Novelty Detection in Community Question Answering Forums

Tirthankar Ghosal
ÚFAL, MFF
Charles University, Czech Republic
ghosal@ufal.mff.cuni.cz

Vignesh Edithal
IIT Patna
Patna, India
edithal.cs14@iitp.ac.in

Tanik Saikh
IIT Patna
Patna, India
1821CS08@iitp.ac.in

Saprativa Bhattacharjee
Government Polytechnic Daman
Daman, India
saprativa.bhatt@gov.in

Asif Ekbal
IIT Patna
Patna, India
asif@iitp.ac.in

Pushpak Bhattacharyya
IIT Bombay
Mumbai, India
pb@cse.iitb.ac.in

Abstract

Community Question Answering (CQA) forums are a popular place for open-ended question-answers and discussions by the public. Popular question answering sites have become a one-stop solution for the diverse information-seeking needs of the netizens. However, it is quite common that users sometimes ask the same question which may have been posted before and appropriately answered by the community members. It is not uncommon for users to perform an improper search and pose the same question again. Community members start responding to that question; however, the answer may already have been proposed previously on a different thread. When two questions are getting at the same problem, community members (moderators) in the forum often flag one as a duplicate of the other to help route traffic to high-quality *novel* questions and their correct answers. However, they usually do this manually based on the novelty or redundancy of the question. In this work, we try to mitigate this problem of detecting semantically equivalent *non-novel* questions automatically and flagging those. We also propose an approach to identify *novel* questions from CQA forums so that only the novel questions and corresponding answers threads stay, and semantic duplicates are removed. We make use of a Dynamic Memory Network (DMN) to assimilate information from multiple source questions to answer whether a new question is *novel* or a *semantically equivalent* question already exists. We introduce a new dataset for semantic-level novelty detection on community question answering. Our proposed approach attains performance improvements of +6.64% in terms of accuracy and +9.09% in terms of averaged F1 score over recent textual novelty detection methods. We would make our newly created dataset and the proposed approach available at `https://github.com/edithal-14/DMN-Novelty`.

1 Introduction

Community Question Answering forums are a convenient source of information for web users. Users post their questions on the forum, and fellow community members help them with pointers or answers to their diverse information needs. However, with the rapid growth of content within those websites, *redundancy* has become a problem. Unaware of existing solutions, users pose questions on a new thread that may already have been resolved in an existing thread. Such activity is widespread, and unaware community members start interacting on the new thread with solutions/pointers that may have already been discussed on an older thread. Usually, CQA forums have moderators who flag such *non-novel* questions and route the users to the existing solutions on the forum. To channel traffic to high quality and *novel* questions and answers on the forum, it is important to weed out *redundancies* and *non-novelties*.

Most of the CQA platforms request the users to go through the previously asked questions before posting a new question and post their question only if a similar question does not exist. But expecting such due diligence from each and every user is a tall ask in itself. Moreover, manually tagging questions as duplicates also requires a lot of effort on the moderators. This necessitates automatic techniques for efficient identification of non-novel or duplicate questions. Such a system could be used to tag and merge the duplicates out of the already asked questions and alert the users to the existence of a potential duplicate question while attempting to post a new question.

We leverage the DMN+ model (Xiong et al., 2016a) which proposed modifications to the *input* module and *memory* module of the original DMN framework (Kumar et al., 2016). Also, we extend the usage of DMN from word embeddings to sentence embeddings. We use the Infersent sentence encoder (Conneau et al., 2017) trained on the semantically rich SNLI corpus using GloVe word vectors. Experimental results show significant improvements over two deep learning-based baselines and two existing comparable systems. The major contributions of our current work are (i). An improved DMN framework for semantic level novelty detection in CQA forums, and (ii). A novelty detection dataset parsed from publicly available Stack Exchange (STE) data dump.

The remainder of the paper is organized as follows: Section 2 describes the related works. We describe the dataset in detail in Section 3. Methodologies adopted in this article are described in Section 4. System evaluation results obtained along with comparisons and rigorous error analysis are presented in Section 5. Finally, we conclude this article with some future research directions in Section 6.

