

# UTCNN: a Deep Learning Model of Stance Classification on Social Media Text

**Wei-Fan Chen**

Institute of Information Science,  
Academia Sinica, Taipei, Taiwan.  
viericwf@iis.sinica.edu.tw

**Lun-Wei Ku**

Institute of Information Science,  
Academia Sinica, Taipei, Taiwan.  
lwku@iis.sinica.edu.tw

## Abstract

Most neural network models for document classification on social media focus on text information to the neglect of other information on these platforms. In this paper, we classify post stance on social media channels and develop UTCNN, a neural network model that incorporates user tastes, topic tastes, and user comments on posts. UTCNN not only works on social media texts, but also analyzes texts in forums and message boards. Experiments performed on Chinese Facebook data and English online debate forum data show that UTCNN achieves a 0.755 macro-average f-score for supportive, neutral, and unsupportive stance classes on Facebook data, which is significantly better than models in which either user, topic, or comment information is withheld. This model design greatly mitigates the lack of data for the minor class without the use of oversampling. In addition, UTCNN yields a 0.842 accuracy on English online debate forum data, which also significantly outperforms results from previous work as well as other deep learning models, showing that UTCNN performs well regardless of language or platform.

## 1 Introduction

Deep neural networks have been widely used in text classification and have achieved promising results (Lai et al., 2015; Ren et al., 2016; Huang et al., 2016). Most focus on content information and use models such as convolutional neural networks (CNN) (Kim, 2014) or recursive neural networks (Socher et al., 2013). However, for user-generated posts on social media like Facebook or Twitter, there is more information that should not be ignored. On social media platforms, a user can act either as the author of a post or as a reader who expresses his or her comments about the post.

In this paper, we classify posts taking into account post authorship, likes, topics, and comments. In particular, users and their “likes” hold strong potential for text mining. For example, given a set of posts that are related to a specific topic, a user’s likes and dislikes provide clues for stance labeling. From a user point of view, users with positive attitudes toward the issue leave positive comments on the posts with praise or even just the post’s content; from a post point of view, positive posts attract users who hold positive stances. We also investigate the influence of topics: different topics are associated with different stance labeling tendencies and word usage. For example we discuss women’s rights and unwanted babies on the topic of abortion, but we criticize medicine usage or crime when on the topic of marijuana (Hasan and Ng, 2014). Even for posts on a specific topic like nuclear power, a variety of arguments are raised: green energy, radiation, air pollution, and so on. As for comments, we treat them as additional text information. The arguments in the comments and the commenters (the users who leave the comments) provide hints on the post’s content and further facilitate stance classification.

In this paper, we propose the user-topic-comment neural network (UTCNN), a deep learning model that utilizes user, topic, and comment information. We attempt to learn user and topic representations which encode user interactions and topic influences to further enhance text classification, and we also incorporate comment information. We evaluate this model on a post stance classification task on forum-style social media platforms. The contributions of this paper are as follows: 1. We propose UTCNN,

---

This work is licenced under a Creative Commons Attribution 4.0 International License. License details: <http://creativecommons.org/licenses/by/4.0/>

a neural network for text in modern social media channels as well as legacy social media, forums, and message boards — anywhere that reveals users, their tastes, as well as their replies to posts. 2. When classifying social media post stances, we leverage users, including authors and likers. User embeddings can be generated even for users who have never posted anything. 3. We incorporate a topic model to automatically assign topics to each post in a single topic dataset. 4. We show that overall, the proposed method achieves the highest performance in all instances, and that all of the information extracted, whether users, topics, or comments, still has its contributions.

## 2 Related Work

### 2.1 Extra-Linguistic Features for Stance Classification

In this paper we aim to use text as well as other features to see how they complement each other in a deep learning model. In the stance classification domain, previous work has showed that text features are limited, suggesting that adding extra-linguistic constraints could improve performance (Bansal et al., 2008; Hasan and Ng, 2013a; Walker et al., 2012). For example, Hasan and Ng as well as Thomas et al. require that posts written by the same author have the same stance (Hasan and Ng, 2013b; Thomas et al., 2006). The addition of this constraint yields accuracy improvements of 1–7% for some models and datasets. Hasan and Ng later added user-interaction constraints and ideology constraints (Hasan and Ng, 2013a): the former models the relationship among posts in a sequence of replies and the latter models inter-topic relationships, e.g., users who oppose abortion could be conservative and thus are likely to oppose gay rights.

For work focusing on online forum text, since posts are linked through user replies, sequential labeling methods have been used to model relationships between posts. For example, Hasan and Ng use hidden Markov models (HMMs) to model dependent relationships to the preceding post (Hasan and Ng, 2013b); Burfoot et al. use iterative classification to repeatedly generate new estimates based on the current state of knowledge (Burfoot et al., 2011); Sridhar et al. use probabilistic soft logic (PSL) to model reply links via collaborative filtering (Sridhar et al., 2015). In the Facebook dataset we study, we use comments instead of reply links. However, as the ultimate goal in this paper is predicting not comment stance but post stance, we treat comments as extra information for use in predicting post stance.

