6945	216.42
PV-DBOW	Adam	logistic	[500, 500, -]	0.6948	0.8206	75.89	0.6493	0.5433	0.5916	0.6951	245.12
PV-DBOW	Adam	logistic	[500, 100, 100]	0.6944	0.82	75.84	0.6271	0.605	0.6159	0.6933	169.33
PV-DBOW	Adam	logistic	[500, 100, 50]	0.6938	0.8192	76	0.6271	0.6035	0.615	0.693	157.06
PV-DBOW	Adam	logistic	[250, 100, 100]	0.6931	0.8187	75.65	0.6536	0.5252	0.5824	0.6939	134.58
PV-DBOW	Adam	logistic	[250, 50, 50]	0.693	0.8196	75.76	0.6325	0.5997	0.6157	0.6957	132.37
PV-DBOW	Adam	logistic	[500, 500, 250]	0.6927	0.8189	75.7	0.5982	0.6855	0.6388	0.685	267.39
PV-DBOW	Adam	logistic	[500, 50, 50]	0.6919	0.8187	75.65	0.6285	0.6035	0.6157	0.6939	156.24
PV-DBOW	SGD	relu	[500, 500, 100]	0.6898	0.8095	76.17	0.5901	0.5839	0.587	0.6661	1081.37
PV-DBOW	Adam	relu	[500, 500, 100]	0.689	0.8085	76.14	0.5895	0.5726	0.5809	0.6642	204.21
PV-DBOW	Adam	relu	[500, 250, 50]	0.6889	0.8052	77.5	0.5756	0.5613	0.5684	0.6535	146.95
PV-DBOW	Adam	relu	[500, 100, -]	0.688	0.8094	75.7	0.5912	0.5636	0.577	0.6642	140.24
PV-DBOW	Adam	relu	[500, 250, 250]	0.6872	0.81	75.47	0.5879	0.6193	0.6032	0.6688	154.9
PV-DBOW	Adam	relu	[500, -, -]	0.6871	0.8081	76.88	0.5729	0.5824	0.5776	0.6538	195.71
PV-DBOW	SGD	relu	[500, 100, 100]	0.6861	0.8071	77.62	0.577	0.5666	0.5718	0.655	572
PV-DBOW	Adam	relu	[250, -, -]	0.6861	0.8053	75.89	0.5904	0.5553	0.5723	0.6627	178.96
PV-DBOW	SGD	relu	[500, 250, -]	0.6855	0.8081	76.23	0.585	0.5515	0.5678	0.6587	923.78
PV-DBOW	SGD	relu	[500, 100, -]	0.6853	0.8042	77.34	0.5876	0.5553	0.571	0.6609	768.49
PV-DBOW	SGD	relu	[500, 500, -]	0.6845	0.8052	75.42	0.587	0.5636	0.575	0.6615	1367.84
PV-DBOW	SGD	relu	[500, 250, 100]	0.6828	0.8012	75.33	0.5753	0.5719	0.5736	0.6544	765.83
PV-DBOW	SGD	relu	[500, 50, -]	0.6823	0.8055	75.46	0.5771	0.5719	0.5745	0.6557	681.09
PV-DBOW	Adam	relu	[250, 250, 100]	0.6823	0.8151	76.37	0.6077	0.5094	0.5542	0.667	132.59
PV-DBOW	Adam	relu	[250, 250, 250]	0.6816	0.8073	76.21	0.6163	0.4786	0.5388	0.667	138.96
PV-DBOW	SGD	relu	[500, 250, 250]	0.6805	0.8013	75.65	0.5747	0.5704	0.5725	0.6538	853.08
PV-DBOW	Adam	relu	[500, 500, 250]	0.68	0.8035	76.37	0.6021	0.5613	0.581	0.6709	207.74
PV-DBOW	Adam	relu	[500, 250, 100]	0.6796	0.7963	75.73	0.5829	0.611	0.5966	0.6642	147.82
PV-DBOW	Adam	relu	[500, 100, 50]	0.6795	0.8025	75.87	0.5899	0.5899	0.5899	0.6667	167.79
PV-DBOW	SGD	relu	[500, 250, 50]	0.679	0.8	74.37	0.5847	0.5636	0.5739	0.6599	769.21
PV-DBOW	Adam	relu	[50, -, -]	0.6789	0.802	74.61	0.5746	0.5388	0.5561	0.6505	150.11
PV-DBOW	Adam	relu	[500, 250, -]	0.6784	0.8015	74.97	0.6037	0.5388	0.5694	0.6688	147.97
PV-DBOW	Adam	relu	[250, 250, 50]	0.6784	0.8029	74.02	0.5892	0.5192	0.552	0.6575	129.22
PV-DBOW	SGD	relu	[500, 100, 50]	0.678	0.8027	75.18	0.5828	0.5771	0.58	0.6602	594.9
PV-DBOW	Adam	relu	[250, 50, -]	0.678	0.8043	74.84	0.5899	0.5335	0.5603	0.6596	126.64
PV-DBOW	Adam	relu	[500, 500, 50]	0.6779	0.7993	75.16	0.6069	0.5192	0.5596	0.6679	193.78
PV-DBOW	Adam	relu	[250, 50, 50]	0.6777	0.8028	74.79	0.6144	0.5011	0.552	0.6694	116.88
PV-DBOW	Adam	relu	[100, -, -]	0.6764	0.7963	75.58	0.5776	0.5628	0.5701	0.655	254.68
PV-DBOW	Adam	relu	[250, 100, 50]	0.6759	0.8006	75.18	0.5827	0.5222	0.5508	0.6538	116.39
PV-DBOW	SGD	relu	[500, 500, 500]	0.6757	0.7998	74.65	0.582	0.5658	0.5738	0.6584	1069.71
PV-DBOW	SGD	relu	[500, 500, 50]	0.675	0.8002	75.24	0.5778	0.5643	0.571	0.6554	1066.17
PV-DBOW	Adam	relu	[250, 100, 100]	0.6749	0.8009	75.43	0.5714	0.626	0.5975	0.6572	120.57
PV-DBOW	Adam	relu	[500, 50, 50]	0.6745	0.798	73.88	0.5675	0.6328	0.5984	0.6547	128.97
PV-DBOW	Adam	relu	[100, 100, -]	0.6736	0.7985	73.98	0.5734	0.5816	0.5775	0.6541	130.35
PV-DBOW	SGD	relu	[500, 50, 50]	0.6735	0.7994	75.36	0.5847	0.5847	0.5847	0.6624	523.52

Preliminary experiments using only (q, c) inputs – All results.

