Modeling Transferable Topics for Cross-Target Stance Detection

Penghui Wei and Wenji Mao
[†]SKL-MCCS, Institute of Automation, Chinese Academy of Sciences, Beijing, China
[‡]University of Chinese Academy of Sciences, Beijing, China
{weipenghui2016,wenji.mao}@ia.ac.cn

ABSTRACT

Targeted stance detection aims to classify the attitude of an opinionated text towards a pre-defined target. Previous methods mainly focus on *in-target* setting that models are trained and tested using data specific to the same target. In practical cases, the target we concern may have few or no labeled data, which restrains us from training a target-specific model. In this paper we study the problem of *cross-target* stance detection, utilizing labeled data of a source target to learn models that can be adapted to a destination target. To this end, we propose an effective method, the core intuition of which is to leverage shared latent topics between two targets as transferable knowledge to facilitate model adaptation. Our method acquires topic knowledge with neural variational inference, and further adopts adversarial training that encourages the model to learn target-invariant representations. Experimental results verify that our proposed method is superior to the state-of-the-art methods.

ACM Reference Format:
Penghui Wei and Wenji Mao. 2019. Modeling Transferable Topics for Cross-Target Stance Detection. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '19), July 21–25, 2019, Paris, France.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3331184.3331367

1 INTRODUCTION

Identifying people's stances and attitudes from textual contents on social media has attracted increasing research attention in recent years [6, 8]. Targeted stance detection is the task of automatically determining the attitude (i.e., in favor of, against, or none) of a text towards a given target[1]. The target may be a proposition, a government policy, a product, a person, etc [10]. Understanding stances in text is critical to many scenarios such as government decision-making, fact checking [11, 16] and product services.

In general, targeted stance detection is different from aspect sentiment classification (ASC) in two ways. First, the same stance can be expressed using positive, negative, or neutral sentiment polarity [10]. Second, a text may not directly mention the given target or explicitly express an attitude towards the target, while aspect terms and categories are assumed to be explicitly discussed in ASC. Therefore, targeted stance detection is a challenging task.

[1]In different research fields, the definitions of *stance detection* have some nuances. In this paper, we focus on *targeted* stance detection in opinion mining field [13]. In fact checking field [4], it is defined as classifying the perspective of a text towards a *claim.*

SIGIR '19, July 21–25, 2019, Paris, France

ACM ISBN 978-1-4503-6172-9/19/07…$15.00
https://doi.org/10.1145/3331184.3331367

Table 1: An example of cross-target stance detection.

Source Target: Feminist Movement **Text**: All humans, male and female, should have equal political, economic and social rights. **Stance**: In favor	**Latent topic** equality
Destination Target: Legalization of Abortion **Text**: It's so brilliant that #lovewins - now extend the equality to women's rights **Stance**: In favor	**Latent topic** equality

Previous methods for targeted stance detection mainly focus on *in-target* setting, that is, training and testing models using data specific to the same target. When we are faced with a new target, it is often the case that we have labeled data concerning an existing target but few or no labeled data for the new one. Thus, the lack of labeled training data for new targets restrains us from building target-specific models. This issue has motivated research on *cross-target* stance detection [1, 18]: building models for a *destination target* by utilizing labeled data from a different but related *source target*, so as to alleviate required annotation of new targets through acquiring knowledge from labeled data of existing targets.

Two previous studies have addressed cross-target stance detection. Augenstein et al. [1] integrate the semantic representation of target for learning target-dependent text representation, and Xu et al. [18] further utilize self-attention to notice important words. However, they do not take any information from destination target into account during training, and consequently their models learn more source target-dependent features. Moreover, they only utilize target information for text representation learning, but do not explicitly model transferable knowledge between two targets, thus the cross-target adaptation ability of them is rather limited.

The key research challenge in cross-target stance detection task is the effective modeling of transferable knowledge to facilitate adaptation across two targets. Intuitively, users often discuss some *subordinated topics* of a target, and expressions about these topics can be used to infer their stances towards the target. Further, the commonly *shared topics* between two targets can be leveraged as transferable knowledge to help model adaptation across targets. Table 1 shows an illustrative example. Although two texts belong to different targets, they both express attitudes towards the topic "equality" to hold their stances. It is our intention to incorporate such topic knowledge into stance detection models for improving their cross-target adaptation ability.

