

---

# Classifying Reddit comments by subreddit

---

**Jee Ian Tam**

Department of Computer Science  
Stanford University  
Stanford, CA 94305  
jeetam@stanford.edu

## Abstract

Reddit.com is a website that is primarily organized by communities called subreddits, where users can post comments to. As subreddits can have very different cultures, we aim to classify comments by subreddit as a means of sentiment analysis. We use a publicly available reddit comment dataset over the year of 2016 and perform a classification on a selection of 20 subreddits among the top 50 by total comment word counts. We investigate the Recurrent Convolutional Neural Network (CNN) and Attention-Pooled CNN as more advanced models in the classification task. We find that the Recurrent CNN with GRU performs the best out of all the models tested, achieving an F1 score of 0.53 and a Cohen's Kappa score of 0.502. Out of all the subreddits, we find that the Recurrent CNN model gives the highest F1 score on the "anime" subreddit. Quantitative and qualitative analysis of the Recurrent CNN with GRU suggests that the model is able to reasonably approach the task. Further work suggests that the trained model can also be used to cluster subreddits based on comment similarity.

## 1 Introduction

Reddit.com is a website that is primarily organized by communities called subreddits. Users can post threads to subreddits, and can also post comments inside those threads. Different subreddits focus on different content and have different moderation cultures, resulting in a consistent culture within each subreddit, but different subreddits can have drastically different cultures. Thus, we expect that the comments of a subreddit will be representative of the culture of that subreddit due to the subreddit's demographic and moderation. For example, we would expect that the comments of the "WritingPrompts" subreddit (a subreddit where comments are short stories in response to a prompt of a thread) would be very different than the comments of the "politics" subreddit.

This task is a sentence/sequence classification problem, similar in nature to that of sentence sentiment classification. However, this task is slightly more complex than sentiment classification in that the classes in sentiment classification are usually over a spectrum ("negative", "neutral", "positive"), whereas subreddits do not necessarily lie on some sort of spectrum. However, subreddits can be grouped in clusters of similar culture. Indeed, this has been done at [1], [2] and [3]. We will later explain our method of choosing which subreddits as prediction classes in section 3.1

We will explore various models in the literature that have been used for sentence/sequence classification. These models are described in more detail in section 3. The general approach will be to use GloVe word embeddings and a fixed window of words as input to a deep neural network that either uses recurrent and/or convolutional units with a softmax layer as the final layer.

## 2 Background / Related Work

Previous similar work using Reddit data based on subreddits can be found at [4], where subreddit recommendations are predicted given a user’s existing preferred subreddit. Our work differs in that we do not work with user data, but rather predict the subreddit that a comment was posted to.

Sentence / Sequence classification is a common task that has been well-explored in the literature. One of the simpler yet effective models is [5] by Kim Yoon. The CNN model employed is a shallow CNN model with multiple filters. Max-over-time pooling (pooling over the entire sequence window) is used to summarize each filter.

An extension to the work of [5] is that of [6], where the CNN model is extended to include a recurrent neural network with the convolutional aspect of the model. The motivation behind this is that a CNN model can capture local contexts well around any particular word location, but has more difficulty capture sequential relationships of a word in a sentence relative to the words before and after it along the entire sentence. [6] uses two recurrent network layers (one forwards, one backwards) to provide a left and right convolutional context for each word in the window. Max-pooling is then used to obtain a final feature representation for each word.

Another extension is that of [7], where attention-pooling is used as a form of attention to reduce the convolutional filters to word features for each word. The model uses the cosine similarity between the output of a bidirectional LSTM and the convolutional filters to compute attention for each word, and combine the attention with the convolutional filters to produce a feature representation for each word.

## 3 Approach

### 3.1 Dataset

We first download a dataset of reddit comments, made available by Baumgartner at [9]. Within the initial dataset, we count over 280,000 different subreddits. In order to ease the analysis of results and make the project simpler, we determine the top 50 subreddits by comment word count and select 20 of the most diverse ones among them. The reasoning behind this is that many subreddits do not actually have many comments, and so we want to select subreddits with the most comments (in terms of word count) so that each subreddit (class) has enough data for training. For subreddit selection, we select subreddits that are deemed to have different cultures from other selected subreddits, and we also try to select subreddits such that most of the classes have roughly the same amount of comments. The balance of the classes chosen to be included in the final dataset are shown in table 2 :

The final dataset consists of 21.9 million comments, and 2.86 billion words over 1-year period of 2016. We separate the dataset into 70% for training, 15% for validation and 15% for final testing.

### 3.2 Vocabulary

We use GloVe vectors [8] to represent the embeddings of words into a higher-dimensional space. We particular use the 840B.300d word vectors made available at [8]. We process the dataset to the determine a vocabulary from the dataset, and then filter out words that are not present in the vocabulary of the pretrained word vector embeddings and also have a count of less than 100. This is to filter out infrequent words or errors that were made during processing of the dataset. The filtered out words are replaced with the UNK token. Words that were not initially in the vocabulary of the pretrained word embeddings were initialized as zero vectors. The final vocabulary consists of 746,220 words.

### 3.3 Models

In this section we describe the models that we investigate with regards to the classification task described above. All models were implemented in Tensorflow [10].

#### 3.3.1 Baseline Models

We use two models as reference baseline models to evaluate the more advanced models against.

The first baseline model is a simple Bag-Of-Words (BOW) model where the sentence feature is the mean of all word features (embeddings). A softmax layer is then put on top of it for classification.

The second baseline model is a Dense Window model where we concatenate all of the word embeddings together to form a sentence feature, and put a softmax layer on top of it for classification.

#### 3.3.2 Advanced Models

We first implement the convolutional neural network (CNN) model discussed in [5], which uses convolutional filters to summarize the local context around each word of a sentence. We implement the CNN-non-static version, which is the version where the model and word embeddings are trained, but only a single channel is used (instead of multi-channel where two versions of word embeddings are loaded - One of them is trained and the other is frozen. Adjustments are then made depending on how different the trained and frozen embeddings are). The model is shown in figure 1, taken from [5]. Our implementation uses window sizes of 5,7 and 10, with each window having 100 feature maps each.

We next implement the recurrent convolutional neural network (RCNN) model discussed in [6], which uses two recurrent neural networks to provide features representing contexts to the left and right of each word. The context features are concatenated with the word features, and a linear-transform followed by an activation function is applied to each of those features to perform a convolution over a word with the contexts to the left and right of it. Max-pooling-over-time is done to reduce the convolution filters to a single sentence feature. [6] use simple feed-forward units for the recurrent units, but we will also investigate the performance of the model when GRU units are used instead. The model architecture is show in figure 2. Our implementation uses a recurrent unit state size of 200, and a hidden state size (dimension of max-pooling layer) of 100.