### 2.2 Deep Learning on Extra-Linguistic Features

In recent years neural network models have been applied to document sentiment classification (Socher et al., 2012; Socher et al., 2013; Kalchbrenner et al., 2014; Johnson and Zhang, 2015; Huang et al., 2016). Text features can be used in deep networks to capture text semantics or sentiment. For example, Dong et al. use an adaptive layer in a recursive neural network for target-dependent Twitter sentiment analysis, where targets are topics such as *windows 7* or *taylor swift* (Dong et al., 2014a; Dong et al., 2014b); recursive neural tensor networks (RNTNs) utilize sentence parse trees to capture sentence-level sentiment for movie reviews (Socher et al., 2013); Le and Mikolov predict sentiment by using paragraph vectors to model each paragraph as a continuous representation (Le and Mikolov, 2014). They show that performance can thus be improved by more delicate text models.

Others have suggested using extra-linguistic features to improve the deep learning model. The user-word composition vector model (UWCVM) (Tang et al., 2015b) is inspired by the possibility that the strength of sentiment words is user-specific; to capture this they add user embeddings in their model. In UPNN, a later extension, they further add a product-word composition as product embeddings, arguing that products can also show different tendencies of being rated or reviewed (Tang et al., 2015a). Their addition of user information yielded 2–10% improvements in accuracy as compared to the above-mentioned RNTN and paragraph vector methods. We also seek to inject user information into the neural network model. In comparison to the research of Tang et al. on sentiment classification for product reviews, the difference is two-fold. First, we take into account multiple users (one author and potentially many likers) for one post, whereas only one user (the reviewer) is involved in a review. Second, we add comment information to provide more features for post stance classification. None of these two factors have been considered previously in a deep learning model for text stance classification. Therefore, we

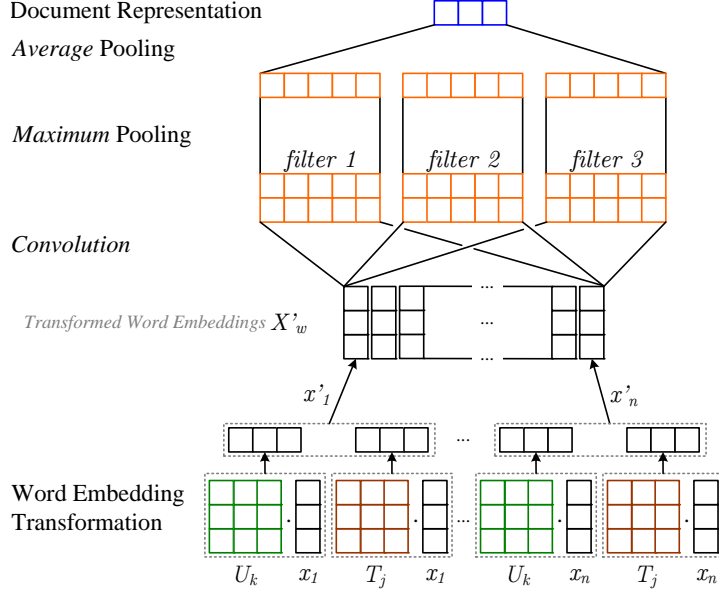


Figure 1: Document composition in a convolutional neural network with three convolutional filters and user- and topic-dependent semantic transformations. Respectively,  $x_w$  is the word embedding of word  $w$ ,  $x'_w$  is the word embedding of word  $w$  after transformation,  $U_k$  and  $T_j$  are user and topic matrix embeddings for user  $k$  and topic  $j$ .

propose UTCNN, which generates and utilizes user embeddings for all users — even for those who have not authored any posts — and incorporates comments to further improve performance.

### 3 Method

In this section, we first describe CNN-based document composition, which captures user- and topic-dependent document-level semantic representation from word representations. Then we show how to add comment information to construct the user-topic-comment neural network (UTCNN).

#### 3.1 User- and Topic-dependent Document Composition

As shown in Figure 1, we use a general CNN (Kim, 2014) and two semantic transformations for document composition<sup>1</sup>. We are given a document with an engaged user  $k$ , a topic  $j$ , and its composite  $n$  words, each word  $w$  of which is associated with a word embedding  $x_w \in \mathbb{R}^d$  where  $d$  is the vector dimension. For each word embedding  $x_w$ , we apply two dot operations as shown in Equation 1:

$$x'_w = [U_k \cdot x_w; T_j \cdot x_w] \quad (1)$$

where  $U_k \in \mathbb{R}^{d_u \times d}$  models the user reading preference for certain semantics, and  $T_j \in \mathbb{R}^{d_t \times d}$  models the topic semantics;  $d_u$  and  $d_t$  are the dimensions of transformed user and topic embeddings respectively. We use  $U_k$  to model semantically what each user prefers to read and/or write, and use  $T_j$  to model the semantics of each topic. The dot operation of  $U_k$  and  $x_w$  transforms the global representation  $x_w$  to a user-dependent representation. Likewise, the dot operation of  $T_j$  and  $x_w$  transforms  $x_w$  to a topic-dependent representation.