To this end, we propose a cross-target stance detection method that integrates topic knowledge acquisition and target-invariant text representation learning into a unified end-to-end framework. To effectively acquire transferable topic knowledge and enable this process to be trained via backpropagation, we employ neural variational inference [9, 14] to yield latent topics using unlabeled data from both source and destination targets, and utilize the acquired topic knowledge to enhance text representations. To further encourage our model to learn more target-invariant representations,

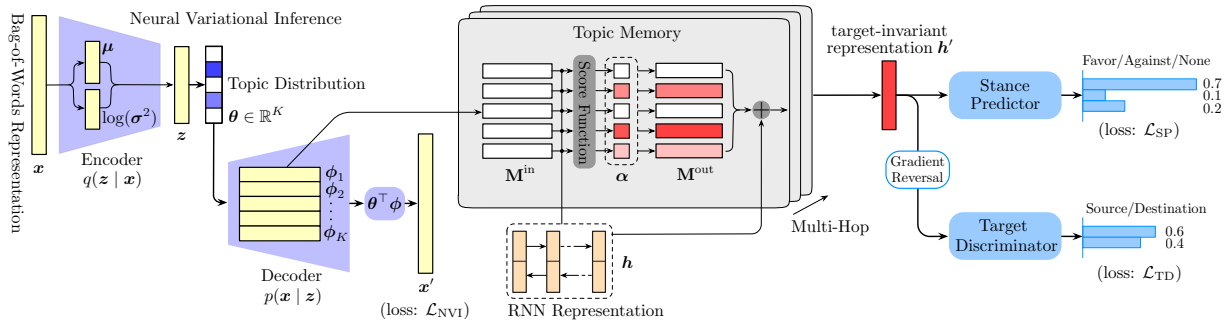

Figure 1: Overall structure of Variational Transfer Network for cross-target stance detection.

we adopt adversarial training technique [5] to optimize our model. The contributions of our work are as follows.

- We are among the first to leverage transferable topic knowledge for model adaptation across targets, and propose an effective method for cross-target stance detection.
- Our method acquires topic knowledge and learns target-invariant representations in an end-to-end framework, exploiting data from both source target and destination target.
- Experimental results show that our method outperforms the state-of-the-art methods in cross-target stance detection.

2 PROBLEM STATEMENT

We are given a set of **labeled** texts from a source target $\mathcal{D}_s^l = \{(x_s^{(i)}, y_s^{(i)})\}_{i=1}^{N_s}$ where x_s is a sentence and y_s is the stance label, and a set of **unlabeled** texts from a destination target $\mathcal{D}_d = \{x_d^{(i)}\}_{i=1}^{N_d}$. We also have a large set of unlabeled texts from both targets $\mathcal{D}_e = \{x_e^{(i)}\}_{i=1}^{N_e}$ (called external data), where x_e belongs to the source target or the destination target. Typically, $N_e \gg N_s \approx N_d$. Cross-target stance detection task is to predict the stance labels of texts in $\mathcal{D}_d$.

3 PROPOSED METHOD

We propose a **V**ariational **T**ransfer **N**etwork (VTN) for cross-target stance detection. Figure 1 shows the overall structure, which consists of two parts. The first part is a topic knowledge acquisition module, with the aim of exploiting large unlabeled data from both targets ($\mathcal{D}_e$) to acquire transferable topic knowledge. The second part is a topic-enhanced memory network with discriminator, which stores topic knowledge in memory and learns target-invariant text representations for stance prediction using $\mathcal{D}_s^l$ and $\mathcal{D}_d$. Two parts are integrated into an end-to-end training manner.

3.1 Topic Knowledge Acquisition with Neural Variational Inference

Transferable knowledge acquisition is the core function of our method, and we apply topic modeling that yields latent topics to acquire the knowledge for effective cross-target adaptation.

3.1.1 Topic modeling based on variational autoencoder. Let K denote the number of topics during the topic modeling process. For each topic k ($k = 1, 2, \ldots, K$), we introduce a topic embedding $\boldsymbol{t}_k \in \mathbb{R}^d$, which is a parameter vector to be learned. Our *topic knowledge* $\mathbf{T}$ consists of these topic embeddings $\mathbf{T} = \{\boldsymbol{t}_1, \boldsymbol{t}_2, \ldots, \boldsymbol{t}_K\}$. Given the word embedding matrix $\mathbf{E} \in \mathbb{R}^{V \times d}$ where V is the vocabulary size, we can obtain word distribution $\boldsymbol{\phi}_k \in \mathbb{R}^V$ for each topic k by computing the semantic similarity between the topic and words:

$$\boldsymbol{\phi}_k = \textsc{Softmax}(\mathbf{E}\boldsymbol{t}_k), \quad k = 1, 2, \ldots, K \tag{1}$$

We denote all topics' word distributions as $\boldsymbol{\phi} = (\boldsymbol{\phi}_1^\top, \boldsymbol{\phi}_2^\top, \ldots, \boldsymbol{\phi}_K^\top)$.