The last model that we investigate is the attention-pooled CNN (ACNN) model as presented in [7], which uses a bidirectional LSTM to provide attention to the convolutional filters for each word. This is done by having the dimensions of the LSTM outputs be the same as the number of convolutional feature maps, and computing attention weights for each word via the cosine similarity between the LSTM outputs and the convolutional feature map outputs. The sentence representation is a contatenation of the bidirectional LSTM outputs with the attention-weighted sum of convolutional filter outputs over the input sentence for each window size. We do not make any significant modifications to the model. The model architecture is shown in figure 3. Our implementation uses convolutional window sizes of 3,5 and 7 with 100 feature maps each. The bi-LSTM has a recurrent unit state size of 100.

### 3.4 Evaluation

To evaluate and summarize the performance of each of the models presented above, we calculate the following metrics at the end of each epoch :

F1 score, which is the geometric mean of precision (proportion of selected subreddits that were correct) and recall (proportion of correct subreddits that were selected). We calculate precision, recall and F1 score for each class, and summarize by averaging those scores across all classes.

Cross-Entropy loss is used as the loss function to train all of the models, and represents a measure of the Kullback-Leibler divergence between the distribution of the class labels in the (training) data and the distribution as learnt by the model.

Cohen’s Kappa [11] is a measure the agreement between raters on items. In this context, it measures how well the predictions of the model agree with the ground truth labels. It is a more robust measure than a macro accuracy calculation, as it takes into account the possibility of an agreement happening by chance. It is defined as  $\kappa = \frac{p_o - p_e}{1 - p_e}$ , where  $p_o$  is the relative observed agreement between the model and the ground truth, and  $p_e$  is the probability of chance agreement.

### 3.5 Training Details

As mentioned above, we use cross-entropy loss to train all of the implemented models. We explore various training strategies, including SGD with momentum [12] on a learning schedule, Adam[13], AdaDelta [14] and RMSProp [15]. However, we find that Adam gives the best results with slightly easier training, and so the results below are presented for models trained with Adam using an initial learning rate of 0.0001 with a batch size of 2048.

For all models, we use a fixed sentence window of 50, padding sentences that are shorter and truncating sentences that are longer. The initial GloVe word embeddings are trained (backpropagated into) simultaneously with all models. We experimented with holding the word embeddings fixed, but as expected model performance suffered as a result. One important detail is that we train all word embeddings except the one corresponding to the PAD token (which is initialized to the zero vector). The reasoning is that some sentences will contain many PAD tokens, and it should not contribute in any way to the model’s classification decision.

For the RCNN and ACNN models, we clip gradients to a max of 100.

Note that we don’t use any form of regularization for our models, due to the finding that our models were underfitting more than overfitting during training by comparing validation loss over training epochs.

## 4 Experiments and Results

### 4.1 Performance of various models

The performance of the various models on the test set is shown in table 1 :

We see that the RCNN with GRU model performs the best out of all the models tested. This is slightly surprising as we expected the ACNN model to perform the best due to having a more complex model using attention via bidirectional LSTMs with convolutional layers, whereas the RCNN with GRU consists of a simpler structure. This could be due to the ACNN being harder to train due to larger model complexity, and thus not being able to achieve a better performance within the same number of training epochs as the RCNN. This is a possibility that can be inferred by looking at the convergence plot in figure 4.

Another possibility is that the RCNN model is actually able to capture a larger context than the ACNN model. [6] claim that the RCNN model structure allows the model to capture (and

Model	Precision	Recall	F1	Loss	Cohen Kappa
Random Guesser	0.07	0.05	0.05	3.035	-1.03e-5
Majority Classifier	0.01	0.12	0.03	2.996	0.0
Baseline - BOW	0.52	0.49	0.48	1.708	0.4432
Baseline - Dense Window	0.51	0.51	0.50	1.627	0.4708
CNN	0.52	0.53	0.52	1.548	0.4860
RCNN with linear recurrent unit	0.51	0.52	0.51	1.570	0.4812
<b>RCNN with GRU</b>	<b>0.54</b>	<b>0.54</b>	<b>0.53</b>	<b>1.493</b>	<b>0.5022</b>
ACNN	0.52	0.52	0.51	1.556	0.4856

Table 1: Performance metrics for various models

pool) the whole sentence context to the left and right of each word, whereas the ACNN model makes use of attention but only with local convolutional filters, and does not perform the same global-sentence-context capturing that the RCNN model does. It could be possible that the RCNN model’s ability to capture global context (combined with the GRU) is better than the ACNN model’s ability to provide attention to local contexts, thus giving the observed results.

The convergence plot (of validation loss) of the various models over training epochs is shown in figure 4 :

We also note that the Dense Window baseline model is easier to train and performs surprisingly well, just slightly worse than all of the advanced models that were trained within the time available. However, it seems that the dense window baseline model starts to overfit after a couple of epochs, and this could be improved with regularization such as Dropout. In comparison, the more complex models show no sign of overfitting, though they are harder to train and take longer to converge to an optimum.

## 4.2 Analysis of best model (RCNN with GRU)

We now analyze the performance of the best-performing model, the RCNN model with GRU.

### 4.2.1 Confusion Matrices and F1 scores

The confusion matrices (Absolute, row-normalized and column-normalized) for the RCNN with GRU model on the test set are shown in figures 5, 6 and 7 :

The F1 scores for each class are given in table 3.

Looking at the confusion matrices, we note the following :

- The model has difficulty distinguishing the comments within clusters such as the political cluster (“The\_Donald”, “SandersForPresident”, “PoliticalDiscussion” and “worldnews”) and the relationship cluster (“AskMen”, “AskWomen” and “relationships”). This suggests a method of using comment similarity to cluster subreddits together - If the model confuses a subreddit for another in terms of comments, it suggests that those subreddits are similar to each other.
- The model is also able to classify comments subreddits that are culturally different from all of the other subreddits, as it achieves a high precision and recall (> 75%) on the “anime” and “Fitness” subreddits, which are all very different from all the other subreddits in terms of culture and content. This can also be seen from the subreddits that the model achieves the highest F1 scores on in table 3.
- The “worldnews” and “todayilearned” subreddits seem to give the model the most confusion in terms of precision and recall relative to the other subreddits. This can be seen by

looking at the rows for those subreddits in the column-normalized confusion matrix (figure 6) and the columns for those subreddits in the row-normalized confusion matrix (figure 7). This could be because both of those subreddits have comments that span a large variety of topics that would be encountered in all of the other subreddits.