Category	**Solver**	**Activation**	**Hidden Layer**	**MAP**	**AvgRec**	**MRR**	**P**	**R**	**F_1**	**Acc**	**Runtime(sec)**
PV-DBOW	SGD	relu	[50, 50, -]	0.6731	0.7978	74.39	0.5764	0.5621	0.5691	0.6541	259.25
PV-DBOW	SGD	relu	[250, 250, 50]	0.6726	0.7935	73.95	0.5671	0.5598	0.5634	0.6474	602.45
PV-DBOW	Adam	tanh	[500, 500, 500]	0.6724	0.7953	75.14	0.5931	0.5824	0.5877	0.6679	547.86
PV-DBOW	SGD	relu	[500, 500, 250]	0.6723	0.7955	73.58	0.5784	0.5636	0.5709	0.6557	1322
PV-DBOW	SGD	relu	[250, 250, 100]	0.6721	0.7967	75.01	0.5756	0.5643	0.5699	0.6538	654.96
PV-DBOW	SGD	relu	[250, 100, -]	0.6707	0.7989	74.35	0.5655	0.5327	0.5486	0.6437	552.36
PV-DBOW	Adam	relu	[500, 500, -]	0.6707	0.7923	73.13	0.5747	0.5786	0.5767	0.6547	148.62
PV-DBOW	SGD	relu	[50, 50, 50]	0.67	0.8021	74.36	0.5796	0.5997	0.5895	0.6606	263.43
PV-DBOW	SGD	relu	[250, 250, -]	0.6693	0.7942	74.6	0.578	0.5523	0.5648	0.6541	741.41
PV-DBOW	SGD	relu	[250, 100, 50]	0.6685	0.7909	73.52	0.5592	0.547	0.5531	0.6407	489.11
PV-DBOW	Adam	relu	[250, 250, -]	0.6684	0.8003	73.9	0.5709	0.5756	0.5732	0.6517	121.24
PV-DBOW	SGD	relu	[250, 50, 50]	0.668	0.7915	73.88	0.5735	0.5666	0.57	0.6526	453.22
PV-DBOW	Adam	relu	[250, 100, -]	0.6673	0.7936	73.37	0.5725	0.5764	0.5744	0.6529	119.39
PV-DBOW	Adam	tanh	[500, 500, 250]	0.6658	0.7878	73.56	0.5874	0.5764	0.5818	0.6633	418.31
PV-DBOW	Adam	tanh	[500, 250, 100]	0.6652	0.7888	72.88	0.5847	0.5508	0.5672	0.6584	269.1
PV-DBOW	Adam	relu	[500, 100, 100]	0.6651	0.7938	72.41	0.5822	0.6102	0.5959	0.6636	135.75
PV-DBOW	Adam	relu	[50, 50, -]	0.6643	0.7886	73.6	0.5562	0.5583	0.5573	0.6394	160.04
PV-DBOW	SGD	relu	[250, 100, 100]	0.6642	0.7908	73.86	0.5731	0.5455	0.559	0.6502	491.03
PV-DBOW	SGD	relu	[100, 50, -]	0.664	0.7927	74.39	0.5575	0.5688	0.5631	0.6413	369.27
PV-DBOW	SGD	relu	[250, 50, -]	0.6629	0.7963	73.97	0.5704	0.5455	0.5577	0.6483	508.34
PV-DBOW	SGD	relu	[250, 250, 250]	0.6627	0.7916	72.39	0.5732	0.5365	0.5542	0.6492	794.86
PV-DBOW	Adam	tanh	[500, 500, 100]	0.662	0.789	73.43	0.5696	0.5786	0.5741	0.6511	374.78
PV-DBOW	Adam	tanh	[250, 250, 250]	0.6615	0.7823	73.69	0.5816	0.5628	0.5721	0.6578	289.35
PV-DBOW	Adam	tanh	[500, 500, -]	0.6602	0.7872	74.35	0.5988	0.5816	0.5901	0.6716	581.17
PV-DBOW	Adam	tanh	[250, 250, 100]	0.6601	0.7876	72.89	0.5831	0.5598	0.5712	0.6584	264.34
PV-DBOW	Adam	relu	[100, 50, -]	0.6596	0.7862	74.06	0.5573	0.5455	0.5513	0.6391	138.45
PV-DBOW	Adam	tanh	[500, 250, 50]	0.6582	0.7858	72.06	0.6067	0.5606	0.5827	0.6737	256.13
PV-DBOW	SGD	relu	[100, 100, -]	0.6581	0.7863	72.44	0.5478	0.5515	0.5497	0.6327	518.07
PV-DBOW	SGD	relu	[100, 50, 50]	0.6581	0.787	72.68	0.5545	0.5591	0.5568	0.6382	448.03
PV-DM	SGD	relu	[100, -, -]	0.6578	0.7855	74.58	0.5793	0.5357	0.5567	0.6532	406.38
PV-DBOW	Adam	tanh	[250, 50, -]	0.6574	0.7864	72.75	0.5678	0.5922	0.5797	0.6511	277.48
PV-DBOW	Adam	tanh	[500, 500, 50]	0.6572	0.7909	72.68	0.5879	0.5862	0.587	0.6648	452.67
PV-DM	SGD	relu	[50, -, -]	0.6566	0.7812	74.01	0.5468	0.5139	0.5299	0.6294	336.11
PV-DBOW	Adam	tanh	[500, 50, -]	0.6556	0.784	73.92	0.5698	0.5741	0.572	0.6508	400.86
PV-DBOW	Adam	tanh	[500, 250, 250]	0.655	0.7826	73.79	0.5745	0.5922	0.5832	0.656	328.39
PV-DBOW	Adam	relu	[100, 100, 100]	0.6544	0.7836	72.81	0.5582	0.5591	0.5586	0.641	127.78
PV-DBOW	SGD	relu	[100, 100, 100]	0.6542	0.7831	73.9	0.5391	0.5553	0.5471	0.6263	511.62
PV-DBOW	Adam	tanh	[100, 100, 50]	0.6532	0.782	73.38	0.5583	0.5583	0.5583	0.641	207.32
PV-DBOW	Adam	tanh	[250, 250, 50]	0.653	0.7853	74.63	0.5734	0.5553	0.5642	0.6514	252.29
PV-DBOW	SGD	relu	[100, 100, 50]	0.6527	0.7798	72.78	0.5398	0.5508	0.5453	0.6266	453.5
PV-DM	SGD	relu	[250, -, -]	0.651	0.7788	72.85	0.5512	0.5019	0.5254	0.6315	647.54
PV-DBOW	Adam	tanh	[100, 50, 50]	0.65	0.779	71.93	0.565	0.5726	0.5688	0.6471	198.89
PV-DBOW	Adam	tanh	[500, 100, 100]	0.6497	0.7739	71.12	0.5695	0.5764	0.5729	0.6508	243.46
PV-DM	SGD	relu	[500, -, -]	0.6489	0.7816	73.25	0.5672	0.5719	0.5695	0.6486	1023.4
PV-DBOW	Adam	tanh	[250, 100, -]	0.6483	0.7797	72.33	0.5628	0.5801	0.5713	0.6462	292.92
PV-DBOW	Adam	tanh	[250, 100, 50]	0.6455	0.7772	70.64	0.5734	0.5493	0.5611	0.6508	211.19
PV-DBOW	Adam	tanh	[250, 100, 100]	0.6433	0.7745	71.85	0.5766	0.5944	0.5854	0.6578	216.4
PV-DBOW	Adam	tanh	[500, 250, -]	0.6432	0.7718	72.2	0.543	0.5651	0.5538	0.63	421.79
PV-DBOW	Adam	relu	[100, 100, 50]	0.6431	0.7801	72.71	0.5613	0.5613	0.5613	0.6434	121.1
PV-DBOW	Adam	tanh	[100, 100, -]	0.643	0.7732	71.16	0.5537	0.5741	0.5637	0.6388	270.55
PV-DBOW	Adam	tanh	[500, 100, 50]	0.6416	0.7717	71.75	0.5449	0.5388	0.5418	0.6297	250.93
PV-DBOW	Adam	tanh	[100, 100, 100]	0.6414	0.7705	71.56	0.5532	0.5282	0.5404	0.6349	224.08
PV-DBOW	Adam	tanh	[100, 50, -]	0.6413	0.7726	72.96	0.5376	0.5591	0.5481	0.6254	258.9
PV-DBOW	Adam	tanh	[500, 100, -]	0.6397	0.7683	70.39	0.566	0.5485	0.5571	0.6456	335.9
PV-DBOW	Adam	tanh	[50, 50, -]	0.6376	0.7578	70.44	0.5467	0.5636	0.555	0.6327	285.37
PV-DBOW	Adam	relu	[100, 50, 50]	0.6364	0.7727	70.43	0.5684	0.6095	0.5882	0.6532	114.89
PV-DBOW	Adam	tanh	[500, 50, 50]	0.6358	0.7675	70.43	0.5545	0.5515	0.553	0.6376	285.94
PV-DBOW	Adam	tanh	[250, 250, -]	0.6347	0.7682	70.07	0.5676	0.5719	0.5697	0.6489	354.91
PV-DM	SGD	relu	[50, 50, -]	0.6345	0.7582	72.47	0.5159	0.5132	0.5145	0.6064	285.52
PV-DBOW	Adam	relu	[50, 50, 50]	0.6333	0.7639	70.17	0.5441	0.5613	0.5526	0.6306	178.9
PV-DBOW	Adam	tanh	[50, 50, 50]	0.631	0.7669	70.62	0.5428	0.5636	0.553	0.6297	234.92
PV-DBOW	Adam	tanh	[250, 50, 50]	0.6292	0.7619	70.82	0.5508	0.5591	0.5549	0.6355	225
PV-DM	SGD	relu	[500, 250, 50]	0.6286	0.7586	72.08	0.5227	0.5546	0.5382	0.6131	538.98
PV-DM	SGD	relu	[500, 500, 250]	0.6237	0.7542	71.01	0.5135	0.544	0.5283	0.6052	878.02
PV-DM	SGD	relu	[500, 100, 100]	0.6235	0.7531	71.02	0.5145	0.547	0.5303	0.6061	480.21
PV-DM	SGD	relu	[500, 50, -]	0.6215	0.756	70.78	0.5138	0.5613	0.5365	0.6058	718.87
PV-DM	SGD	relu	[500, 500, 500]	0.6204	0.754	71.33	0.4986	0.5252	0.5115	0.5924	1504.31
PV-DM	SGD	relu	[500, 100, 50]	0.6194	0.7524	70.9	0.5223	0.538	0.53	0.6122	614.61
PV-DM	SGD	relu	[500, 250, 100]	0.6193	0.7517	70.47	0.5032	0.532	0.5172	0.5963	767.04
PV-DM	SGD	relu	[500, 500, 100]	0.6192	0.7505	70.94	0.5158	0.529	0.5223	0.6067	1065.57
PV-DM	SGD	relu	[500, 100, -]	0.6184	0.7496	70.66	0.5207	0.5485	0.5343	0.6113	735.53
PV-DM	SGD	relu	[500, 500, -]	0.6175	0.7533	70.17	0.513	0.5485	0.5302	0.6049	1426.78
PV-DM	SGD	relu	[500, 500, 50]	0.6174	0.7525	71.03	0.5246	0.5613	0.5423	0.615	780.77
PV-DM	SGD	relu	[250, 250, -]	0.6168	0.7424	70.98	0.4877	0.553	0.5183	0.5823	743.93
PV-DM	SGD	relu	[500, 250, 250]	0.6166	0.7523	71.18	0.5228	0.535	0.5288	0.6125	633.24
PV-DM	SGD	relu	[500, 250, -]	0.6164	0.7513	70.24	0.5217	0.5598	0.5401	0.6125	932.22
PV-DM	SGD	relu	[500, 50, 50]	0.6161	0.7498	69.61	0.5162	0.5388	0.5272	0.6073	392.48