After the two dot operations on  $x_w$ , we have user-dependent and topic-dependent word vectors  $U_k \cdot x_w$  and  $T_j \cdot x_w$ , which are concatenated to form a user- and topic-dependent word vector  $x'_w$ . Then the transformed word embeddings  $X'_w = [x'_1; x'_2; \dots; x'_n]$  are used as the CNN input. Here we apply three convolutional layers on the concatenated transformed word embeddings  $x'_c = [x'_m; x'_{m+1}; \dots; x'_{m+l_{cf}-1}] \in \mathbb{R}^{d \cdot l_{cf}}$ :

$$h_{cf} = f(W_{cf} \cdot x'_c + b_{cf}) \quad (2)$$

<sup>1</sup>Here by saying *document*, we mean the user-generated content in a post or a comment.

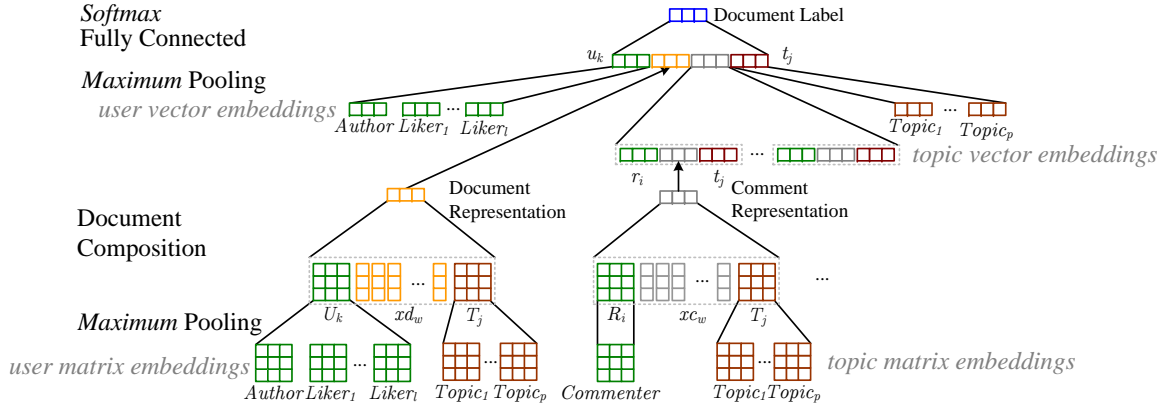


Figure 2: The UTCNN model. Assuming one post author,  $l$  likers and  $p$  topics,  $x_{d_w}$  is the word embedding of word  $w$  in the document;  $x_{c_w}$  is the word embedding of word  $w$  in the comments;  $U_k$  and  $u_k$  are the moderator matrix and vector embedding for moderator  $k$ ;  $T_j$  and  $t_j$  are the topic matrix and vector embedding for topic  $j$ ;  $R_i$  and  $r_i$  are the commenter matrix and vector embedding for commenter  $i$ . For simplicity we do not explicitly plot the topic vector embedding part for comments, but it does include a maximum pooling layer as with documents.

where  $m$  is the index of words;  $f$  is a non-linear activation function (we use  $\tanh^2$ );  $W_{cf} \in \mathbb{R}^{len \times d \cdot l_{cf}}$  is the convolutional filter with input length  $d \cdot l_{cf}$  and output length  $len$ , where  $l_{cf}$  is the window size of the convolutional operation; and  $h_{cf}$  and  $b_{cf}$  are the output and bias of the convolution layer  $cf$ , respectively. In our experiments, the three window sizes  $l_{cf}$  in the three convolution layers are one, two, and three, encoding unigram, bigram, and trigram semantics accordingly.

After the convolutional layer, we add a maximum pooling layer among convolutional outputs to obtain the unigram, bigram, and trigram  $n$ -gram representations. This is succeeded by an average pooling layer for an element-wise average of the three maximized convolution outputs.