#### 4.2.2 Score-maximizing comment for each subreddit

In order to visualize the model, we next find the comments in the test set that maximize the scores for each class. The score is the activation of the model before it is passed through the softmax layer for classification. The class-maximizing comments are shown in tables 4 and 5. We had to perform some filtering of the comments to filter out comments that were made by bots, as the model learnt to assign the maximal scores for comments made by bots in each subreddit (Some subreddits use bots for automatic moderation and user notification). This is probably because the comments made by bots are the most consistent over time, and so the model learnt to associate comments made by a subreddit's bots to that particular subreddit. We filter out high-scoring comments containing the words "submission", "removed", "bot", "rule", and "automatically".

We see that the comments that maximize the score for each subreddit are representative (perhaps too stereotypically so?) of the content and culture that one would find in that subreddit. For example, the score-maximizing comment for the "AskMen" subreddit is about friends with benefits, the one for the "AskWomen" subreddit is about lipstick, the one for the "gaming" subreddit is a complaint about a game developer, the one for "personalfinance" is one about money management, the one for the "movies" subreddit is talking about the plot line of a movie and so on. The model even picks up on a meme that is prevalent on "The\_Donald" subreddit.

#### 4.2.3 Highest-scoring incorrect prediction for each subreddit

We would also like to make sure that the model makes reasonable mistakes. That is, if the model is very confident about its prediction but is actually incorrect, we want to make sure that the model's mistake is something that a human might make. The most confident incorrect predictions are shown in tables 6 and 7. Here, the confidence is measured by the score that the model assigns to each class, where the score is the activation of the model before it is passed through the softmax layer for classification.

We see that the model makes reasonable mistakes for its most confident incorrect predictions. For example, a comment in the "explainlikeimfive" subreddit explaining how to pronounce japanese words is mistakenly predicted to be in the "anime" subreddit. In another example, a comment in the "SandersForPresident" subreddit about insurance and income is mistakenly predicted to be in the "personalfinance" subreddit. In general the incorrect predictions mirror the statistics presented in the confusion matrices.

## 5 Conclusion

We investigate the performance of various text classification models on the task of classifying Reddit comments by subreddit. We find that the Recurrent Convolutional Neural Network with GRU model gives the best performance, demonstrating a clear gains in performance over the baseline models. We quantitatively and qualitatively verify that the model performs adequately at the task.

One idea for future work would be to refine the dataset to incorporate clusters of subreddits instead of subreddits themselves. For example, instead of having all of the video gaming-related subreddits be individually labeled, we can instead group them under a "gaming" label. Doing so would allow us to incorporate more subreddits into the task without needing to add a large amount of classes as outputs. As mentioned above, it would be possible to use the confusion matrices to cluster

subreddits together based on some threshold - This is something that is worth exploring further.

Given more time, another avenue for exploration would be to train the advanced models for longer to see how much better they can perform over the baseline. The advanced models as trained only give a small performance gain over the baseline models. We have observed that the advanced models are hard to train, in that loss decreases slowly and the models do not come close to overfitting the training dataset despite several epochs of training. More training time or a better optimization strategy / techniques might be able to give a higher performance for the advanced models.

Lastly, it would be interesting to apply the model to a separate dataset that is not based on Reddit data, and use the separate dataset to indirectly infer additional information about subreddits. For example, we could apply the model to the Stanford Sentiment Treebank [16], and use the sentiment labels to infer which subreddits are more associated with a positive sentiment, and which ones are more associated with a negative sentiment. For example, if the model classifies a sentence as belonging to the "politics" subreddit, and that sentence has a negative sentiment, that we can indirectly infer that the "politics" subreddit is more aligned to negative sentiment.

## References

- [1] <http://www.arimorcos.com/blog/post/Clustering>
- [2] <http://dwieker.github.io/Reddit/>
- [3] <http://www.shorttails.io/interactive-map-of-reddit-and-subreddit-similarity-calculator/>
- [4] Sundaresan, Hsu and Chang. Subreddit recommendations within Reddit communities. <http://snap.stanford.edu/class/cs224w-2014/projects2014/cs224w-16-final.pdf>
- [5] Kim, Yoon. "Convolutional neural networks for sentence classification." arXiv preprint arXiv:1408.5882 (2014).
- [6] Lai, Siwei, et al. "Recurrent Convolutional Neural Networks for Text Classification." AAAI. Vol. 333. 2015.
- [7] Er, Meng Joo, et al. "Attention pooling-based convolutional neural network for sentence modelling." Information Sciences 373 (2016): 388-403.
- [8] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation.
- [9] Baumgartner, Jason. "https://files.pushshift.io/"
- [10] Abadi, Martn, et al. "Tensorflow: Large-scale machine learning on heterogeneous distributed systems." arXiv preprint arXiv:1603.04467 (2016).
- [11] Byrt, Ted, Janet Bishop, and John B. Carlin. "Bias, prevalence and kappa." Journal of clinical epidemiology 46.5 (1993): 423-429.
- [12] Qian, Ning. "On the momentum term in gradient descent learning algorithms." Neural networks 12.1 (1999): 145-151.
- [13] Kingma, Diederik, and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).
- [14] Zeiler, Matthew D. "ADADELTA: an adaptive learning rate method." arXiv preprint arXiv:1212.5701 (2012).
- [15] Tieleman, Tijmen, and Geoffrey Hinton. "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude." COURSE: Neural networks for machine learning 4.2 (2012).
- [16] Socher, Richard, et al. "Recursive deep models for semantic compositionality over a sentiment treebank." Proceedings of the conference on empirical methods in natural language processing (EMNLP). Vol. 1631. 2013.