Preliminary experiments using only (q, c) inputs – All results.

Category	**Solver**	**Activation**	**Hidden Layer**	**MAP**	**AvgRec**	**MRR**	**P**	**R**	$\mathbf{F_1}$	**Acc**	**Runtime(sec)**
PV-DM	SGD	relu	[250, 50, -]	0.6154	0.7413	72.24	0.51	0.5546	0.5314	0.6024	531.63
PV-DM	SGD	relu	[250, 250, 50]	0.6146	0.7442	69.09	0.478	0.5305	0.5029	0.5737	628.25
PV-DM	SGD	relu	[100, 50, 50]	0.6131	0.7449	69.56	0.5068	0.5598	0.532	0.5997	503.87
PV-DM	SGD	relu	[250, 50, 50]	0.6128	0.7386	69.05	0.4951	0.5357	0.5146	0.5893	432.66
PV-DM	SGD	relu	[100, 50, -]	0.6108	0.7475	69.63	0.498	0.553	0.5241	0.5917	434.61
PV-DM	SGD	relu	[250, 100, -]	0.6093	0.7437	68.79	0.5153	0.5583	0.5359	0.607	557.81
PV-DM	SGD	relu	[100, 100, -]	0.6089	0.7471	68.59	0.5058	0.5869	0.5434	0.5991	447.16
PV-DM	SGD	relu	[250, 100, 100]	0.6077	0.7405	69.81	0.51	0.5365	0.5229	0.6021	509.99
PV-DM	SGD	relu	[250, 250, 100]	0.6047	0.7417	68.62	0.5048	0.5553	0.5288	0.5979	597.51
PV-DM	SGD	relu	[50, 50, 50]	0.6009	0.7415	68.04	0.5065	0.5862	0.5434	0.5997	306.17
PV-DM	SGD	relu	[250, 100, 50]	0.5989	0.7338	69.07	0.4858	0.5267	0.5054	0.581	502.23
PV-DM	SGD	relu	[250, 250, 250]	0.5984	0.7365	67.62	0.4975	0.5312	0.5138	0.5914	698.78
PV-DM	SGD	relu	[100, 100, 50]	0.5956	0.7339	66.41	0.5123	0.5342	0.523	0.604	510.46
PV-DM	SGD	relu	[100, 100, 100]	0.5931	0.7225	66.82	0.4963	0.5523	0.5228	0.5902	573.52
PV-DBOW	SGD	logistic	[500, 100, -]	0.586	0.7138	65.75	0	0	0	0.5936	132.45
PV-DBOW	SGD	logistic	[250, 50, -]	0.5839	0.7233	65.44	0	0	0	0.5936	135.94
PV-DBOW	SGD	logistic	[500, 250, 100]	0.581	0.7154	65.69	0	0	0	0.5936	155.95
PV-DBOW	SGD	logistic	[500, 100, 100]	0.5751	0.7053	64.64	0	0	0	0.5936	139.51
PV-DBOW	SGD	logistic	[250, 250, 50]	0.5749	0.7005	63.55	0	0	0	0.5936	133.74
PV-DM	SGD	logistic	[500, 250, -]	0.5671	0.6922	64.02	0	0	0	0.5936	167.21
PV-DM	SGD	logistic	[250, 250, 50]	0.566	0.6933	64.56	0	0	0	0.5936	147.52
PV-DBOW	SGD	logistic	[500, 250, -]	0.5642	0.6977	61.58	0	0	0	0.5936	150.1
PV-DBOW	SGD	logistic	[500, 500, -]	0.5574	0.6849	62.9	0	0	0	0.5936	168.45
PV-DBOW	SGD	logistic	[100, 100, -]	0.5558	0.6908	61.06	0	0	0	0.5936	110.29
PV-DM	SGD	logistic	[100, 100, 100]	0.5535	0.6856	61.61	0	0	0	0.5936	127.31
PV-DM	SGD	logistic	[250, 100, 50]	0.5511	0.6856	61.34	0	0	0	0.5936	132.9
PV-DM	SGD	logistic	[250, 250, 100]	0.548	0.6785	63	0	0	0	0.5936	147.96
PV-DBOW	SGD	logistic	[500, 250, 250]	0.5469	0.684	61.51	0	0	0	0.5936	166.18
PV-DM	SGD	logistic	[250, 100, 100]	0.5449	0.679	62.94	0	0	0	0.5936	134.04
PV-DM	SGD	logistic	[250, 250, -]	0.5421	0.6772	60.01	0	0	0	0.5936	137.88
PV-DM	SGD	logistic	[500, 100, 50]	0.5417	0.6665	59.95	0	0	0	0.5936	156.08
PV-DBOW	SGD	logistic	[250, 100, -]	0.5412	0.6712	62.1	0	0	0	0.5936	141.42
PV-DM	SGD	logistic	[500, 100, -]	0.5407	0.6747	60.5	0	0	0	0.5936	153.05
PV-DBOW	SGD	logistic	[50, 50, -]	0.5397	0.6761	60.53	0	0	0	0.5936	96.45
PV-DBOW	SGD	logistic	[250, 50, 50]	0.5392	0.675	60.45	0	0	0	0.5936	142.97
PV-DBOW	SGD	logistic	[500, 100, 50]	0.5377	0.6729	60.24	0	0	0	0.5936	142.11
PV-DM	SGD	logistic	[500, 50, -]	0.5376	0.6686	60.73	0	0	0	0.5936	149.02
PV-DM	SGD	logistic	[250, 50, 50]	0.5366	0.6711	61.68	0	0	0	0.5936	129.84
PV-DM	SGD	logistic	[100, 50, 50]	0.5351	0.6675	59.26	0	0	0	0.5936	115.25
PV-DBOW	SGD	logistic	[250, 100, 100]	0.5343	0.657	58.2	0	0	0	0.5936	126.65
PV-DM	SGD	logistic	[100, 50, -]	0.5327	0.6644	59.19	0	0	0	0.5936	112.61
PV-DM	SGD	logistic	[500, 500, 100]	0.5312	0.6575	59.8	0	0	0	0.5936	196.87
PV-DM	SGD	logistic	[500, 500, 500]	0.53	0.6616	61.47	0	0	0	0.5936	238.96
PV-DBOW	SGD	logistic	[50, 50, 50]	0.53	0.6744	59.74	0	0	0	0.5936	98.65
PV-DBOW	SGD	logistic	[500, 50, -]	0.5299	0.6692	58.93	0	0	0	0.5936	133.22
PV-DBOW	SGD	logistic	[100, 50, 50]	0.5282	0.6548	58.24	0	0	0	0.5936	102.35
PV-DM	SGD	logistic	[250, 50, -]	0.5277	0.6572	58.57	0	0	0	0.5936	127.24
PV-DBOW	SGD	logistic	[100, 100, 100]	0.5264	0.66	57.68	0	0	0	0.5936	127.72
PV-DM	SGD	logistic	[500, 500, -]	0.5256	0.6514	57.24	0	0	0	0.5936	188.14
PV-DM	SGD	logistic	[50, 50, 50]	0.5231	0.6516	58.16	0	0	0	0.5936	111.9
PV-DM	SGD	logistic	[500, 500, 50]	0.523	0.6601	57.51	0	0	0	0.5936	201.83
PV-DBOW	SGD	logistic	[500, 500, 100]	0.5229	0.6494	59.03	0	0	0	0.5936	172.3
PV-DM	SGD	logistic	[50, 50, -]	0.5224	0.6564	58.23	0	0	0	0.5936	109.46
PV-DM	SGD	logistic	[500, 250, 50]	0.5218	0.6586	57.39	0	0	0	0.5936	175.63
PV-DBOW	SGD	logistic	[500, 500, 250]	0.5218	0.6445	56.53	0	0	0	0.5936	202.79
PV-DBOW	SGD	logistic	[100, 100, 50]	0.5211	0.65	56.6	0	0	0	0.5936	133.63
PV-DM	SGD	logistic	[250, 250, 250]	0.5208	0.6514	59.21	0	0	0	0.5936	155.23
PV-DM	SGD	logistic	[500, 100, 100]	0.5195	0.6525	57.21	0	0	0	0.5936	155.54
PV-DBOW	SGD	logistic	[250, 250, -]	0.5173	0.6502	56.13	0	0	0	0.5936	132.11
PV-DM	SGD	logistic	[500, 50, 50]	0.5157	0.642	56.98	0	0	0	0.5936	154.23
PV-DM	SGD	logistic	[100, 100, 50]	0.5151	0.6481	57.97	0	0	0	0.5936	127.28
PV-DBOW	SGD	logistic	[500, 50, 50]	0.5151	0.6501	56.55	0	0	0	0.5936	129.06
PV-DBOW	SGD	logistic	[100, 50, -]	0.5151	0.6553	58.1	0	0	0	0.5936	99.76
PV-DBOW	SGD	logistic	[250, 250, 250]	0.5132	0.6458	57.58	0	0	0	0.5936	162.82
PV-DM	SGD	logistic	[250, 100, -]	0.5127	0.6427	57.32	0	0	0	0.5936	129.53
PV-DM	SGD	logistic	[500, 250, 100]	0.5116	0.6428	57.82	0	0	0	0.5936	168.03
PV-DM	SGD	logistic	[500, 500, 250]	0.5112	0.6492	55.82	0	0	0	0.5936	224.96
PV-DM	SGD	logistic	[100, 100, -]	0.5098	0.649	57.46	0	0	0	0.5936	115.69
PV-DM	SGD	logistic	[100, -, -]	0.508	0.6317	55.97	0	0	0	0.5936	114.61
PV-DBOW	SGD	logistic	[500, 500, 50]	0.5049	0.6364	55.15	0	0	0	0.5936	173.42
PV-DM	SGD	logistic	[500, 250, 250]	0.4981	0.6294	54.73	0	0	0	0.5936	187.67
PV-DM	SGD	logistic	[50, -, -]	0.4979	0.6309	53.81	0	0	0	0.5936	111.51
PV-DBOW	SGD	logistic	[250, 250, 100]	0.493	0.6307	55.28	0	0	0	0.5936	138.22
PV-DBOW	SGD	logistic	[500, 250, 50]	0.4918	0.6242	52.28	0	0	0	0.5936	144.43
PV-DM	SGD	logistic	[250, -, -]	0.491	0.6178	53.62	0	0	0	0.5936	133.34
PV-DM	SGD	logistic	[500, -, -]	0.4804	0.6082	51.04	0	0	0	0.5936	163.81
PV-DBOW	SGD	logistic	[500, 500, 500]	0.4698	0.6025	51.18	0	0	0	0.5936	236.12
PV-DBOW	SGD	logistic	[250, 100, 50]	0.4678	0.5966	49.34	0	0	0	0.5936	125.11