2 Related Works

The problem of novelty mining is a long-standing problem in Information Retrieval (IR). The task has matured through several shared tasks, workshops, etc. Starting from Wayne (1997) to the novelty detection track as a part of The Text Retrieval Conference (TREC) workshop organized by NIST in the year of 2002 (Voorhees, 2002), 2003 (Voorhees, 2003) and 2004 (Clarke et al., 2004). Allan et al. (2003) investigated the tasks defined in the TREC 2020 novelty track, i.e., given a topic and list of documents relevant to the topic. The task is to first find the relevant sentences from the collection of documents and then find the novel sentences from the collection of relevant sentences.

The task of novelty detection can also be performed by Textual Entailment (TE). This idea had taken shape through the shared tasks organized in the year 2006 (Bentivogli, 2010), and 2007 (Bentivogli, 2011). In this era of deep learning, the availability of high-quality benchmark data has been the key bottleneck in advancing the novelty detection field. Ghosal et al. (2018b) first came up with a considerable amount of data, namely *Document Level Novelty Detection (DLND) TAP-DLND 1.0* and later extended version *TAP-DLND 2.0* to feed data-hungry deep neural models and adapted the novelty detection task from sentence level to document level.

Duplicate questions detection is a sub-task in QA. A pair of questions are assumed to be similar if both the questions can be satisfied with the same answer. It is a challenging task in two aspects: *viz. (i). There are many ways of asking a question, i.e., (question paraphrasing)* and *(ii). Asking a question has an implicit purpose, so even if two different questions seem like looking for the same solution, they can have entirely different purposes*. Being able to detect such questions leads to an increase in the accuracy of a QA system. Bogdanova et al. (2015) defined two questions can be considered as duplicates if the same answer can answer them. Robertson et al. (1994) considered two questions as a bag of words and computes scores between them. Further, a weighted matching (Inverse Document Frequency) between the tokens is performed for determining if two questions are close. The system proposed in Prabowo and Budi Herwanto (2019) detects duplicate questions in QA Website. The proposed system is equipped with GloVe pre-trained word embeddings, Convolutional Neural Network (CNN), and siamese network. Labeled data is precious, Rücklé et al. (2019) tried to mitigate this problem. They framed this problem into a zero-shot setting.

In contrast to the prior work, we frame the problem of identifying duplicate questions from a different viewpoint. We employ a novelty detection approach for this. There are almost no studies that address this problem from a novelty perspective. We consider duplication as an opposite characteristic of novelty. So given a pair of input questions, novelty detection system predicts as *novel* (non-duplicate) or *non-novel* (duplicate). With this intuition, we carry out the experiments performed in this article. Our model is based on the Dynamic Memory Network (DMN) (Kumar et al., 2016) technique. DMN has a special property of having a *memory component* and *an attention mechanism*. This kind of network-aided technique has been used for QA, but we are unaware of any such methods that use DMN for novel question detection in online CQA forums.

3 Dataset Description

We create a new 'novel-question' detection dataset from the Stack Exchange family of websites. All of the community-contributed questions and answers, along with comments, upvotes, downvotes, tags, and other metadata from all these sites, are published publicly by the Stack Exchange network on the *Internet Archive*[1] regularly. As of this writing, the latest available data dump is from 06 June 2022. In the data dump, each question is connected to a set of related questions through the *PostLinks* entity. One of the attributes of this entity is the *LinkTypeId*, possible values of which are 1 and 3, depending on whether the present question is or is not a duplicate of the related question, respectively. This information comes from the act of marking questions as potential *duplicates* of related questions by the CQA forum moderators. The Stack Exchange (STE) novelty dataset is thus automatically created by extracting pairs of related questions, and the ground truth of novelty is established based on the value of the *LinkTypeId* for each such pair (NOVEL, NON-NOVEL). The dataset thus created spans 50 different topics with an average of 4312 question pairs for each topic resulting in 215667 question pairs. In Figure 1 we present the statistics for the top ten topics in the STE dataset.

4 Methodology

Our proposed novelty-detection method is based on Dynamic Memory Networks (DMN), which proved to be very effective in question answering.