### 3.2 UTCNN Model Description

Figure 2 illustrates the UTCNN model. As more than one user may interact with a given post, we first add a maximum pooling layer after the user matrix embedding layer and user vector embedding layer to form a moderator matrix embedding  $U_k$  and a moderator vector embedding  $u_k$  for moderator  $k$  respectively, where  $U_k$  is used for the semantic transformation in the document composition process, as mentioned in the previous section. The term *moderator* here is to denote the pseudo user who provides the overall semantic/sentiment of all the engaged users for one document. The embedding  $u_k$  models the moderator stance preference, that is, the pattern of the revealed user stance: whether a user is willing to show his preference, whether a user likes to show impartiality with neutral statements and reasonable arguments, or just wants to show strong support for one stance. Ideally, the latent user stance is modeled by  $u_k$  for each user. Likewise, for topic information, a maximum pooling layer is added after the topic matrix embedding layer and topic vector embedding layer to form a joint topic matrix embedding  $T_j$  and a joint topic vector embedding  $t_j$  for topic  $j$  respectively, where  $T_j$  models the semantic transformation of topic  $j$  as in users and  $t_j$  models the topic stance tendency. The latent topic stance is also modeled by  $t_j$  for each topic.

As for comments, we view them as short documents with authors only but without likers nor their own comments<sup>3</sup>. Therefore we apply document composition on comments although here users are commenters (users who comment). It is noticed that the word embeddings  $x_w$  for the same word in the posts and comments are the same, but after being transformed to  $x'_w$  in the document composition process

<sup>2</sup>Some papers suggest using *ReLU* as the activation function in deep CNNs with many layers. Nevertheless, we use *tanh* as the activation function, as our model is moderately deep and empirically we found the impact to be limited.

<sup>3</sup>Recently Facebook released a function allowing likes and comments on comments, but it was not available during the time we collected data. However, UTCNN works on this richer data, as comments are treated as posts under this framework.

Dataset	FBFans				CreateDebate							
Type	<i>Sup</i>	<i>Neu</i>	<i>Uns</i>	All	ABO		GAY		OBA		MAR	
					<i>F</i>	<i>A</i>	<i>F</i>	<i>A</i>	<i>F</i>	<i>A</i>	<i>F</i>	<i>A</i>
Training	7,097	19,412	245	26,754	770.4	622.4	700.8	400.0	420.8	367.2	355.2	145.6
Development	155	2,785	11	2,951	-	-	-	-	-	-	-	-
Testing	252	2,619	19	2,890	192.6	155.6	175.2	100.0	105.2	91.8	88.8	36.4
All	7,504	24,816	275	32,595	963.0	778.0	876.0	500.0	526.0	459.0	444.0	182.0

Table 1: Annotation results of FBFans and CreateDebate dataset.

Author \ Post	Post		
	<i>Sup</i>	<i>Neu</i>	<i>Uns</i>
<i>Sup</i>	<b>58.5%</b>	<b>51.3%</b>	29.4%
<i>Neu</i>	33.9%	43.5%	9.3%
<i>Uns</i>	7.6%	5.2%	<b>61.3%</b>

Table 2: Distribution of like behavior.

shown in Figure 1, they might become different because of their different engaged users. The output comment representation together with the commenter vector embedding  $r_i$  and topic vector embedding  $t_j$  are concatenated and a maximum pooling layer is added to select the most important feature for comments. Instead of requiring that the comment stance agree with the post, UTCNN simply extracts the most important features of the comment contents; they could be helpful, whether they show obvious agreement or disagreement. Therefore when combining comment information here, the maximum pooling layer is more appropriate than other pooling or merging layers. Indeed, we believe this is one reason for UTCNN’s performance gains.

Finally, the pooled comment representation, together with user vector embedding  $u_k$ , topic vector embedding  $t_j$ , and document representation are fed to a fully connected network, and softmax is applied to yield the final stance label prediction for the post.

## 4 Experiment

We start with the experimental dataset and then describe the training process as well as the implementation of the baselines. We also implement several variations to reveal the effects of features: authors, likers, comment, and commenters. In the results section we compare our model with related work.

### 4.1 Dataset

We tested the proposed UTCNN on two different datasets: FBFans and CreateDebate. FBFans is a privately-owned<sup>4</sup>, single-topic, Chinese, unbalanced, social media dataset, and CreateDebate is a public, multiple-topic, English, balanced, forum dataset. Results using these two datasets show the applicability and superiority for different topics, languages, data distributions, and platforms.

The FBFans dataset contains data from anti-nuclear-power Chinese Facebook fan groups from September 2013 to August 2014, including posts and their author and liker IDs. There are a total of 2,496 authors, 505,137 likers, 33,686 commenters, and 505,412 unique users. Two annotators were asked to take into account only the post content to label the stance of the posts in the whole dataset as *supportive*, *neutral*, or *unsupportive* (hereafter denoted as *Sup*, *Neu*, and *Uns*). *Sup/Uns* posts were those in support of or against anti-reconstruction; *Neu* posts were those evincing a neutral standpoint on the topic, or were irrelevant. Raw agreement between annotators is 0.91, indicating high agreement. Specifically, Cohen’s Kappa for *Neu* and not *Neu* labeling is 0.58 (moderate), and for *Sup* or *Uns* labeling is

<sup>4</sup>Currently not released due to copyright and privacy issues.