Preliminary experiments using only (q, c) inputs – All results.

Appendix C

Experiments post inclusion of Average Answer

Category	Solver	Activation	Hidden Layer	MAP	AvgRec	MRR	P	R	F_1	Acc	Runtime(sec)
PV-DBOW	SGD	relu	[250, -, -]	0.7306	0.8416	79.61	0.6607	0.5786	0.6169	0.708	1006.57
PV-DBOW	SGD	relu	[100, -, -]	0.7296	0.8418	80.81	0.6678	0.5688	0.6144	0.7098	617.66
PV-DBOW	SGD	relu	[500, -, -]	0.7253	0.8429	78.88	0.6536	0.6035	0.6275	0.7089	1475.27
PV-DBOW	SGD	tanh	[500, 100, -]	0.7188	0.8343	79.11	0.6684	0.5658	0.6129	0.7095	260.33
PV-DBOW	SGD	tanh	[250, 250, 250]	0.7182	0.834	78.87	0.6655	0.5658	0.6116	0.708	303.5
PV-DBOW	SGD	logistic	[100, -, -]	0.7179	0.8346	79.13	0.6652	0.553	0.6039	0.7052	253.01
PV-DBOW	Adam	logistic	[50, -, -]	0.7177	0.8342	78.96	0.6635	0.5726	0.6147	0.7083	131.9
PV-DBOW	SGD	tanh	[500, 500, 500]	0.7172	0.8337	78.78	0.6675	0.5696	0.6147	0.7098	543.58
PV-DBOW	SGD	tanh	[250, 100, 50]	0.7171	0.8329	78.48	0.6655	0.5673	0.6125	0.7083	424.85
PV-DBOW	SGD	tanh	[50, -, -]	0.7171	0.8339	79.09	0.6628	0.5606	0.6074	0.7055	177.75
PV-DBOW	SGD	tanh	[250, 250, 100]	0.7169	0.8333	78.79	0.6661	0.5613	0.6092	0.7073	240.19
PV-DBOW	SGD	tanh	[100, 100, -]	0.7169	0.8338	79.13	0.6652	0.5576	0.6066	0.7061	311.23
PV-DBOW	SGD	tanh	[50, 50, 50]	0.7169	0.8325	78.47	0.6673	0.5658	0.6124	0.7089	290.3
PV-DBOW	Adam	tanh	[250, -, -]	0.7169	0.8339	78.64	0.655	0.5771	0.6136	0.7046	351.16
PV-DBOW	SGD	logistic	[250, -, -]	0.7168	0.8338	78.83	0.6664	0.5576	0.6071	0.7067	371.12
PV-DBOW	SGD	tanh	[100, 50, 50]	0.7167	0.8326	78.84	0.6687	0.5636	0.6117	0.7092	338.32
PV-DBOW	Adam	logistic	[250, -, -]	0.7167	0.834	78.68	0.6676	0.538	0.5958	0.7034	145.1
PV-DBOW	SGD	tanh	[500, 500, 50]	0.7166	0.8333	78.75	0.6676	0.5681	0.6138	0.7095	436.02
PV-DBOW	SGD	tanh	[250, -, -]	0.7166	0.8336	79.01	0.6649	0.5598	0.6078	0.7064	490.53
PV-DBOW	SGD	tanh	[250, 250, 50]	0.7165	0.8333	78.64	0.6652	0.5621	0.6093	0.707	373.19
PV-DBOW	SGD	tanh	[250, 250, -]	0.7165	0.8324	78.77	0.6661	0.5628	0.6101	0.7076	366.54
PV-DBOW	SGD	tanh	[100, 100, 100]	0.7164	0.8328	78.71	0.6667	0.5658	0.6121	0.7086	360.34
PV-DBOW	Adam	logistic	[250, 250, -]	0.7164	0.8334	78.65	0.6562	0.5959	0.6246	0.7089	151.33
PV-DBOW	Adam	logistic	[100, 100, -]	0.7163	0.8331	78.76	0.6598	0.5779	0.6161	0.7073	143.98
PV-DBOW	SGD	logistic	[500, -, -]	0.7162	0.8336	78.93	0.6634	0.553	0.6032	0.7043	654.97
PV-DBOW	Adam	logistic	[500, -, -]	0.7162	0.8334	78.79	0.6603	0.5704	0.612	0.7061	152.61
PV-DBOW	Adam	logistic	[100, 100, 50]	0.7162	0.8331	78.73	0.6519	0.5862	0.6173	0.7046	131.12
PV-DBOW	SGD	tanh	[500, 250, 50]	0.7161	0.8332	78.96	0.667	0.5621	0.61	0.708	283.91
PV-DBOW	SGD	tanh	[250, 50, -]	0.7161	0.8335	78.69	0.6602	0.5613	0.6068	0.7043	434.61
PV-DBOW	Adam	relu	[500, 250, 250]	0.7161	0.8263	79.45	0.6298	0.6005	0.6148	0.6942	207.78
PV-DBOW	SGD	tanh	[250, 100, -]	0.716	0.8328	78.31	0.6655	0.5628	0.6099	0.7073	529.11
PV-DBOW	SGD	tanh	[500, 500, -]	0.7158	0.8328	78.42	0.6652	0.5621	0.6093	0.707	512.39
PV-DBOW	SGD	tanh	[500, 100, 50]	0.7158	0.8321	78.53	0.667	0.5681	0.6136	0.7092	229.37
PV-DBOW	SGD	tanh	[100, 50, -]	0.7157	0.8328	78.88	0.6649	0.5613	0.6087	0.7067	345.05
PV-DBOW	Adam	logistic	[50, 50, 50]	0.7157	0.8325	78.64	0.6472	0.6087	0.6274	0.7061	142.91
PV-DBOW	SGD	tanh	[500, 500, 250]	0.7156	0.8327	78.65	0.6664	0.5651	0.6116	0.7083	497.31
PV-DBOW	SGD	tanh	[500, 250, 250]	0.7156	0.8329	78.52	0.6681	0.5666	0.6132	0.7095	388.15
PV-DBOW	SGD	tanh	[500, -, -]	0.7156	0.8328	78.76	0.6676	0.5591	0.6085	0.7076	289.94
PV-DBOW	SGD	tanh	[500, 100, 100]	0.7154	0.8329	78.52	0.6697	0.5628	0.6116	0.7095	327.82
PV-DBOW	SGD	tanh	[250, 50, 50]	0.7154	0.8329	78.78	0.6611	0.5651	0.6093	0.7055	567.72