4.1 Dataset Used

SNLI: SNLI (Bowman et al., 2015) is a widely used, well-recognized Natural Language Inference

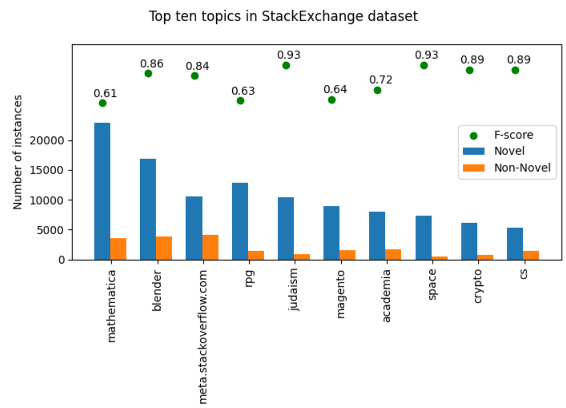

Figure 1: Statistics of top 10 topics by number of instances in the Stack Exchange dataset. The F-score corresponds to the performance of our DMN+ based model on that topic.

(NLI) corpus. It has 570k human-written English example pairs that are labeled manually for balanced classification with the labels *entailment, contradiction, and neutral*. We use SNLI to train our sentence encodings method (Infersent).

Stack Exchange Dataset: We parse the Stack Exchange data dump to create a Novelty detection dataset as described in Section 3 and run our proposed model on this newly created dataset.

4.2 Data pre-processing

We convert the source and target documents into a document matrix by encoding each sentence in the question into a 2048 dimensional vector. We employ *InferSent*(Conneau et al., 2017) for this purpose.

4.3 An improved DMN framework: DMN+

The basic DMN framework (Kumar and Irsoy, 2015) was proposed for dealing with QA problems. DMN consists of four independent modules *viz.* Input Module, Question Module, Episodic Memory Module, and Answer Module that can be improved independently. We re-implement the DMN+ model (Xiong et al., 2016b) for our task, which proposes to change the Input and Memory module of the original DMN framework.

Input Module: This module is responsible for encoding the inputs and adding contextual information to the inputs to produce fact vectors (f) used for further processing. The vanilla DMN model works on

[1]https://archive.org/details/stackexchange

the word level. In contrast, in the DMN+ model, we work on document level. We compute all sentences (contained in a document) embedding beforehand (using Infersent) and pass them through a Bi-GRU, and consider the output at each time-step is one fact. Let us consider our input consists of n sentences, and representation for each sentence from Infersent as $S_i, i = 1....n$; so the input is a sequence of n such representations as follows: $S = (s_1, s_2, s_3......s_n)$. This sequence is fed into a Bi-GRU, At each time step t, the network updates its hidden state $h_t = BiGRU(M[w_t], h_{t-1})$, where M is the embedding matrix and w_t is the sentence index of the t^{th} sentence of the input sequence. This module outputs the hidden states of the recurrent network. So we obtain f_n facts representation. The DMN+ approach solves two problems: (i). it allows for direct interaction between sentences that might be related to each other; at the word level, this type of interaction is difficult to capture, and (ii). a Bi-GRU allows for incorporating more contextual information from preceding and succeeding sentences, improving the quality of the fact vectors. We opt for Bi-GRU as it is lightweight, has fewer parameters, requires fewer computational resources, and is much faster than LSTM.

Question Module: We encode the target document's sentence representations via a Bi-GRU. Given the target question of T_Q sentences, hidden states for the question encoder at time t is given by $q_t = Bi - GRU(M[w_t^Q], q_{t1})$, M represents the embedding matrix as in the Input Module and w_t^Q represents the sentence index of the t^{th} sentence in the question. The word embedding matrix is shared across the input and question modules. This module produces the final hidden state of the Bi-GRU encoder, $q = q_{T_Q}$.

Memory Module: The outputs from the previous two modules are fed into the memory module. We attend the facts with respect to the question vector and the memory state (m) (initial memory state is the question vector itself) using a modified GRU called Attention GRU to create a context vector (c). The vanilla model sets the next memory state equal to this context vector, however the DMN+ model uses a memory update step to compute the next memory state using a ReLU layer over c, Q, and m. The final memory state after a defined number of memory state updates (also known as episodes or hops) is sent to the answer module. The final memory state should have enough information to answer the question after multiple attention passes over the incoming facts.