0.84 (almost perfect). Posts with inconsistent labels were filtered out, and the development and testing sets were randomly selected from what was left. Posts in the development and testing sets involved at least one user who appeared in the training set. The number of posts for each stance is shown on the left-hand side of Table 1. About twenty percent of the posts were labeled with a stance, and the number of supportive (*Sup*) posts was much larger than that of the unsupportive (*Uns*) ones: this is thus highly skewed data, which complicates stance classification. On average, 161.1 users were involved in one post. The maximum was 23,297 and the minimum was one (the author). For comments, on average there were 3 comments per post. The maximum was 1,092 and the minimum was zero.

To test whether the assumption of this paper – posts attract users who hold the same stance to like them – is reliable, we examine the likes from authors of different stances. Posts in FBFans dataset are used for this analysis. We calculate the like statistics of each distinct author from these 32,595 posts. As the numbers of authors in the *Sup*, *Neu* and *Uns* stances are largely imbalanced, these numbers are normalized by the number of users of each stance. Table 4 shows the results. Posts with stances (i.e., not neutral) attract users of the same stance. Neutral posts also attract both supportive and neutral users, like what we observe in supportive posts, but just the neutral posts can attract even more neutral likers. These results do suggest that users prefer posts of the same stance, or at least posts of no obvious stance which might cause annoyance when reading, and hence support the user modeling in our approach.

The CreateDebate dataset was collected from an English online debate forum<sup>5</sup> discussing four topics: abortion (ABO), gay rights (GAY), Obama (OBA), and marijuana (MAR). The posts are annotated as *for* (F) and *against* (A). Replies to posts in this dataset are also labeled with stance and hence use the same data format as posts. The labeling results are shown in the right-hand side of Table 1. We observe that the dataset is more balanced than the FBFans dataset. In addition, there are 977 unique users in the dataset. To compare with Hasan and Ng’s work, we conducted five-fold cross-validation and present the annotation results as the average number of all folds (Hasan and Ng, 2013b; Hasan and Ng, 2014).

The FBFans dataset has more integrated functions than the CreateDebate dataset; thus our model can utilize all linguistic and extra-linguistic features. For the CreateDebate dataset, on the other hand, the like and comment features are not available (as there is a stance label for each reply, replies are evaluated as posts as other previous work) but we still implemented our model using the content, author, and topic information.

## 4.2 Settings

In the UTCNN training process, cross-entropy was used as the loss function and *AdaGrad* as the optimizer. For FBFans dataset, we learned the 50-dimensional word embeddings on the whole dataset using GloVe<sup>6</sup> (Pennington et al., 2014) to capture the word semantics; for CreateDebate dataset we used the publicly available English 50-dimensional word embeddings, pre-trained also using GloVe. These word embeddings were fixed in the training process. The learning rate was set to 0.03. All user and topic embeddings were randomly initialized in the range of [-0.1 0.1]. Matrix embeddings for users and topics were sized at 250 ( $5 \times 50$ ); vector embeddings for users and topics were set to length 10.

We applied the LDA topic model (Blei et al., 2003) on the FBFans dataset to determine the latent topics with which to build topic embeddings, as there is only one general known topic: nuclear power plants. We learned 100 latent topics and assigned the top three topics for each post. For the CreateDebate dataset, which itself constitutes four topics, the topic labels for posts were used directly without additionally applying LDA.

For the FBFans data we report class-based f-scores as well as the macro-average f-score ( $F_1^{SNU}$ ) shown in equation 3.

$$F_1^{SNU} = 2 \cdot \frac{P^{SNU} \cdot R^{SNU}}{P^{SNU} + R^{SNU}} \quad (3)$$

<sup>5</sup><http://www.createdebate.com/>

<sup>6</sup><http://nlp.stanford.edu/projects/glove/>

Method	Features				F-score			$F_1^{SNU}$
	Content	User	Topic	Comment	Sup	Neu	Uns	
Majority					.000	.841	.000	.280
SVM -UniBiTrigram	✓				.721	.967	.091	.640
SVM -UniBiTrigram	✓			✓	.610	.938	.156	.621
SVM -AvgWordVec	✓				.631	.952	.114	.579
SVM -AvgWordVec	✓			✓	.526	.100	.165	.336
SVM -AvgWordVec (transformed)	✓	✓	✓		.571	.920	.229	.637
SVM -AvgWordVec (transformed)	✓	✓	✓	✓	.597	.963	.210	.642
CNN (Kim, 2014)	✓				.738	.967	.171	.637
CNN (Kim, 2014)	✓			✓	.726	.964	.222	.648
RCNN (Lai et al., 2015)	✓				.669	.951	.079	.606
RCNN (Lai et al., 2015)	✓			✓	.628	.944	.096	.605
UTCNN without user	✓		✓	✓	<b>.748</b>	<b>.973</b>	.000	.580
UTCNN without topic	✓	✓		✓	.643	.944	.476	.706
UTCNN without comment	✓	✓	✓		.632	.940	.480	.707
UTCNN shared user embedding	✓	✓	✓	✓	.625	.969	.531	.732
UTCNN (full)	✓	✓	✓	✓	.698	.957	<b>.571</b>	<b>.755*</b>

Table 3: Performance of post stance classification on the FBFans dataset.