## 6 Appendix

### 6.1 Tables

Subreddit Name	% of dataset (comments)
The_Donald	9.03
SandersForPresident	4.75
PoliticalDiscussion	4.83
worldnews	6.46
movies	5.34
anime	6.80
gaming	3.79
Fitness	2.89
personalfinance	3.83
technology	3.73
relationships	5.49
WTF	2.53
AskMen	4.38
AskWomen	3.14
Showerthoughts	2.92
explainlikeimfive	4.99
todayilearned	5.42
changemyview	6.16
atheism	4.92
WritingPrompts	8.57

Table 2: Dataset class distribution

Subreddit Name	F1 Score
anime	0.77
Fitness	0.76
relationships	0.72
movies	0.69
personalfinance	0.69
WritingPrompts	0.68
gaming	0.63
The_Donald	0.61
SandersForPresident	0.55
worldnews	0.55
technology	0.45
PoliticalDiscussion	0.43
AskWomen	0.42
atheism	0.40
explainlikeimfive	0.38
WTF	0.37
todayilearned	0.31
changemyview	0.31
AskMen	0.30
Showerthoughts	0.24

Table 3: Class F1 Scores for RCNN with GRU on test set



Subreddit Name	Score-maximizing comment
The_Donald	if if if if if if if if if if if if if if if if if if okie doke okie doke okie doke okie doke okie doke if if if if if ...
SandersForPresident	... this needs to happen make it happen we need incentive and a goal is an incentive it helps reduce the bystander effect and it gives a tangible goal other than everybody phonebank ill be bugging you until this happens or until you respond
PoliticalDiscussion	quinnipiac senate for florida north carolina ohio and pennsylvania florida rubio rinc 48 2 murphy d 44 1 north carolina ross d 46 3 burr rinc 46 3 ohio portman rinc 55 4 strickland d 38 2 pennsylvania toomey rinc 50 4 mcginty d 42 3 in the north carolina governors race roy cooper leads gov pat mccrory by 2 cooper d 48 3 mccrory rinc 46 2 changes are from the september 9th q
worldnews	unian is a ukrainian media company with a large stake owned by ihor <unk> who also finances gt the aidar battalion is publicly backed by ukrainian oligarch ihor <unk> who also allegedly funds the azov donbas dnepr 1 dnepr 2 volunteer battalions operating under orders from kiev last spring <unk> offered a bounty of 10000 of his own money for <unk> russian <unk> gt armed volunteers ...
movies	... but the general meeting louise at the party doesnt make sense he was just about to attack the heptapods when louise picks up the satphone if she doesnt know the generals phone number at that point how does the natural flow of events lead us to the party where everythings resolved its paradoxical she doesnt get any of this information until the future but the future doesnt happen that way unless she knows the information
anime	suwa didnt read it until after the fact and naho just didnt believe it in full at that point as for the other three then it could be a mix of the two azusa was the one to invite kakeru to hang out and takako instead after along with suwa they probably didnt read their letters at that point...
gaming	starforge founder customer here words wont describe how sorry i am originally it showed such promise then they made it blatantly obvious that they didnt give two fucks about their game and ditched it to release reign of kings a month or so later after the release of starforge
Fitness	phul is a methodology not a program 10 people will have 10 different phul programs upper what you do on those days is completely up to you and as a newbie you can completely miss the mark on a complete workout sl icf etc are all programs with specific exercises on specific days
personalfinance	can you explain your 15k efund a little more is that not included in your net worth screenshot what was your efund at before starting ynab im going to make an assumption that your efund has grown it seems like youre doing pretty well but not accounting for efund also if im reading this correctly you have 185k total ...
technology	what hdhomerun does not allow is for <unk> channels to be streamed as i mentioned i have yet to find a solution that supports all channels including <unk> channels aside from the dinosaur that is wmc unless hdhomerun manages to get the cable industry to agree on allowing <unk> channels to stream then im not onboard
relationships	bil owes your husband an apology it really is that simple even if bil was truly offended by husbands comment how he dealt with it is completely inappropriate and juvenile not talking for a week sure months nope your husband should tell your mil that your brother has broken his trust...
WTF	if that was a real wasps nest and not a mud daubers nest you would have been stung at least a once if there was a wasp on the nest which there is usually at least one unless it is in the very beginning stage mud daubers are generally the <unk> of the wasp world not anywhere near as aggressive as other wasps
AskMen	fwb barely ever works imo its possible that she stayed casual with you because she liked you too much to just not be with you at all seriously ive never been in a fwb thing before but i was in contact long distance sexting etc for a long time with a guy who didnt want a girlfriend i thought i was okay with it the reason being that i liked him a lot it took months for me to realize he really wasnt going to change his mind ...
AskWomen	colourpop lippie stix i hate the name but these things are fantastic i wish i could own them in every single color rimmel lasting finish lipstick by kate moss for under 10 id recommend the nyx natural shadow palette coming in at slightly higher but worth it in my opinion and still on the low end of things would be the neutrogena foundations and powders
Showerthoughts	this isnt a legit shower thought even if as you commented elsewhere that you thought of it in the shower i thought in the shower this morning i need more soap if i were to then come here post that thought get downvoted then claim theres too many <unk> id be annoying as fuck
explainlikeimfive	this post is not asking for a laymanfriendly explanation to something complicated or technical so it doesnt belong here entirely subjective questions generally belong in <unk> post

Table 4: Score-maximizing test-set comments for each class

Subreddit Name	Score-maximizing comment
todayilearned	might even didnt say it would be likely but at least it might be possible and contrary to the current state of affairs where they ticket groden and then he defends it in court if they would deny groden some required license and he went to court to appeal for it there almost certainly are decisions made by officials you cant appeal in any way btw as screwed up as that is it forces groden to take the initiative in the courts which would probably cost him more in legal fees and also changes the narrative im not agreeing with the city here just wondering why theyre being so dumb about the whole deal
changemyview	i cant find any strong evidence for which major cities will or wont be susceptible to disastrous flooding therefore any argument i could make against you would merely be hearsay if what you say is true that the flooding of these cities would cause a global crisis that would directly affect me then i would also have to say that my view should change
atheism	read this most likely the ffrf would if you request send a letter to your counselor giving the legal reasons why you should be excused from attending meetings with religious content at the ffrf website you can read examples of various kinds of letters that the ffrf has sent to various government officials who have violated separation of church and state in various ways the letters are quite impressive here are a couple of examples heres a story of a recent victory so you can see how these things work
WritingPrompts	dropcap empty as it had ever been in the short time which id possessed the bottle of whiskey thick glass tumbled with a hollow clunk and clamor as the bottle rolled free from the <unk> surface of my worn folding table a dramatic echo of its landing soon accompanied by the following journey along the floor carried with unpleasant effect in my ears only ceasing its complaints by an excessively loud impact at the base of the metal legs of the table and the rug whatever normal circumstances that could have spawned such an act of drunken clumsiness much akin to that which just unfolded might be enough to force consideration demanding a prompt and immediate ceasing of my nightly ritual ...