Attention GRU: The *attention GRU* is a traditional GRU with its update gate modified, which represents the importance of the incoming fact at the current time step. Following is a mathematical representation of a traditional GRU (here, x is the incoming fact, h is the hidden state of the GRU, i is the current time step, and $\bullet$ symbolizes the element-wise product)

$$u_i = \sigma(W^{(u)}x_i + U^{(u)}h_{i-1} + b^{(u)}) \tag{1}$$

$$r_i = \sigma(W^{(r)}x_i + U^{(r)}h_{i-1} + b^{(r)}) \tag{2}$$

$$\tilde{h_i} = tanh(Wx_i + r_i \bullet Uh_{i-1} + b^{(h)}) \tag{3}$$

$$h_i = u_i \bullet \tilde{h_i} + (1 - u_i) \bullet h_{i-1} \tag{4}$$

We replace the update gate u with an attention gate g as follows (here, f is the incoming fact, q is the question vector, m is the memory state, i is the current time step and t is the current memory update step)

$$z_i^t = [f_i \bullet q; f_i \bullet m^{t-1}; |f_i - q|; |f_i - m^{t-1}|] \tag{5}$$

$$Z_i^t = W^{(2)}tanh(W^{(1)}z_i^t + b^{(1)}) + b^{(2)} \tag{6}$$

$$g_i^t = \frac{exp(Z_i^t)}{\Sigma_{k=1}^{n} exp(Z_k^t)} \tag{7}$$

$$h_i = g_i^t \bullet \tilde{h_i} + (1 - g_i^t) \bullet h_{(i-1)} \tag{8}$$

Note that g is the output of a softmax layer which is essentially a probability distribution of how important each feature of the incoming fact is. The final hidden state of the Attention GRU is called the context vector (c).

Memory update step: This step is used only in the DMN+ model to compute the memory state based on the context vector. Note that the initial memory state m^0 is the question vector itself.

$$m^t = ReLU(W^t[m^{t-1}; c^t; q] + b) \tag{9}$$

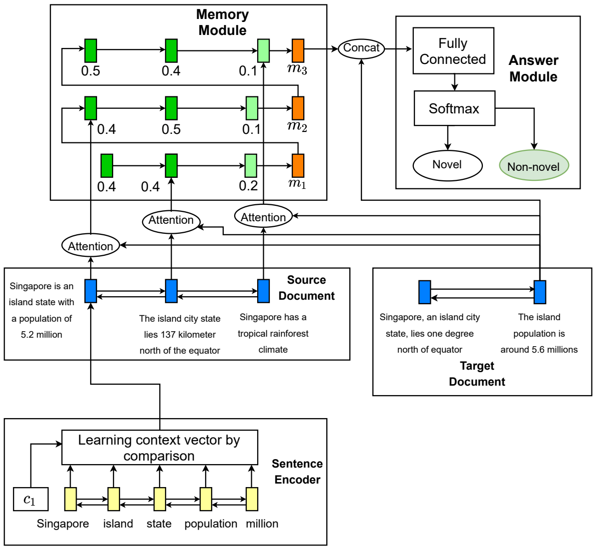

Figure 2: Novelty detection model based on DMN+ framework, colored blocks represent various tensors (e.g word/sentence embedding, attention and memory tensors).

Answer Module: This module differs according to the task at hand. We can use a linear layer with softmax for classification or one-word answer problems. We can use an RNN-based decoder network; if the answer is expected to be a sentence. We use the former approach, as our task is to classify the target document (question) as 'Novel' or 'Non-Novel.' The final memory state is concatenated with the question vector and passed through a ReLU layer followed by a softmax layer for classifying the target document as *novel* or *non-novel*.

5 Evaluation

5.1 Training

We use the Pytorch library for the implementation of the proposed model. We use the cross-entropy loss function since we have the softmax layer as the final layer to classify into Novel or Non-Novel classes. We initialize the model parameters using the Xavier initialization method, and we optimize them using Adam optimizer (Kingma and Ba, 2017). We decrease the learning rate (LR) by a factor of 10 on the validation accuracy plateau with the patience of 3 epochs (if validation accuracy does not improve by 1% in 3 epochs); this allows the optimizer to escape the local minima when it is stuck in one and prevents the model from over-fitting. We use a batch size of 32. We train the model for 25 epochs with an early stopping of 10 epochs (stop if the validation accuracy does not improve in 10 epochs). We chose the model with the best validation accuracy for testing. We perform model hyper-parameters tuning manually. The original DMN+ model uses three memory update iterations (hops) in its memory module; however, we observe that four hops provides the optimal result, further increasing the number of hops reduces the accuracy.