In our framework, to enable the topic modeling process to be trained via backpropagation, we employ neural variational inference (NVI) [9, 14] that implements LDA-style topic modeling [2] by variational autoencoder (VAE) [7].

Let $\boldsymbol{x} \in \mathbb{Z}_+^V$ denote the bag-of-words representation of a text $x \in \mathcal{D}_e$. Based on VAE, the generative process for each text $\boldsymbol{x}$ (with a length of N words) is:

1. Draw a latent variable $\boldsymbol{z} \in \mathbb{R}^K$ from the prior distribution (standard Gaussian): $\boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

2. Obtain the topic distribution $\boldsymbol{\theta} = \textsc{Softmax}(\mathbf{W}\boldsymbol{z}) \in \mathbb{R}^K$, where $\mathbf{W}$ is a learnable matrix.

3. For each word at position n ($n = 1, 2, \ldots, N$) of the sentence, draw a word $w_n \sim \boldsymbol{\phi}^\top \boldsymbol{\theta}$.

Here, the probability of drawing the word w_n is computed by:

$$p(w_n \mid \boldsymbol{\theta}, \boldsymbol{\phi}) = \sum_{k=1}^{K} p(k \mid \boldsymbol{\theta}) \cdot p(w_n \mid \boldsymbol{\phi}_k) = [\boldsymbol{\theta}^\top \boldsymbol{\phi}]_{w_n} \tag{2}$$

and thus we have $w_n \sim \boldsymbol{\phi}^\top \boldsymbol{\theta}$, where $\boldsymbol{\phi}^\top \boldsymbol{\theta} \in \mathbb{R}^V$.

Because posterior inference for $\boldsymbol{z}$ is intractable, VAE introduces a variational distribution $q(\boldsymbol{z} \mid \boldsymbol{x})$ to approximate the true posterior. Formally, $q(\boldsymbol{z} \mid \boldsymbol{x})$ is a diagonal Gaussian: $q(\boldsymbol{z} \mid \boldsymbol{x}) = \mathcal{N}(\boldsymbol{z}; \boldsymbol{\mu}, \boldsymbol{\sigma}^2 \mathbf{I})$, and $\boldsymbol{\mu}$, $\boldsymbol{\sigma}^2$ are parameterized by multi-layer perceptrons (MLP): $\boldsymbol{\mu} = \textsc{Mlp}_\mu(\boldsymbol{x})$, $\log(\boldsymbol{\sigma}^2) = \textsc{Mlp}_\sigma(\boldsymbol{x})$.

3.1.2 Objective function. During training process, VAE-based topic model aims to maximize the variational lower bound:

$$\text{ELBO}(\boldsymbol{x}) = \mathop{\mathbb{E}}_{\boldsymbol{z} \sim q(\boldsymbol{z}|\boldsymbol{x})} [\boldsymbol{x}^\top \log(\boldsymbol{\phi}^\top \boldsymbol{\theta})] - \mathbb{D}_{\text{KL}}[q(\boldsymbol{z} \mid \boldsymbol{x}) \parallel p(\boldsymbol{z})] \tag{3}$$

where the first term is to reconstruct the input $\boldsymbol{x}$, and the second KL-divergence term works like a regularizer that matches the variational posterior to the prior. Therefore, the objective function of the NVI network is: $\min_{\{\mathbf{T}, \mathbf{W}, \textsc{Mlp}_\mu, \textsc{Mlp}_\sigma\}} \mathcal{L}_{\text{NVI}} = -\sum_{x \in \mathcal{D}_e} \text{ELBO}(\boldsymbol{x})$.

3.2 Knowledge Transfer with Topic Memory

The NVI network induces topic knowledge $\mathbf{T}$. To take full advantage of it for effective knowledge transfer across targets, we store the topic knowledge in an external memory, and adopt a multi-hop memory network to learn text representation and predict stance.