TABLE 8: Experiments using (q, c, avg_com_q) inputs – All results.

Category	Solver	Activation	Hidden Layer	MAP	AvgRec	MRR	P	R	F_1	Acc	Runtime(sec)
PV-DBOW	SGD	tanh	[100, -, -]	0.7154	0.8324	78.62	0.6625	0.5523	0.6024	0.7037	327.56
PV-DBOW	SGD	tanh	[50, 50, -]	0.7154	0.8325	78.58	0.6658	0.5666	0.6122	0.7083	327.6
PV-DBOW	Adam	tanh	[500, -, -]	0.7152	0.8335	78.54	0.685	0.468	0.5561	0.6963	328.31
PV-DBOW	Adam	logistic	[100, 100, 100]	0.7152	0.8327	78.52	0.6649	0.5538	0.6043	0.7052	140.71
PV-DBOW	SGD	tanh	[500, 50, -]	0.715	0.8329	78.6	0.6667	0.5673	0.613	0.7089	250.67
PV-DBOW	SGD	tanh	[500, 250, -]	0.7148	0.8325	78.69	0.6664	0.5636	0.6107	0.708	330.26
PV-DBOW	SGD	tanh	[500, 250, 100]	0.7147	0.8324	78.41	0.6646	0.5621	0.6091	0.7067	364.63
PV-DBOW	SGD	tanh	[250, 100, 100]	0.7147	0.8318	78.47	0.664	0.5636	0.6097	0.7067	385.25
PV-DBOW	SGD	logistic	[50, -, -]	0.7147	0.8328	78.62	0.6673	0.5598	0.6088	0.7076	216.07
PV-DBOW	Adam	logistic	[100, -, -]	0.7147	0.8327	78.63	0.6686	0.5312	0.592	0.7024	135.51
PV-DBOW	SGD	relu	[50, -, -]	0.7145	0.8328	78.38	0.6611	0.5636	0.6084	0.7052	204.36
PV-DBOW	Adam	logistic	[50, 50, -]	0.7144	0.8325	78.38	0.6676	0.544	0.5995	0.7046	132.09
PV-DBOW	SGD	tanh	[500, 50, 50]	0.7143	0.8322	78.5	0.669	0.5643	0.6122	0.7095	250.62
PV-DBOW	SGD	tanh	[500, 500, 100]	0.7142	0.8315	78.38	0.6681	0.5651	0.6123	0.7092	450.76
PV-DBOW	Adam	relu	[500, 100, -]	0.714	0.8281	79.44	0.6341	0.5816	0.6068	0.6936	168.03
PV-DBOW	Adam	logistic	[500, 500, 250]	0.714	0.832	78.36	0.6703	0.5613	0.611	0.7095	172.65
PV-DBOW	SGD	tanh	[100, 100, 50]	0.7139	0.8316	78.34	0.6673	0.5673	0.6133	0.7092	376.71
PV-DBOW	Adam	logistic	[500, 250, -]	0.7138	0.8314	78.68	0.6785	0.5192	0.5882	0.7046	227.55
PV-DBOW	Adam	logistic	[250, 250, 100]	0.7138	0.8319	78.73	0.6524	0.6072	0.629	0.7089	210.29
PV-DBOW	Adam	logistic	[500, 100, 100]	0.7133	0.8315	78.77	0.6623	0.5711	0.6133	0.7073	186.39
PV-DBOW	Adam	tanh	[50, -, -]	0.7131	0.8317	78.16	0.6599	0.5606	0.6062	0.704	128.91
PV-DBOW	SGD	relu	[500, 50, -]	0.7128	0.8242	79.71	0.618	0.5989	0.6083	0.6865	1003.3
PV-DBOW	Adam	logistic	[500, 500, 50]	0.7128	0.831	78.52	0.655	0.6072	0.6302	0.7104	233.55
PV-DBOW	Adam	logistic	[250, 250, 50]	0.7125	0.8311	78.58	0.6591	0.5892	0.6222	0.7092	199.44
PV-DBOW	Adam	logistic	[500, 50, 50]	0.7123	0.8328	78.68	0.6357	0.6343	0.635	0.7037	175.17
PV-DBOW	Adam	logistic	[250, 50, 50]	0.7122	0.8317	78.76	0.6722	0.5139	0.5825	0.7006	172.98
PV-DBOW	Adam	logistic	[500, 500, -]	0.7117	0.8315	78.78	0.6548	0.608	0.6305	0.7104	314.9
PV-DBOW	Adam	logistic	[500, 250, 50]	0.7114	0.8301	78.31	0.666	0.5222	0.5854	0.6994	197.5
PV-DBOW	Adam	logistic	[250, 50, -]	0.7113	0.8312	78.68	0.6515	0.6035	0.6266	0.7076	170.79
PV-DBOW	Adam	logistic	[500, 500, 500]	0.711	0.8313	78.21	0.6691	0.5568	0.6078	0.708	389.5
PV-DBOW	Adam	logistic	[500, 100, -]	0.711	0.8314	78.68	0.6635	0.5711	0.6138	0.708	220.66
PV-DBOW	Adam	logistic	[250, 100, -]	0.711	0.8304	78.53	0.6648	0.5403	0.5961	0.7024	175.26
PV-DBOW	Adam	logistic	[500, 100, 50]	0.7109	0.8311	78.61	0.6899	0.4838	0.5688	0.7018	191.68
PV-DBOW	Adam	logistic	[250, 100, 50]	0.7108	0.8309	78.67	0.6617	0.5711	0.6131	0.707	182.06
PV-DBOW	Adam	logistic	[500, 250, 250]	0.7106	0.8304	78.17	0.6562	0.5801	0.6158	0.7058	251.89
PV-DBOW	Adam	logistic	[250, 250, 250]	0.7103	0.8298	78.17	0.6676	0.5275	0.5893	0.7012	228.06
PV-DBOW	Adam	logistic	[500, 250, 100]	0.71	0.8307	78.6	0.6658	0.5591	0.6078	0.7067	280.17
PV-DBOW	Adam	logistic	[500, 50, -]	0.7093	0.8296	78.19	0.6728	0.5199	0.5866	0.7021	195
PV-DBOW	Adam	logistic	[250, 100, 100]	0.7093	0.8297	78.31	0.655	0.6042	0.6286	0.7098	189.62
PV-DBOW	Adam	logistic	[500, 500, 100]	0.7088	0.8305	78.39	0.6654	0.5478	0.6009	0.7043	296.11