Table 5: Score-maximizing test-set comments for each class (continued)

<b>Predicted Subreddit (Actual Subreddit)</b>	<b>Score-maximizing comment</b>
The_Donald (todayilearned)	still better than if okie doke
sandersforpresident (The_Donald)	donate phonebank ...
PoliticalDiscussion (The_Donald)	i echo this and would add that any supporting kaines <unk> as corey booker of nj who immediately popped out a tweet with a resounding endorsement something like i know tim kaine very well and he is an excellent choice upon the vp announcement should be closely scrutinized might want a place at the table
worldnews (PoliticalDiscussion)	turkey violations of greece airspace is part of the aegean dispute and a whole other thing on the other hand russia have slowly been ramping up the provocations bombing ethnic turks closer and closer to the border i mean it was just friday turkey summoned the russian ambassador over this issue of course this is a message from turkey but i am afraid that it is a message that putin have forced them to send ...
movies (todayilearned)	nicolas winding refn also uses this to his advantage the financiers of drive absolutely fucking hated the movie when they saw the near final cut and asked refn to change things around to which he just responded theres no material its honestly a pretty great way to protect your vision as a filmmaker but obviously you have to know what youre doing
anime (explainlikeimfive)	tou and toh are representing two different things the name <unk> the touma part of kamijou touma from to aru majutsu no index is two characters pronounced <unk> tou and <unk> ma respectively the long o sound comes from the first character only so we write it as touma the name <unk> the tohsaka part of tohsaka rin from fatestay night is two characters pronounced...
gaming (movies)	dae shitty hahaha circlejerk
Fitness (AskMen)	a linear progression based ppl program for pull deadlifts <unk> rows 4x5 1x5 alternate so if you did deadlifts on monday you would do rows on thursday and so on 3x812 pulldowns or pullups or chinups 3x812 seated cable rows or chest supported rows <unk> face pulls 4x812 hammer curls 4x812 dumbbell curls push 4x5 1x5 bench <unk> 1x5 overhead press alternate in the same fashion as the rows and deadlifts
personalfinance (SandersForPresident)	apply on the state exchange it will let you know what youre eligible for basically you and your spouse are going to need to make less than 184230 per month to qualify for medicaid after that you can qualify for aptc to help pay for a full insurance plan from the exchange if you qualify for medicaid and report an increase in income that makes you ineligible you will have to enroll in insurance ...
technology (explainlikeimfive)	each docsis qam256 channel takes up 6mhz <unk> is available for 158 channels total each channel has about <unk> of downstream throughput thats about <unk> total capacity without any tv channels per cmnts one cmnts typically serves up to 200 houses a single 10gb fiber ...
relationships (WTF)	dude she sounds like a super entitled child good on you for breaking up with her you dont need to deal with that sort of drama especially when youre trying to better yourself by going to school and working to support yourself ditch the immature beezy and focus on yourself ...
WTF (todayilearned)	at first i thought it was because the gator slapped him as he went past but right at the start of the gif you can see he is in position with his mouth open just waiting for that other one to come past this was no chance incident this gator premeditated that shit he even had a bro to the right ready to step in just in case i bet 3arm fucked godzillas mate after eating too many fermented fruits
AskMen (AskWomen)	more askmen than this one but kind of interesting questions that would make for a great discussion in general but just to men or just to women stuff that i feel like would be more informative if it had been asked on askreddit especially noticeable when its <unk> questions aimed at men and weird personal advice ones romance stuff usually in <unk>
AskWomen (relationships)	microgestin made me an anxious <unk> over the course of the 15 years that i took it it crept up too those effects were not immediate aubra was okay i still had mood swings quick to anger etc not okay ortho tri cyclen made me suicidal feel starving all the time and i could sleep 16 hours a day and still feel exhausted i failed a semester of college taking otc never again i got an iud and will never go back
Showertoughts (changemyview)	op saw a popular shower thought had a counter shower thought of his own and now has come on here to discuss if he should keep his counter shower thought whether or not it happens much is unknown and for the purposes of this post it is also unimportant
explainlikeimfive (movies)	why do so many people on reddit preface simple questions with eli5 all the time not everything is a complicated concept that needs to be simplified in laymans terms which is what the idea behind eli5 originally was now eli5 even in the sub itself has just become synonymous with i have a question on reddit

Table 6: Most confident incorrect predictions for each class

Predicted Subreddit (Actual Subreddit)	Score-maximizing comment
todayilearned (The_Donald)	from gtthe electrically driven wing flap jackscrew actuator was not recovered the flap switch in the cockpit was destroyed the throttle and propeller controls were found in the fullforward position the mixture control was broken the alternate air control was found in the closed position the key in the magneto switch was found in the both position gtthe tachometer needle was found intact fixed in place and pointed to 2750 rpm the red line on the tachometer began at 2700 rpm the hour
changemyview (relationships)	this right here my local workforce paid for my class books amp testing fees it just took some paperwork verifying income amp a couple of hours for an orientation class amp i headed off to class also ive worked as an aide amp it was by far the most physically gruelling job ive ever had my best friends are my back brace amp icy hot i have some friends who gotten jobs as sitters sometimes in <unk> others in private homes not near as harsh on the body best of luck to you im hoping to see an awesome update from you
atheism (worldnews)	i know im 17 and the same thing happened to me last year i was locked in a small office with the two very large and intimidating male dean and principal and harassed by them while i cried and tried to defend myself i contacted the ffrf and they informed my school officials that they would open a case against them for violating students constitutional rights if they did not stop never let yourself feel small and never let someone take away your inalienable rights always stand up for yourself fight ignorance with knowledge know your rights
WritingPrompts (AskWomen)	the morning that rachel comes to his apartment she even says i want to get back together she knew that they had broken up she just couldnt stand the fact that ross had gone out and found another woman for a one night stand she expected the breakup to wreck him emotionally which would make him come crawling back and apologize she used the break to be emotionally manipulative and then went even further by blaming him when he went out and rebounded she refused to recognize her own role in what had happened and held onto the grudge for years though so did he tldr they were on a break

Table 7: Most confident incorrect predictions for each class (continued)