5.2 Results

We test our model on the newly introduced Stack Exchange (STE) novelty dataset. We compare the results with two deep learning-based models that we consider as the baselines and two other existing comparable systems, *viz.* (i) RDV-CNN (Ghosal et al., 2018a) and (ii) Decomposable attention-based model (Ghosal and Edithal, 2020). For each of the 50 topics in the dataset, we split the available document pairs into 80:20 ratios for training and testing, respectively. We then individually train and evaluate the model on each topic separately. Finally, we average the performance of the model across all the topics. We show a comparison of the results in *Table 1*. From the Table we conclude that leveraging the DMN+ framework to obtain a joint representation of a pair of source and target documents and then using it for Novelty judgment provides much better accuracy than simply passing the sentence embeddings through a BiLSTM.

5.3 Analysis

We now show an example **non-novel** (redundant) source/target document pair, where our proposed model can capture document redundancy. In contrast, its close competitor, the decomposable attention model (Ghosal and Edithal, 2020) fails to detect the redundancy and classifies the document pair incorrectly as Novel. We present the model's predictions along with a heatmap to explain the predictions in *Figure 3*.

Source **[s1]** Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus. **[s2]** Most people who fall sick with COVID-19 will experience mild to moderate symp-

Model	Accuracy	P (N)	R (N)	P (NN)	R (NN)	F1 (avg)
Doc2Vec + BiLSTM + MLP	64.88%	0.66	0.62	0.64	0.67	0.65
Inner attention + BiLSTM + MLP	70%	0.60	0.70	0.75	0.75	0.70
RDV-CNN (Ghosal et al., 2018a)	72%	0.71	0.61	0.70	0.71	0.72
Decomposable Attention (Ghosal and Edithal, 2020)	77%	0.82	0.68	0.73	0.85	0.77
Inner attention + DMN+ + MLP	**83.64%**	0.86	0.80	0.81	0.87	**0.84**

Table 1: Comparing our model with two baselines and two comparing systems (RDV-CNN and Decomposable attention) on the STE dataset. Average performance across all the 50 topics in the STE dataset. N→Novel, NN→Non-Novel

Topic	Link to Source Document	Target Document	Gold class	Predicted class	Comments
blender /08761.json	https://blender .stackexchange.com /questions/16267/i-dont-know-how-i-locked-view-offset-but-how-do-i-unlock-it	https://blender .stackexchange.com /questions/47935/how-to-remove-revert-blender-object-centre-view	Novel	Non-Novel	Domain specific parlance and overlapping named entities caused the model to predict Non-Novel
cs /00632.json	https://cs.stackexchange. com/questions/4800/the-order-of-growth-analysis-for-simple-loop	https://cs.stackexchange. com/questions/ 10813/decreasing-runs-of-inner-loop-in-outer-loop	Non-Novel	Novel	Inability of the model to understand mathematical formatting and programming language syntax causes wrong prediction
academia /00074.json	https://academia .stackexchange.com /questions/1190/what-are-some-good-project-management-tools-for-academics	https://academia .stackexchange.com /questions/1273/use-cases-of-org-mode-as-a-scientific-productivity-tool-for-academics-without-pr	Non-Novel	Novel	There are instance of incorrect gold labels in this dataset, since this dataset is created by an algorithm which uses linked posts and post metadata, it is prone to mistakes. In this case the given document pair is clearly Novel, however, the gold label states Non-Novel

Table 2: Error analysis of the DMN+ model using STE instances, topic corresponds to the topic specific Stack Exchange forum where the questions were asked. The comment column explains the cause of the miss-classification.

toms and recover without special treatment. **[s3]** The virus that causes COVID-19 is mainly transmitted through droplets generated when an infected person coughs, sneezes, or exhales. **[s4]** These droplets are too heavy to hang in the air, and quickly fall on floors or surfaces. **[s5]** You can be infected by breathing in the virus if you are within close proximity of someone who has COVID-19, or by touching a contaminated surface and then your eyes, nose or mouth. **[s6]** You can reduce your chances of being infected or spreading COVID-19 by regularly and thoroughly cleaning your hands with an alcohol based hand rub or wash them with soap and water. **[s7]** Washing your hands with soap and water or using alcohol based hand rub kills viruses that may be on your hands.