\*UTCNN (full) results are statistically significant ( $p$ -value  $< 0.005$ ) with respect to all other methods except for UTCNN shared user embedding.

where  $P^{SNU}$  and  $R^{SNU}$  are the average precision and recall of the three class. We adopted the macro-average f-score as the evaluation metric for the overall performance because (1) the experimental dataset is severely imbalanced, which is common for contentious issues; and (2) for stance classification, content in minor-class posts is usually more important for further applications. For the CreateDebate dataset, accuracy was adopted as the evaluation metric to compare the results with related work (Hasan and Ng, 2013a; Hasan and Ng, 2013b; Sridhar et al., 2015).

### 4.3 Baselines

We pit our model against the following baselines: 1) SVM with unigram, bigram, and trigram features, which is a standard yet rather strong classifier for text features; 2) SVM with average word embedding, where a document is represented as a continuous representation by averaging the embeddings of the composite words; 3) SVM with average transformed word embeddings (the  $x'w$  in equation 1), where a document is represented as a continuous representation by averaging the transformed embeddings of the composite words; 4) two mature deep learning models on text classification, CNN (Kim, 2014) and Recurrent Convolutional Neural Networks (RCNN) (Lai et al., 2015), where the hyperparameters are based on their work; 5) the above SVM and deep learning models with comment information; 6) UTCNN without user information, representing a pure-text CNN model where we use the same user matrix and user embeddings  $U_k$  and  $u_k$  for each user; 7) UTCNN without the LDA model, representing how UTCNN works with a single-topic dataset; 8) UTCNN without comments, in which the model predicts the stance label given only user and topic information. All these models were trained on the training set, and parameters as well as the SVM kernel selections (linear or RBF) were fine-tuned on the development set. Also, we adopt oversampling on SVMs, CNN and RCNN because the FBFans dataset is highly imbalanced.

### 4.4 Results on FBFans Dataset

In Table 3 we show the results of UTCNN and the baselines on the FBFans dataset. Here Majority yields good performance on *Neu* since FBFans is highly biased to the neutral class. The SVM models

perform well on *Sup* and *Neu* but perform poorly for *Uns*, showing that content information in itself is insufficient to predict stance labels, especially for the minor class. With the transformed word embedding feature, SVM can achieve comparable performance as SVM with  $n$ -gram feature. However, the much fewer feature dimension of the transformed word embedding makes SVM with word embeddings a more efficient choice for modeling the large scale social media dataset. For the CNN and RCNN models, they perform slightly better than most of the SVM models but still, the content information is insufficient to achieve a good performance on the *Uns* posts. As to adding comment information to these models, since the commenters do not always hold the same stance as the author, simply adding comments and post contents together merely adds noise to the model.

Among all UTCNN variations, we find that user information is most important, followed by topic and comment information. UTCNN without user information shows results similar to SVMs — it does well for *Sup* and *Neu* but detects no *Uns*. Its best f-scores on both *Sup* and *Neu* among all methods show that with enough training data, content-based models can perform well; at the same time, the lack of user information results in too few clues for minor-class posts to either predict their stance directly or link them to other users and posts for improved performance. The 17.5% improvement when adding user information suggests that user information is especially useful when the dataset is highly imbalanced. All models that consider user information predict the minority class successfully. UTCNN without topic information works well but achieves lower performance than the full UTCNN model. The 4.9% performance gain brought by LDA shows that although it is satisfactory for single topic datasets, adding that latent topics still benefits performance: even when we are discussing the same topic, we use different arguments and supporting evidence. Lastly, we get 4.8% improvement when adding comment information and it achieves comparable performance to UTCNN without topic information, which shows that comments also benefit performance. For platforms where user IDs are pixelated or otherwise hidden, adding comments to a text model still improves performance. In its integration of user, content, and comment information, the full UTCNN produces the highest f-scores on all *Sup*, *Neu*, and *Uns* stances among models that predict the *Uns* class, and the highest macro-average f-score overall. This shows its ability to balance a biased dataset and supports our claim that UTCNN successfully bridges content and user, topic, and comment information for stance classification on social media text. Another merit of UTCNN is that it does not require a balanced training data. This is supported by its outperforming other models though no oversampling technique is applied to the UTCNN related experiments as shown in this paper. Thus we can conclude that the user information provides strong clues and it is still rich even in the minority class.