## 6.2 Figures

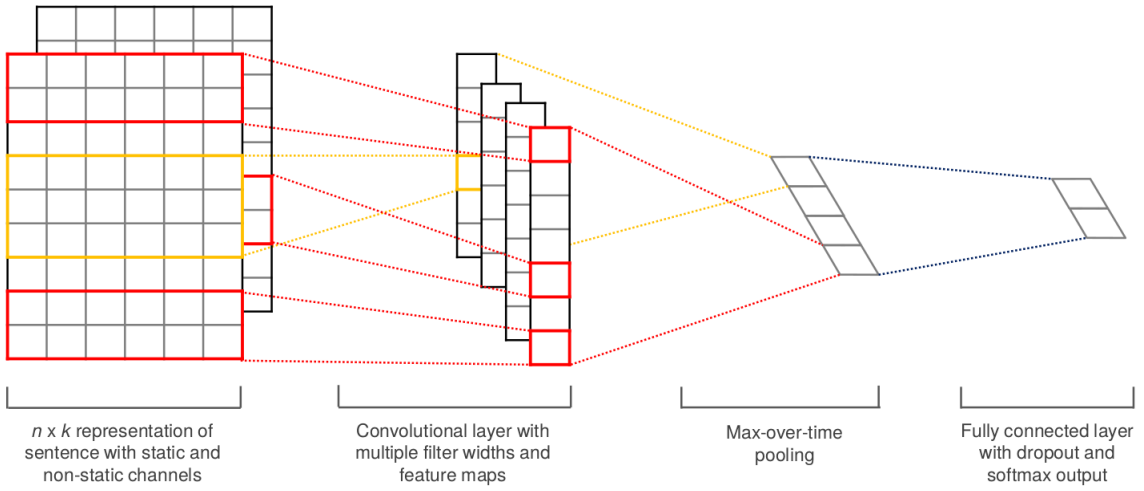


Figure 1: Architecture of CNN model

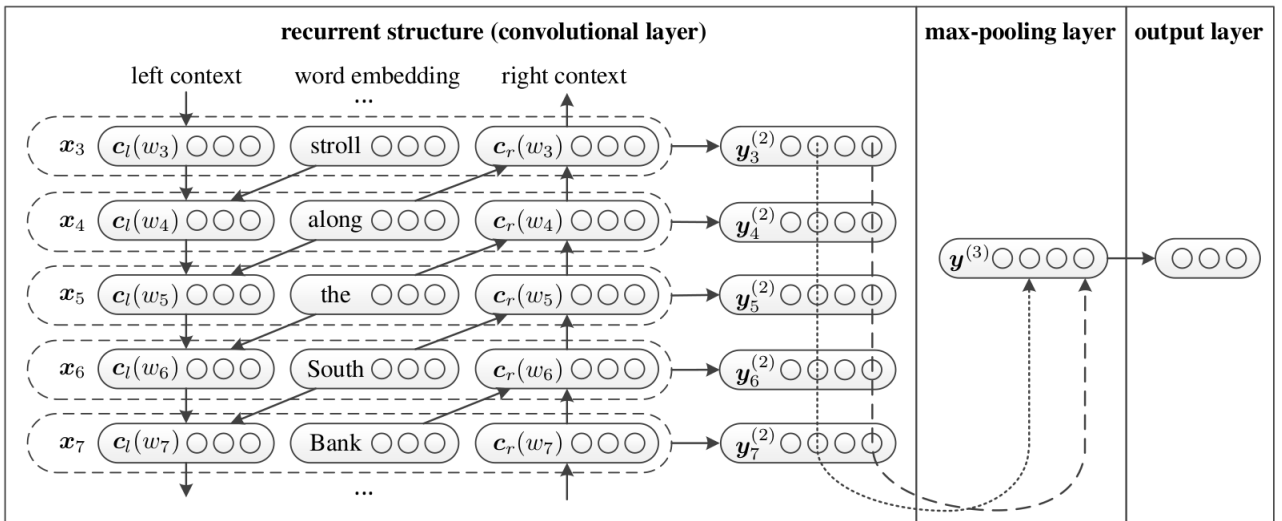


Figure 2: Architecture of RCNN model

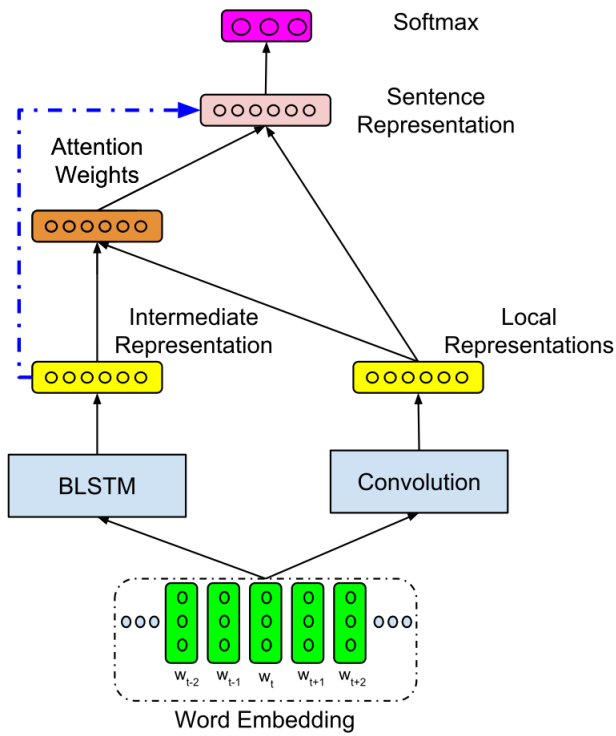


Figure 3: Architecture of ACNN model