PV-DBOW	Adam	logistic	[100, 50, 50]	0.7083	0.83	78.22	0.6437	0.6253	0.6344	0.707	162.92
PV-DBOW	Adam	relu	[250, 100, 100]	0.7015	0.8172	77.38	0.5863	0.6494	0.6162	0.6713	167
PV-DBOW	SGD	relu	[500, 500, -]	0.7006	0.8179	78.32	0.6204	0.6027	0.6115	0.6887	1784.41
PV-DBOW	SGD	relu	[500, 250, 250]	0.7	0.8199	77.32	0.6109	0.6012	0.606	0.6823	1248.84
PV-DBOW	SGD	relu	[500, 100, -]	0.699	0.8198	78.02	0.6145	0.5937	0.6039	0.6835	1106.38
PV-DBOW	SGD	relu	[250, 250, 50]	0.6988	0.8125	77.23	0.6056	0.5696	0.587	0.6743	793.21
PV-DBOW	SGD	relu	[500, 250, 100]	0.6977	0.82	76.84	0.6097	0.5854	0.5973	0.6792	1065.68
PV-DBOW	SGD	relu	[250, 250, 250]	0.6976	0.8139	77.73	0.6029	0.5862	0.5944	0.6749	1091.9
PV-DBOW	SGD	relu	[250, 250, -]	0.6974	0.8184	76.81	0.6016	0.5771	0.5891	0.6728	803.06
PV-DBOW	Adam	relu	[500, -, -]	0.6973	0.8177	76.76	0.6268	0.5839	0.6046	0.6896	223.56
PV-DBOW	Adam	relu	[500, 50, 50]	0.6946	0.8204	77.34	0.6198	0.6035	0.6115	0.6884	181.34
PV-DBOW	SGD	relu	[500, 250, -]	0.6944	0.8152	76.18	0.613	0.5816	0.5969	0.6807	1282.36
PV-DBOW	SGD	relu	[500, 500, 250]	0.6941	0.8162	77.64	0.6005	0.5711	0.5854	0.6713	1787.21
PV-DBOW	SGD	relu	[250, 100, -]	0.6939	0.8152	76.73	0.59	0.5922	0.5911	0.667	814.86
PV-DBOW	Adam	relu	[250, 100, -]	0.6938	0.8163	76.97	0.6092	0.5666	0.5871	0.6761	165.17
PV-DBOW	SGD	relu	[250, 50, -]	0.6933	0.8117	78.16	0.5957	0.5877	0.5917	0.6703	730.03
PV-DBOW	Adam	relu	[250, -, -]	0.6932	0.8148	76.87	0.6046	0.5741	0.589	0.6743	223.95
PV-DBOW	Adam	relu	[500, 500, -]	0.6927	0.8124	76.81	0.5747	0.6599	0.6144	0.6633	220.47
PV-DBOW	Adam	relu	[500, 500, 500]	0.6924	0.8171	77.13	0.6233	0.5839	0.603	0.6875	252.46
PV-DBOW	Adam	relu	[100, 100, -]	0.6917	0.8121	76.6	0.6	0.5688	0.584	0.6706	168.93
PV-DBOW	Adam	relu	[500, 500, 250]	0.6908	0.8185	76.13	0.6068	0.6561	0.6305	0.6875	274.75
PV-DBOW	Adam	relu	[100, 100, 100]	0.6908	0.8105	76.42	0.5951	0.5463	0.5696	0.6645	162.52
PV-DBOW	Adam	relu	[250, 250, 50]	0.6902	0.8047	76.05	0.6003	0.5538	0.5761	0.6688	190.65
PV-DBOW	SGD	relu	[500, 50, 50]	0.689	0.8114	76.18	0.6038	0.5974	0.6006	0.6771	573.03
PV-DBOW	Adam	relu	[500, 100, 100]	0.6889	0.8091	75.52	0.6233	0.5403	0.5788	0.6804	176.74
PV-DBOW	Adam	relu	[500, 50, -]	0.6885	0.8147	76.52	0.6255	0.5944	0.6096	0.6905	170.29
PV-DBOW	SGD	relu	[250, 250, 100]	0.6883	0.8116	74.95	0.6159	0.5756	0.5951	0.6817	850.23
PV-DBOW	SGD	relu	[250, 50, 50]	0.688	0.807	76.93	0.5986	0.5688	0.5833	0.6697	647.14
PV-DBOW	Adam	relu	[500, 250, 100]	0.6879	0.8109	76.69	0.608	0.5229	0.5623	0.6691	212.82
PV-DBOW	Adam	logistic	[100, 50, -]	0.6877	0.8134	75.84	0.622	0.5275	0.5708	0.6777	317.94
PV-DBOW	SGD	relu	[100, 100, -]	0.687	0.8113	77.42	0.5821	0.5764	0.5792	0.6596	615.04
PV-DBOW	Adam	relu	[500, 500, 50]	0.6869	0.8108	76.01	0.595	0.6125	0.6036	0.6731	211.89
PV-DBOW	Adam	relu	[250, 250, 250]	0.6864	0.8077	75.48	0.6141	0.5688	0.5906	0.6795	178.08
PV-DBOW	Adam	relu	[250, 50, 50]	0.6861	0.8103	76.95	0.6116	0.5771	0.5939	0.6792	162.42
PV-DBOW	Adam	relu	[250, 250, 100]	0.6859	0.8082	76.68	0.5777	0.6072	0.5921	0.6599	175.58
PV-DBOW	Adam	relu	[250, 50, -]	0.6859	0.8124	76.08	0.6308	0.5463	0.5855	0.6856	162.23
PV-DBOW	Adam	tanh	[500, 250, 100]	0.6858	0.804	77.83	0.6071	0.5801	0.5933	0.6768	656.93
PV-DBOW	SGD	relu	[250, 100, 100]	0.6857	0.8066	75.94	0.5995	0.5824	0.5908	0.6722	679.61
PV-DBOW	Adam	relu	[500, 500, 100]	0.6855	0.8143	74.75	0.6569	0.5372	0.5911	0.6979	206.64
PV-DBOW	SGD	relu	[500, 100, 50]	0.6841	0.8064	77.55	0.5872	0.5749	0.581	0.663	925.53
PV-DBOW	Adam	relu	[250, 250, -]	0.6834	0.802	75.26	0.6039	0.5162	0.5566	0.6657	181.22
PV-DBOW	Adam	relu	[50, -, -]	0.683	0.8092	76.38	0.5855	0.5485	0.5664	0.6587	408.89

Experiments using (q, c, avg_com_q) inputs – All results.