Target (Non-Novel) **[t1]** COVID-19 symptoms are usually mild and begin gradually. **[t2]** Some people become infected but don't develop any symptoms and don't feel unwell. **[t3]** Most people (about 80%) recover from the disease without needing special treatment. **[t4]** Older people, and those with underlying medical problems like high blood pressure, heart problems or diabetes, are more likely to develop serious illness.

5.4 Error Analysis

In *Table 2* we present a few instances from the STE dataset wherein our model failed to classify the novelty of the document pair correctly. In some cases, our model predicts a pair of source and target documents as non-novel or duplicate due to a signif-

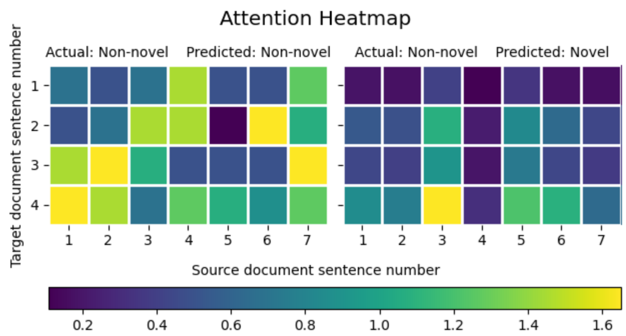

Figure 3: Heatmap denoting attention values between source and target sentences for the document pair mentioned in the Analysis section. *Left:* attention values from the last hop of the episodic memory module in our DMN based model. *Right:* attention values from the sentence comparison step of the Decomposable attention model.

icant overlap of named entities between the two. For instance, consider the pair for the topic *blender*, our model incorrectly predicts it as a non-novel pair mainly because of the use of domain-specific terminologies and due to the occurrence of named entities such as *blender*, *numpad*, *view* and others in both the source and the target.

In some other cases, the model makes wrong predictions due to the inability to understand the mathematical formatting and programming language syntax used in the questions' text. An example of such a pair is the one under the topic of *cs*, in which both the source and target consist of code snippets and inline mathematical expressions. Moreover, there are also instances of erroneous duplicate tagging of the question pairs in the gold set. This might be attributable to the subjective nature of the job, as often in CQA forums, the questions marked as duplicates and merged by moderators are later unmerged after reviewing the appeals from the original user who posted the question or even fellow community members. This is illustrated by the pair from the topic *academia*.

6 Conclusion and Future Work

In this work, we address the problem of duplicate question identification in community question-answering forums from the perspective of textual novelty detection. To the best of our knowledge, no prior work has leveraged textual novelty detection for tackling this problem. We use a deep Dynamic Memory Network, specifically the DMN+ for assimilating information from multiple source questions to detect the novelty of a target question at the semantic-level. Our method outperforms the deep learning-based baselines and recently proposed textual novelty detection methods. We also propose a new dataset consisting of 215K novel and non-novel question pairs over 50 different topics. We automatically create this dataset from the publicly available data dumps of the Stack Exchange network of websites.

In the future, we would like to validate our approach on other CQA forums such as Reddit and Quora questions/posts. We would also like to investigate how the recent large contextual language models would perform for this problem. We envisage that our novel investigation of associating textual novelty detection to detect semantic duplicates on the web would aid in several downstream tasks to alleviate the quality of information available.

References

James Allan, Courtney Wade, and Alvaro Bolivar. 2003. Retrieval and novelty detection at the sentence level. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, SIGIR '03, page 314–321, New York, NY, USA. Association for Computing Machinery.