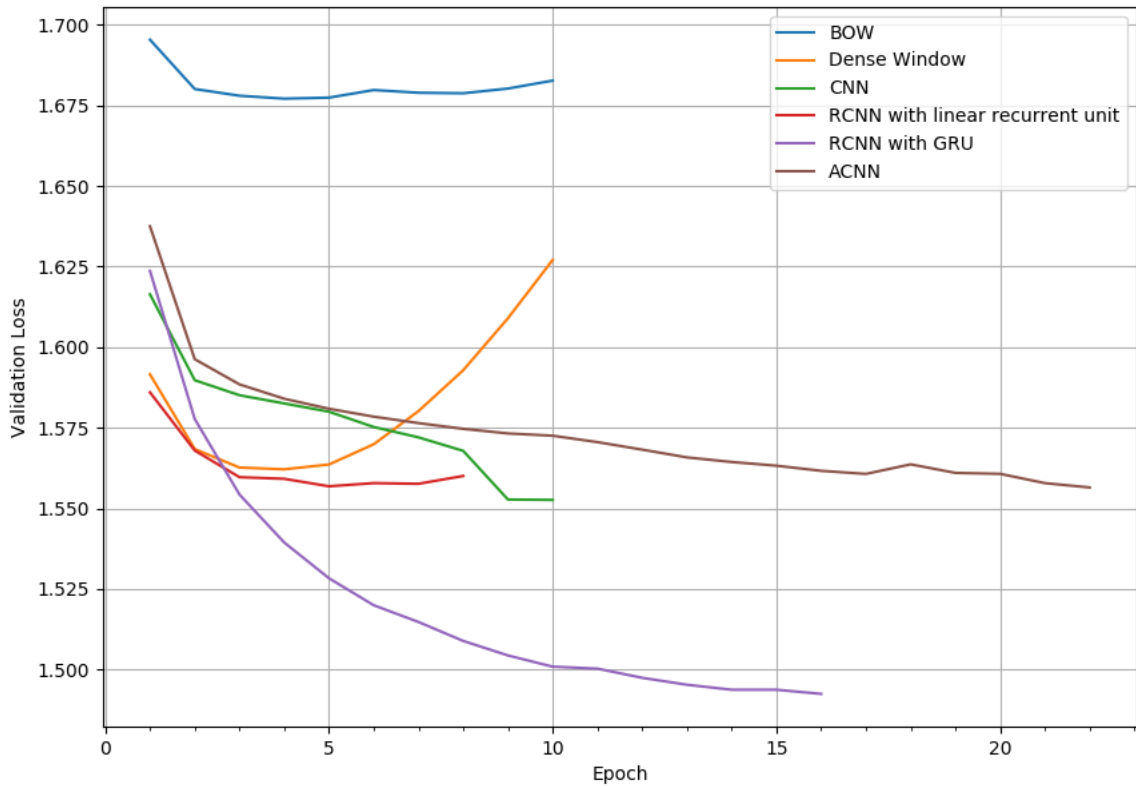


Figure 4: Validation loss after each epoch for various models

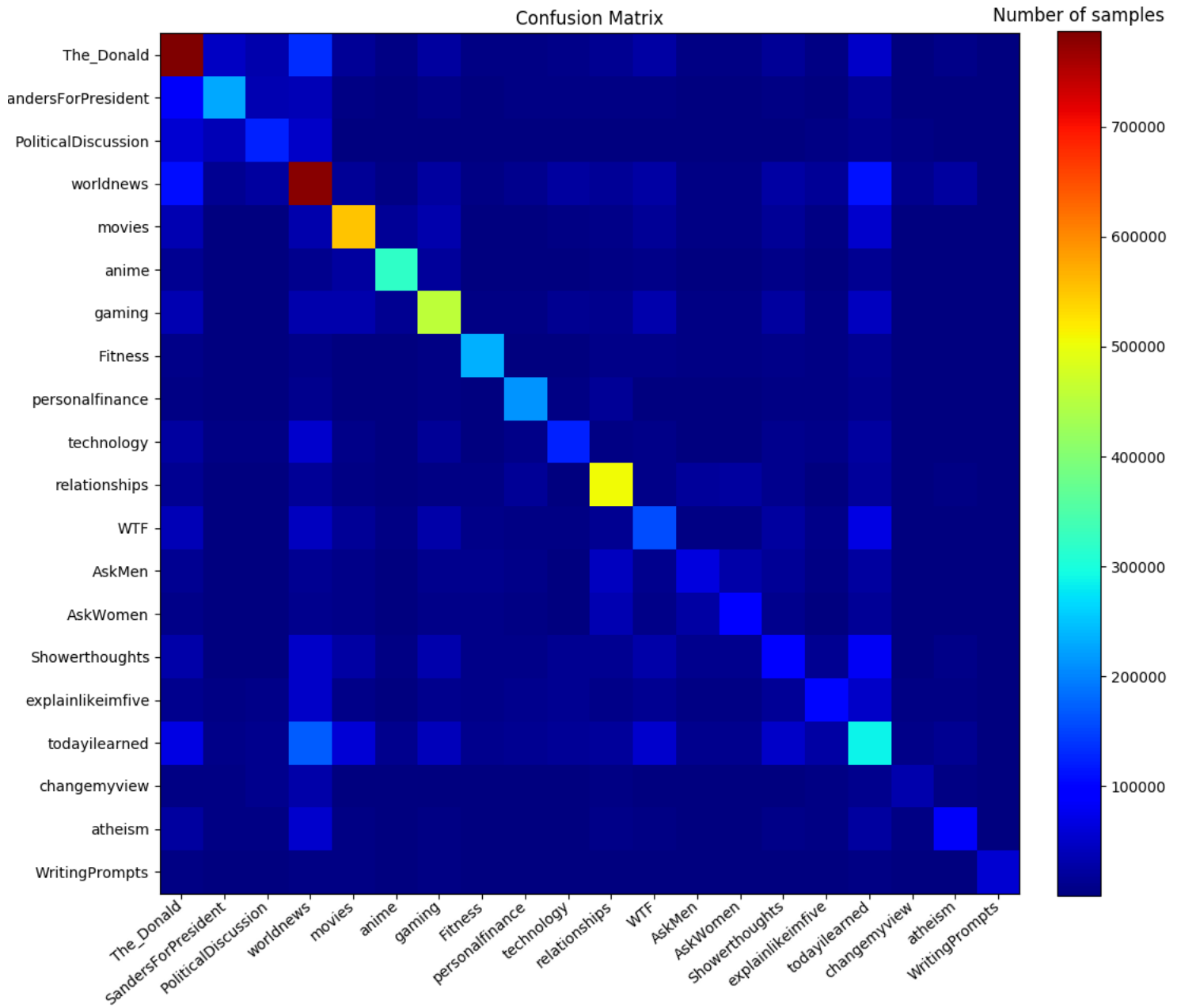


Figure 5: Test Set Confusion Matrix for RCNN-GRU : Number of Samples

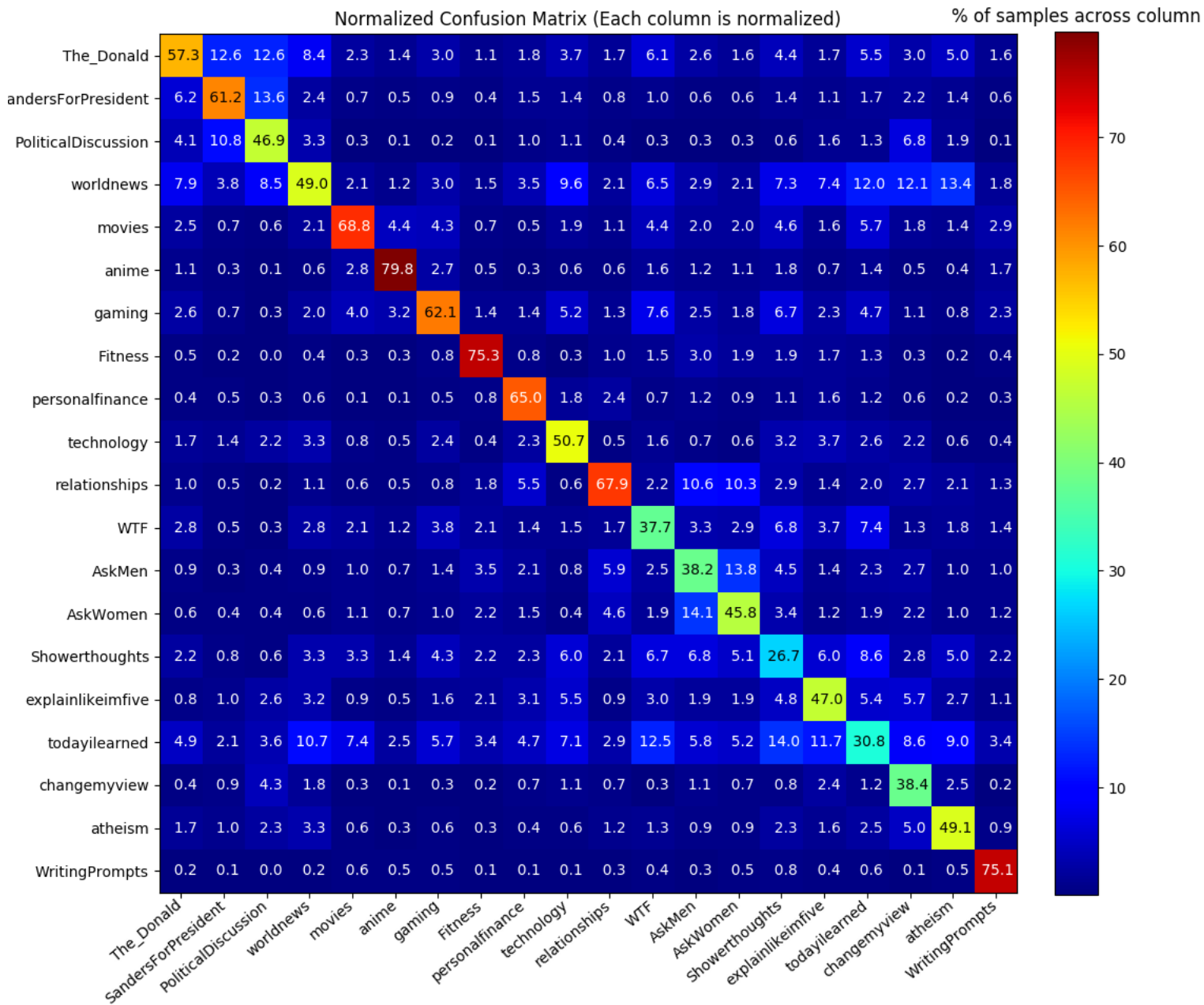