We also investigate the semantic difference when a user acts as an author/liker or a commenter. We evaluated a variation in which all embeddings from the same user were forced to be identical (this is the UTCNN shared user embedding setting in Table 3). This setting yielded only a 2.5% improvement over the model without comments, which is not statistically significant. However, when separating authors/likers and commenters embeddings (i.e., the UTCNN full model), we achieved much greater improvements (4.8%). We attribute this result to the tendency of users to use different wording for different roles (for instance author vs commenter). This is observed when the user, acting as an author, attempts to support her argument against nuclear power by using improvements in solar power; when acting as a commenter, though, she interacts with post contents by criticizing past politicians who supported nuclear power or by arguing that the proposed evacuation plan in case of a nuclear accident is ridiculous. Based on this finding, in the final UTCNN setting we train two user matrix embeddings for one user: one for the author/liker role and the other for the commenter role.

#### 4.5 Results on CreateDebate Dataset

Table 4 shows the results of UTCNN, baselines as we implemented on the FBFans dataset and related work on the CreateDebate dataset. We do not adopt oversampling on these models because the CreateDebate dataset is almost balanced. In previous work, integer linear programming (ILP) or linear-chain conditional random fields (CRFs) were proposed to integrate text features, author, ideology, and user-interaction constraints, where text features are unigram, bigram, and POS-dependencies; the author



Method	Features		Topics				AVG
	Text	User	ABO	GAY	OBA	MAR	
Majority			.549	.634	.539	.695	.604
SVM -UniBiTrigram	✓		.592	.569	.565	.673	.600
SVM -AvgWordVec	✓		.559	.637	.548	.708	.613
SVM -AvgWordVec (transformed)	✓	✓	.859	.830	.800	.741	.808
CNN (Kim, 2014)	✓		.553	.636	.557	.709	.614
RCNN (Lai et al., 2015)	✓		.553	.637	.534	.709	.608
ILP (Hasan and Ng, 2013a)	✓		.614	.626	.581	.669	.623
ILP (Hasan and Ng, 2013a)	✓	✓	.749	.709	.727	.754	.735
CRF (Hasan and Ng, 2013b)	✓	✓	.747	.699	.711	.754	.728
PSL (Sridhar et al., 2015)	✓	✓	.668	.727	.635	.690	.680
UTCNN without topic	✓	✓	.824	<b>.851</b>	.743	<b>.814</b>	.808
UTCNN without user	✓		.617	.627	.599	.685	.632
UTCNN (full)	✓	✓	<b>.878</b>	.850	<b>.857</b>	.782	<b>.842*</b>

Table 4: Accuracies of post stance classification on CreateDebate dataset.

\*UTCNN results were statistically significant ( $p$ -value  $< 0.001$ ) with respect to other UTCNN settings.

constraint tends to require that posts from the same author for the same topic hold the same stance; the ideology constraint aims to capture inferences between topics for the same author; the user-interaction constraint models relationships among posts via user interactions such as replies (Hasan and Ng, 2013a; Hasan and Ng, 2013b).

The SVM with  $n$ -gram or average word embedding feature performs just similar to the majority. However, with the transformed word embedding, it achieves superior results. It shows that the learned user and topic embeddings really capture the user and topic semantics. This finding is not so obvious in the FBFans dataset and it might be due to the unfavorable data skewness for SVM. As for CNN and RCNN, they perform slightly better than most SVMs as we found in Table 3 for FBFans.

Compared to the ILP (Hasan and Ng, 2013a) and CRF (Hasan and Ng, 2013b) methods, the UTCNN user embeddings encode author and user-interaction constraints, where the ideology constraint is modeled by the topic embeddings and text features are modeled by the CNN. The significant improvement achieved by UTCNN suggests the latent representations are more effective than overt model constraints.

The PSL model (Sridhar et al., 2015) jointly labels both author and post stance using probabilistic soft logic (PSL) (Bach et al., 2015) by considering text features and reply links between authors and posts as in Hasan and Ng’s work. Table 4 reports the result of their best AD setting, which represents the full joint stance/disagreement collective model on posts and is hence more relevant to UTCNN. In contrast to their model, the UTCNN user embeddings represent relationships between authors, but UTCNN models do not utilize link information between posts. Though the PSL model has the advantage of being able to jointly label the stances of authors and posts, its performance on posts is lower than the that for the ILP or CRF models. UTCNN significantly outperforms these models on posts and has the potential to predict user stances through the generated user embeddings.