Magnini B. Dagan I. Dang H.T. Giampiccolo D. Bentivogli, L. 2010. The Sixth PASCAL Recognizing Textual Entailment Challenge. In *Proceedings of the Text Analysis Conference (TAC 2010), November 15-16, 2010 National Institute of Standards and Technology Gaithersburg, Maryland, USA.*

Clark P. Dagan I. Dang H. T. Giampiccolo D. Bentivogli, L. 2011. The Seventh PASCAL Recognizing Textual Entailment Challenge. In *In TAC 2011 Notebook Proceedings, November 14-15, 2011, Gaithersburg, Maryland, USA.*

Dasha Bogdanova, Cícero dos Santos, Luciano Barbosa, and Bianca Zadrozny. 2015. Detecting semantically equivalent questions in online user forums. In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning*, pages 123–131, Beijing, China, July. Association for Computational Linguistics.

Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. 2015. A large annotated corpus for learning natural language inference. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, Lisbon, Portugal, September 17-21, 2015*, pages 632–642.

Charles LA Clarke, Nick Craswell, and Ian Soboroff. 2004. Overview of the trec 2004 terabyte track. In *TREC*, volume 4, page 74.

Alexis Conneau, Douwe Kiela, Holger Schwenk, Loïc Barrault, and Antoine Bordes. 2017. Supervised learning of universal sentence representations from natural language inference data. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 670–680, Copenhagen, Denmark, September. Association for Computational Linguistics.

Tirthankar Ghosal and Vignesh Edithal. 2020. Is your document novel? let attention guide you. an attention-based model for document-level novelty detection. apr.

Tirthankar Ghosal, Vignesh Edithal, Asif Ekbal, Pushpak Bhattacharyya, George Tsatsaronis, and Srinivasa Satya Sameer Kumar Chivukula. 2018a. Novelty goes deep. a deep neural solution to document level novelty detection. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2802–2813, Santa Fe, New Mexico, USA, August. Association for Computational Linguistics.

Tirthankar Ghosal, Amitra Salam, Swati Tiwari, Asif Ekbal, and Pushpak Bhattacharyya. 2018b. TAP-DLND 1.0 : A corpus for document level novelty detection. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May. European Language Resources Association (ELRA).

Diederik P. Kingma and Jimmy Ba. 2017. Adam: A method for stochastic optimization.

Ankit Kumar and Ozan Irsoy. 2015. Ask me anything: Dynamic memory networks for natural language processing. arXiv:1506.07285. version 5.

Ankit Kumar, Ozan Irsoy, Peter Ondruska, Mohit Iyyer, James Bradbury, Ishaan Gulrajani, Victor Zhong, Romain Paulus, and Richard Socher. 2016. Ask me anything: Dynamic memory networks for natural language processing. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, ICML'16, page 1378–1387.

Damar Adi Prabowo and Guntur Budi Herwanto. 2019. Duplicate question detection in question answer website using convolutional neural network. In *2019 5th International Conference on Science and Technology (ICST)*, volume 1, pages 1–6.

Stephen E. Robertson, Steve Walker, Susan Jones, Micheline Hancock-Beaulieu, and Mike Gatford. 1994. Okapi at TREC-3. In Donna K. Harman, editor, *Proceedings of The Third Text REtrieval Conference, TREC 1994, Gaithersburg, Maryland, USA, November 2-4, 1994*, volume 500-225 of *NIST Special Publication*, pages 109–126. National Institute of Standards and Technology (NIST).

Andreas Rücklé, Nafise Sadat Moosavi, and Iryna Gurevych. 2019. Neural duplicate question detection without labeled training data. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1607–1617, Hong Kong, China, November. Association for Computational Linguistics.

Ellen M Voorhees. 2002. Overview of trec 2002. In *Trec*.

Ellen M Voorhees. 2003. Overview of trec 2003. In *TREC*, pages 1–13.

Charles L Wayne. 1997. Topic detection and tracking (tdt). In *Workshop held at the University of Maryland on*, volume 27, page 28. Citeseer.

Caiming Xiong, Stephen Merity, and Richard Socher. 2016a. Dynamic memory networks for visual and textual question answering. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, ICML'16, page 2397–2406. JMLR.org.

Caiming Xiong, Stephen Merity, and Richard Socher. 2016b. Dynamic memory networks for visual and textual question answering. *CoRR*, abs/1603.01417.