Category	Solver	Activation	Hidden Layer	MAP	AvgRec	MRR	P	R	F_1	Acc	Runtime(sec)
PV-DBOW	SGD	relu	[500, 100, 100]	0.6827	0.8053	75.99	0.5991	0.5824	0.5906	0.6719	858.49
PV-DBOW	SGD	relu	[500, 500, 100]	0.6826	0.8076	75.83	0.6103	0.5726	0.5908	0.6777	1449.54
PV-DBOW	Adam	relu	[250, 100, 50]	0.6825	0.8119	75.69	0.5909	0.5967	0.5938	0.6682	166.28
PV-DBOW	SGD	relu	[500, 250, 50]	0.6824	0.8122	76.21	0.6137	0.6132	0.6135	0.6859	1099
PV-DBOW	Adam	relu	[100, -, -]	0.6824	0.8063	75.7	0.5995	0.5824	0.5908	0.6722	212.28
PV-DBOW	SGD	relu	[500, 500, 500]	0.6818	0.8054	75.2	0.5922	0.6163	0.604	0.6716	2085.15
PV-DBOW	Adam	relu	[500, 250, 50]	0.6816	0.8152	75.63	0.6303	0.5132	0.5657	0.6798	199.51
PV-DBOW	Adam	relu	[500, 250, -]	0.6802	0.8125	74.16	0.6204	0.5312	0.5724	0.6774	192.04
PV-DBOW	Adam	relu	[500, 100, 50]	0.6796	0.8136	74.41	0.6452	0.5282	0.5809	0.6902	178.77
PV-DBOW	SGD	relu	[500, 500, 50]	0.6794	0.8061	75.36	0.6008	0.5809	0.5907	0.6728	1523.85
PV-DBOW	Adam	tanh	[50, 50, -]	0.6792	0.801	75.67	0.5701	0.5688	0.5695	0.6505	275.89
PV-DBOW	SGD	relu	[50, 50, 50]	0.6789	0.8029	76.84	0.5881	0.5877	0.5879	0.6651	426.89
PV-DBOW	Adam	tanh	[100, -, -]	0.678	0.8026	74.38	0.5924	0.5741	0.5831	0.6664	383.1
PV-DBOW	SGD	relu	[250, 100, 50]	0.6767	0.8005	74.68	0.5882	0.5568	0.5721	0.6615	639.07
PV-DBOW	Adam	relu	[100, 100, 50]	0.6767	0.7996	74.64	0.5998	0.4929	0.5411	0.6602	159.97
PV-DBOW	Adam	tanh	[500, 500, 50]	0.6748	0.799	75.93	0.5947	0.5929	0.5938	0.6703	726.37
PV-DBOW	Adam	tanh	[500, 500, 100]	0.673	0.7966	73.96	0.5973	0.5749	0.5859	0.6697	1295.65
PV-DBOW	Adam	tanh	[500, 250, 250]	0.6729	0.794	73.76	0.6064	0.5959	0.6011	0.6786	836.25
PV-DBOW	SGD	relu	[50, 50, -]	0.671	0.7981	75.27	0.584	0.5673	0.5756	0.6599	432.26
PV-DBOW	SGD	relu	[100, 50, -]	0.6708	0.7932	75.14	0.5707	0.5651	0.5679	0.6505	583.72
PV-DBOW	Adam	tanh	[250, 100, 50]	0.6704	0.7912	74.4	0.599	0.5372	0.5664	0.6657	810.11
PV-DBOW	Adam	relu	[100, 50, -]	0.6697	0.7927	75.04	0.56	0.541	0.5503	0.6407	161.75
PV-DBOW	SGD	relu	[100, 50, 50]	0.6696	0.797	74.92	0.5833	0.553	0.5678	0.6578	494.08
PV-DBOW	Adam	tanh	[500, 250, 50]	0.6691	0.7966	74.67	0.6011	0.5862	0.5935	0.6737	764.32
PV-DBOW	Adam	tanh	[500, 100, 100]	0.6689	0.7968	75.45	0.5827	0.5621	0.5722	0.6584	712.82
PV-DBOW	Adam	tanh	[250, 250, -]	0.6673	0.796	74.55	0.6018	0.5673	0.584	0.6716	1160.42
PV-DBOW	SGD	relu	[100, 100, 100]	0.6661	0.7933	74.06	0.5867	0.5756	0.5811	0.6627	686.74
PV-DBOW	Adam	tanh	[500, 250, -]	0.6657	0.7898	76.67	0.5843	0.5425	0.5626	0.6572	1068.04
PV-DBOW	Adam	tanh	[100, 100, 100]	0.6656	0.7921	74.25	0.5873	0.5719	0.5795	0.6627	764.41
PV-DBOW	Adam	relu	[50, 50, 50]	0.6655	0.7913	73.14	0.5826	0.5809	0.5818	0.6606	190.87
PV-DBOW	Adam	relu	[50, 50, -]	0.665	0.7927	73.82	0.5687	0.5229	0.5449	0.645	386.55
PV-DBOW	Adam	tanh	[250, 100, -]	0.6641	0.7907	73.35	0.5932	0.541	0.5659	0.6627	905.42
PV-DBOW	SGD	relu	[100, 100, 50]	0.6635	0.7998	73.89	0.5994	0.5809	0.59	0.6719	676.15
PV-DBOW	Adam	tanh	[250, 250, 250]	0.6633	0.7923	74.53	0.5917	0.5779	0.5847	0.6664	935.49
PV-DBOW	Adam	relu	[100, 50, 50]	0.6615	0.7899	70.95	0.5711	0.6072	0.5886	0.655	160.35
PV-DBOW	Adam	tanh	[500, 500, 250]	0.6614	0.7936	74.14	0.6085	0.5719	0.5896	0.6765	873.68
PV-DBOW	Adam	tanh	[500, 100, -]	0.6611	0.783	73.24	0.5685	0.5591	0.5637	0.6483	888.85
PV-DBOW	Adam	tanh	[100, 100, -]	0.6609	0.7917	74.31	0.5969	0.5214	0.5566	0.6624	1306.94
PV-DBOW	Adam	tanh	[250, 250, 50]	0.6607	0.7914	73.87	0.586	0.5335	0.5585	0.6572	880.54
PV-DBOW	Adam	tanh	[500, 50, -]	0.6588	0.7872	73.37	0.5839	0.5523	0.5677	0.6581	953.45
PV-DBOW	Adam	tanh	[250, 250, 100]	0.6588	0.7905	73.31	0.5886	0.5448	0.5658	0.6602	881.99
PV-DBOW	Adam	tanh	[500, 100, 50]	0.6586	0.7934	74.55	0.5929	0.5786	0.5857	0.6673	678.64
PV-DBOW	Adam	tanh	[500, 500, 500]	0.6576	0.7848	73.9	0.6056	0.5719	0.5882	0.6746	1325.17
PV-DBOW	Adam	tanh	[500, 500, -]	0.6567	0.7841	72.29	0.575	0.5681	0.5715	0.6538	1076.54
PV-DBOW	Adam	tanh	[250, 100, 100]	0.6567	0.7816	72.83	0.556	0.5673	0.5616	0.6401	703.61
PV-DBOW	Adam	tanh	[250, 50, 50]	0.6551	0.7845	73.46	0.5917	0.5606	0.5757	0.6642	843.5
PV-DBOW	Adam	tanh	[100, 100, 50]	0.6537	0.7766	72.16	0.5758	0.5546	0.565	0.6529	861.98
PV-DBOW	Adam	tanh	[100, 50, -]	0.6537	0.7844	72.16	0.5842	0.5561	0.5698	0.6587	258.91
PV-DBOW	Adam	tanh	[250, 50, -]	0.6497	0.7818	73.01	0.5789	0.541	0.5593	0.6535	1086.33
PV-DBOW	Adam	tanh	[500, 50, 50]	0.6494	0.7765	72.93	0.5772	0.5342	0.5549	0.6517	728.79
PV-DBOW	Adam	tanh	[100, 50, 50]	0.6445	0.7727	71.17	0.5549	0.5666	0.5607	0.6391	229.08
PV-DBOW	Adam	tanh	[50, 50, 50]	0.6403	0.7722	71.21	0.576	0.5561	0.5658	0.6532	240.61
PV-DBOW	SGD	logistic	[100, 50, -]	0.5868	0.7112	66.37	0	0	0	0.5936	155.98
PV-DBOW	SGD	logistic	[500, 50, -]	0.5815	0.7143	65.25	0	0	0	0.5936	233.43
PV-DBOW	SGD	logistic	[250, 250, -]	0.5815	0.7202	66	0	0	0	0.5936	185.18
PV-DBOW	SGD	logistic	[250, 250, 250]	0.5812	0.7164	64.96	0	0	0	0.5936	279.43
PV-DBOW	SGD	logistic	[500, 250, -]	0.575	0.7099	64.84	0	0	0	0.5936	185.65
PV-DBOW	SGD	logistic	[500, 250, 250]	0.5668	0.6978	63.67	0	0	0	0.5936	232.46
PV-DBOW	SGD	logistic	[250, 50, -]	0.5641	0.699	63.22	0	0	0	0.5936	175.06
PV-DBOW	SGD	logistic	[50, 50, 50]	0.5597	0.6841	61.86	0	0	0	0.5936	145.48
PV-DBOW	SGD	logistic	[500, 500, 500]	0.5592	0.6933	63.31	0	0	0	0.5936	255.18
PV-DBOW	SGD	logistic	[500, 100, 100]	0.5556	0.6884	62.08	0	0	0	0.5936	188.33
PV-DBOW	SGD	logistic	[500, 100, -]	0.5549	0.6852	62.92	0	0	0	0.5936	229.09
PV-DBOW	SGD	logistic	[500, 500, 250]	0.5502	0.6841	61.1	0	0	0	0.5936	228.38
PV-DBOW	SGD	logistic	[250, 250, 50]	0.5492	0.6778	62.2	0	0	0	0.5936	180.28
PV-DBOW	SGD	logistic	[500, 500, -]	0.5468	0.675	62.26	0	0	0	0.5936	219.43
PV-DBOW	SGD	logistic	[250, 50, 50]	0.5447	0.669	60.49	0	0	0	0.5936	170.71
PV-DBOW	SGD	logistic	[500, 500, 50]	0.542	0.6716	62.11	0	0	0	0.5936	201.66
PV-DBOW	SGD	logistic	[100, 100, -]	0.5372	0.6669	59.71	0	0	0	0.5936	152.96
PV-DBOW	SGD	logistic	[500, 250, 50]	0.5359	0.6736	58.91	0	0	0	0.5936	190.89
PV-DBOW	SGD	logistic	[250, 100, 100]	0.5346	0.6725	60.44	0	0	0	0.5936	176.08
PV-DBOW	SGD	logistic	[50, 50, -]	0.5326	0.6675	60.02	0	0	0	0.5936	148.01
PV-DBOW	SGD	logistic	[100, 100, 100]	0.5253	0.6518	57.63	0	0	0	0.5936	152.43
PV-DBOW	SGD	logistic	[500, 50, 50]	0.5234	0.662	56.69	0	0	0	0.5936	228.94
PV-DBOW	SGD	logistic	[500, 500, 100]	0.5196	0.6495	58.16	0	0	0	0.5936	216.98
PV-DBOW	SGD	logistic	[250, 250, 100]	0.5162	0.6514	57.24	0	0	0	0.5936	218.35
PV-DBOW	SGD	logistic	[500, 100, 50]	0.5111	0.6458	56.55	0	0	0	0.5936	232.01
PV-DBOW	SGD	logistic	[100, 100, 50]	0.5018	0.6349	55.44	0	0	0	0.5936	156.15
PV-DBOW	SGD	logistic	[500, 250, 100]	0.4927	0.6252	54.11	0	0	0	0.5936	192.67
PV-DBOW	SGD	logistic	[100, 50, 50]	0.4752	0.6062	51.15	0	0	0	0.5936	143.89
PV-DBOW	SGD	logistic	[250, 100, 50]	0.4724	0.6046	51.24	0	0	0	0.5936	171.7
PV-DBOW	SGD	logistic	[250, 100, -]	0.4675	0.5858	49.18	0	0	0	0.5936	157.22