Figure 6: Test Set Confusion Matrix for RCNN-GRU : Column-normalized (Diagonal shows precision)



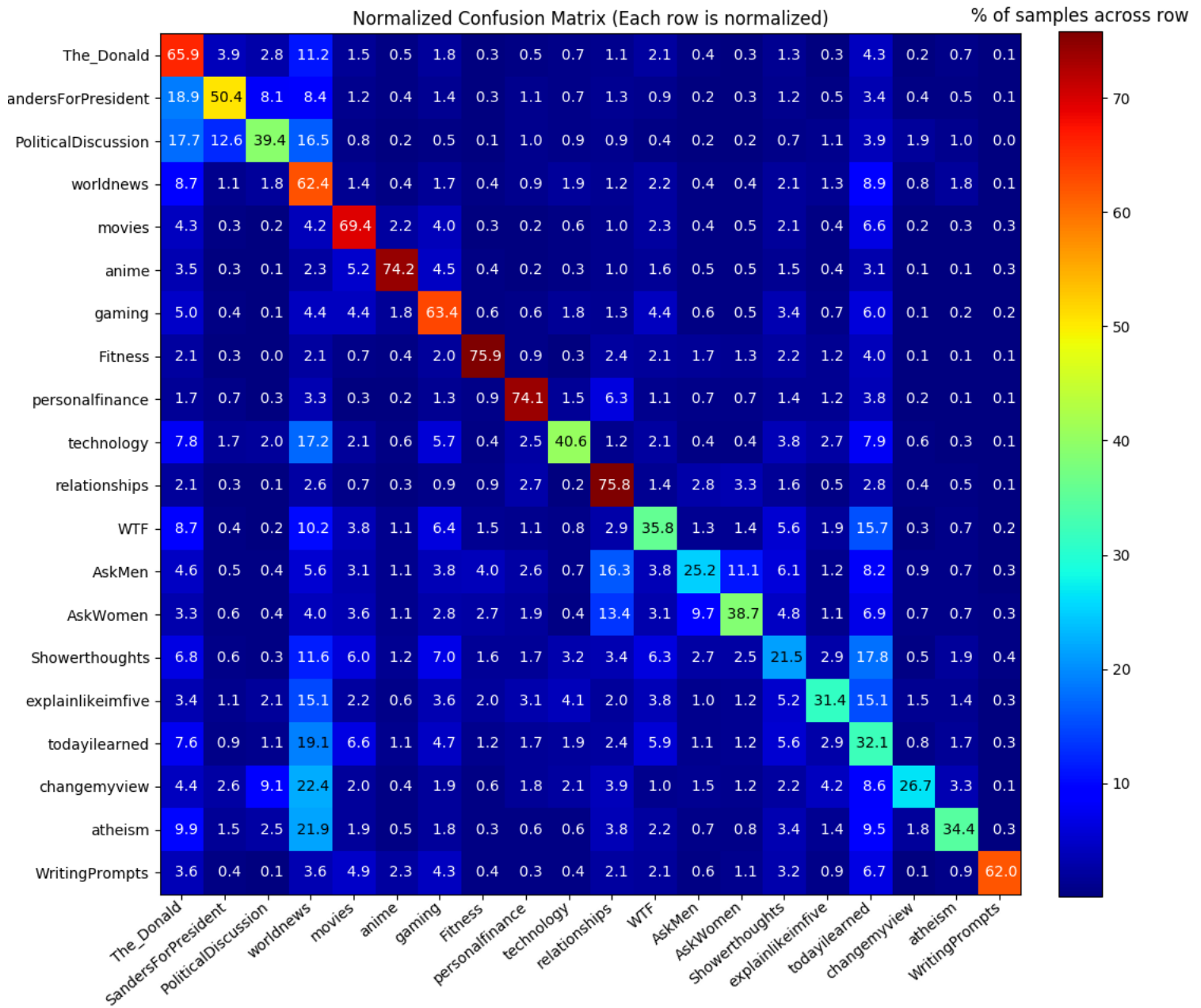


Figure 7: Test set Confusion Matrix for RCNN-GRU - Row-normalized (Diagonal shows recall)