For the CreateDebate dataset, we also evaluated performance when not using topic embeddings or user embeddings; as replies in this dataset are viewed as posts, the setting without comment embeddings is not available. Table 4 shows the same findings as Table 3: the 21% improvement in accuracy demonstrates that user information is the most vital. This finding also supports the results in the related work: user constraints are useful and can yield 11.2% improvement in accuracy (Hasan and Ng, 2013a). Further considering topic information yields 3.4% improvement, suggesting that knowing the subject of debates provides useful information. In sum, Table 3 together with Table 4 show that UTCNN achieves promising performance regardless of topic, language, data distribution, and platform.

## 5 Conclusion

We have proposed UTCNN, a neural network model that incorporates user, topic, content and comment information for stance classification on social media texts. UTCNN learns user embeddings for all users with minimum active degree, i.e., one post or one like. Topic information obtained from the topic model or the pre-defined labels further improves the UTCNN model. In addition, comment information provides additional clues for stance classification. We have shown that UTCNN achieves promising and balanced results. In the future we plan to explore the effectiveness of the UTCNN user embeddings for author stance classification.

## Acknowledgements

Research of this paper was partially supported by Ministry of Science and Technology, Taiwan, under the contract MOST 104-2221-E-001-024-MY2.

## References

- Stephen H Bach, Matthias Broecheler, Bert Huang, and Lise Getoor. 2015. Hinge-loss markov random fields and probabilistic soft logic. *arXiv preprint arXiv:1505.04406*.
- Mohit Bansal, Claire Cardie, and Lillian Lee. 2008. The power of negative thinking: Exploiting label disagreement in the min-cut classification framework. In *COLING (Posters)*, pages 15–18.
- David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3:993–1022.
- Clinton Burfoot, Steven Bird, and Timothy Baldwin. 2011. Collective classification of congressional floor-debate transcripts. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 1506–1515. Association for Computational Linguistics.
- Li Dong, Furu Wei, Chuanqi Tan, Duyu Tang, Ming Zhou, and Ke Xu. 2014a. Adaptive recursive neural network for target-dependent twitter sentiment classification. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 49–54. Association for Computational Linguistics.
- Li Dong, Furu Wei, Ming Zhou, and Ke Xu. 2014b. Adaptive multi-compositionality for recursive neural models with applications to sentiment analysis. In *Proceedings of the Conference of the Association for the Advancement of Artificial Intelligence*. AAAI.
- Kazi Saidul Hasan and Vincent Ng. 2013a. Extra-linguistic constraints on stance recognition in ideological debates. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 816–821. Association for Computational Linguistics.
- Kazi Saidul Hasan and Vincent Ng. 2013b. Stance classification of ideological debates: Data, models, features, and constraints. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 1348–1356. Association for Computational Linguistics.
- Kazi Saidul Hasan and Vincent Ng. 2014. Why are you taking this stance? identifying and classifying reasons in ideological debates. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 751–762.
- Minlie Huang, Yujie Cao, and Chao Dong. 2016. Modeling rich contexts for sentiment classification with lstm. *arXiv preprint arXiv:1605.01478*.
- Rie Johnson and Tong Zhang. 2015. Effective use of word order for text categorization with convolutional neural networks. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics*.
- Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. 2014. A convolutional neural network for modelling sentences. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*.
- Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1746–1751. Association for Computational Linguistics.

- Siwei Lai, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Recurrent convolutional neural networks for text classification. In *Proceedings of the Conference of the Association for the Advancement of Artificial Intelligence*, pages 2267–2273. AAAI.
- Quoc V Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *Proceedings of the 31st International Conference on Machine Learning*.
- Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1532–1543. Association for Computational Linguistics.
- Yafeng Ren, Yue Zhang, Meishan Zhang, and Donghong Ji. 2016. Context-sensitive twitter sentiment classification using neural network. In *Proceedings of the Conference of the Association for the Advancement of Artificial Intelligence*. AAAI.
- Richard Socher, Brody Huval, Christopher D Manning, and Andrew Y Ng. 2012. Semantic compositionality through recursive matrix-vector spaces. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1201–1211. Association for Computational Linguistics.
- Richard Socher, Alex Perelygin, Jean Y Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, volume 1631, page 1642. Association for Computational Linguistics.
- Dhanya Sridhar, James Foulds, Bert Huang, Lise Getoor, and Marilyn Walker. 2015. Joint models of disagreement and stance in online debate. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics.
- Duyu Tang, Bing Qin, and Ting Liu. 2015a. Learning semantic representations of users and products for document level sentiment classification. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pages 1014–1023. Association for Computational Linguistics.
- Duyu Tang, Bing Qin, Ting Liu, and Yuekui Yang. 2015b. User modeling with neural network for review rating prediction. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, pages 1340–1346.
- Matt Thomas, Bo Pang, and Lillian Lee. 2006. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 327–335. Association for Computational Linguistics.
- Marilyn A Walker, Pranav Anand, Robert Abbott, and Ricky Grant. 2012. Stance classification using dialogic properties of persuasion. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 592–596. Association for Computational Linguistics.