Experiments using (q, c, avg_com_q) inputs – All results.

Appendix D

Further experiments with Syntactic & Metadata features

Category	Solver	Activation	Hidden Layer	MAP	AvgRec	MRR	P	R	F_1	Acc	Runtime(sec)

TABLE 9: Experiments using $(q, c, avg\_com_q, ft_{(q,c)})$ inputs – All results.

Category	Solver	Activation	Hidden Layer	MAP	AvgRec	MRR	P	R	F_1	Acc	Runtime(sec)

Experiments using $(q, c, avg\_com_q, ft_{(q,c)})$ inputs – All results.

Category	Solver	Activation	Hidden Layer	MAP	AvgRec	MRR	P	R	F_1	Acc	Runtime(sec)

Experiments using $(q, c, avg\_com_q, ft_{(q,c)})$ inputs – All results.

Bibliography

Barrón-Cedeno, A., Filice, S., Da San Martino, G., Joty, S. R., Màrquez, L., Nakov, P., and Moschitti, A. (2015). Thread-level information for comment classification in community question answering. In *ACL (2)*, pages 687–693. Citeseer.

Bogdanova, D. and Foster, J. (2016). This is how we do it: Answer reranking for open-domain how questions with paragraph vectors and minimal feature engineering.

Heilman, M. and Smith, N. A. (2010). Tree edit models for recognizing textual entailments, paraphrases, and answers to questions. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 1011–1019. Association for Computational Linguistics.

Jansen, P., Surdeanu, M., and Clark, P. (2014). Discourse complements lexical semantics for non-factoid answer reranking. In *ACL (1)*, pages 977–986.

Joty, S. R., Barrón-Cedeno, A., Da San Martino, G., Filice, S., Màrquez, L., Moschitti, A., and Nakov, P. (2015). Global thread-level inference for comment classification in community question answering. In *EMNLP*, pages 573–578.

Lafferty, J., McCallum, A., Pereira, F., et al. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the eighteenth international conference on machine learning, ICML*, volume 1, pages 282–289.

Le, Q. V. and Mikolov, T. (2014). Distributed representations of sentences and documents. In *ICML*, volume 14, pages 1188–1196.

Lin, X. Z. B. H. J. and Wang, Y. X. X. (2015). Icrc-hit: A deep learning based comment sequence labeling system for answer selection challenge. *SemEval-2015*, 210.

Mihaylov, T. and Nakov, P. (2016). Semanticz at semeval-2016 task 3: Ranking relevant answers in community question answering using semantic similarity based on fine-tuned word embeddings. *Proceedings of SemEval*, pages 879–886.

Nakov, P., Màrquez, L., Magdy, W., Moschitti, A., Glass, J., and Randeree, B. (2016). SemEval-2016 task 3: Community question answering. In *Proceedings of the 10th International Workshop on Semantic Evaluation*, SemEval '16, San Diego, California. Association for Computational Linguistics.

Rehurek, R. and Sojka, P. (2010). Software framework for topic modelling with large corpora. In *In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. Citeseer.

Severyn, A. and Moschitti, A. (2012). Structural relationships for large-scale learning of answer re-ranking. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 741–750. ACM.

Severyn, A. and Moschitti, A. (2013). Automatic feature engineering for answer selection and extraction. In *EMNLP*, volume 13, pages 458–467.

Shafiq Joty, L. and Nakov, P. (2016). Joint learning with global inference for comment classification in community question answering. In *Proceedings of NAACL-HLT*, pages 703–713.

Toutanova, K., Klein, D., Manning, C. D., and Singer, Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the 2003*

Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1, pages 173–180. Association for Computational Linguistics.

Tran, Q. H., Tran, V., Vu, T., Nguyen, M., and Pham, S. B. (2015). Jaist: Combining multiple features for answer selection in community question answering. In *Proceedings of the 9th International Workshop on Semantic Evaluation, SemEval*, volume 15, pages 215–219.

Wang, M. and Manning, C. D. (2010). Probabilistic tree-edit models with structured latent variables for textual entailment and question answering. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 1164–1172. Association for Computational Linguistics.

Wang, M., Smith, N. A., and Mitamura, T. (2007). What is the jeopardy model? a quasi-synchronous grammar for qa. In *EMNLP-CoNLL*, volume 7, pages 22–32.

Yao, X., Van Durme, B., Callison-Burch, C., and Clark, P. (2013). Answer extraction as sequence tagging with tree edit distance. In *HLT-NAACL*, pages 858–867. Citeseer.

Zhou, X., Hu, B., Chen, Q., Tang, B., and Wang, X. (2015). Answer sequence learning with neural networks for answer selection in community question answering. *arXiv preprint arXiv:1506.06490.*