3.2.1 Building topic memory. Traditional memory networks contain an input memory $\mathbf{M}^{\text{in}}$ and an output memory $\mathbf{M}^{\text{out}}$ [8, 15]. We employ a simplification version that only contains one memory matrix $\mathbf{M}$, i.e., $\mathbf{M} = \mathbf{M}^{\text{in}} = \mathbf{M}^{\text{out}}$, in our network. Specifically, we

initialize the external memory using $\mathbf{T}$, i.e., $\mathbf{M} = \mathbf{T}$, and further update the memory during training process.

3.2.2 Topic-enhanced text representation learning. Formally, the *topic memory* $\mathbf{M}$ consists of K slots $\{\boldsymbol{m}_1, \boldsymbol{m}_2, \ldots, \boldsymbol{m}_K\}$, and each slot $\boldsymbol{m}_k$ maintains the information of the corresponding topic k. Inspired by [15, 19], we utilize this external memory to incorporate topic knowledge into text representation, improving the model's adaptation ability across source target and destination target.

Given a sentence from $\mathcal{D}_s^l$, we use a bidirectional LSTM to encode it, and concatenate the hidden states of the last time step in two directions to obtain its vector representation $\boldsymbol{h}$. We then adopt attention mechanism to attend important memory slots (i.e., topics), yielding the enhanced text representation $\boldsymbol{h}'$:

$$\alpha_k = \textsc{Softmax}(\boldsymbol{m}_k^\top \mathbf{W}_1 \boldsymbol{h}), \quad \boldsymbol{o} = \sum\nolimits_{k=1}^{K} \alpha_k \boldsymbol{m}_k \tag{4}$$

$$\boldsymbol{h}' = \mathbf{W}_2 \boldsymbol{h} + \boldsymbol{o} \tag{5}$$

where α_k is the match score between the sentence and the topic k, and $\boldsymbol{h}$ is the query vector to access memory. The output vector $\boldsymbol{o}$ produced by topic memory encodes the topic knowledge used for enhancing $\boldsymbol{h}$. Matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ are parameters to be learned.

If we employ one-hop memory network, $\boldsymbol{h}'$ will be the final representation used to predict stance. We further extend it to multi-hop style, which utilizes the $\boldsymbol{h}'$ produced by preceding hop as the query vector for subsequent hop. We still use $\boldsymbol{h}'$ to denote the output text representation of the last hop.

3.2.3 Stance predictor. We adopt an MLP with softmax function as the stance predictor that outputs the predicted stance distribution $\hat{y} = \textsc{Softmax}(\textsc{Mlp}_{\text{SP}}(\boldsymbol{h}'))$. The memory network is trained by minimizing the cross-entropy loss on source target data $\mathcal{D}_s^l$:

$$\min_{\Theta_{\text{M}}} \mathcal{L}_{\text{SP}} = \sum\nolimits_{x \in \mathcal{D}_s^l} \textsc{Cross-Entropy}(\hat{y}, y) \tag{6}$$

where y is the ground-truth of the text x, and Θ_{M} is the parameter set of memory network. (i.e., LSTM, $\mathbf{M}$, $\mathbf{W}_1$, $\mathbf{W}_2$ and $\textsc{Mlp}_{\text{SP}}$).

3.3 Target-Invariant Representation Learning with Target Discriminator

To further make text representation $\boldsymbol{h}'$ more target-invariant to facilitate model adaptation across targets, we introduce a target discriminator to classify the target label of input text, i.e., belonging to source or destination target, which is also implemented by an MLP with softmax (denoted as $\textsc{Mlp}_{\text{TD}}$). The intuition is that if a strong target discriminator cannot predict a text's target label, its representation is target-invariant. Another advantage of introducing the discriminator is that we can exploit unlabeled data from destination target to train topic memory.

Specifically, given a sentence from $\mathcal{D}_s^l$ or $\mathcal{D}_d$, its representation $\boldsymbol{h}'$ aims to confuse the target discriminator and maximize the cross-entropy loss of target classification on $\mathcal{D}_s^l \cup \mathcal{D}_d$ (denoted as $\mathcal{L}_{\text{TD}}$), while the discriminator itself aims to minimize $\mathcal{L}_{\text{TD}}$. Thus, the training procedure of the memory network with discriminator is an adversarial training process, with the following minimax game:

$$\min_{\Theta_{\text{M}}} \max_{\textsc{Mlp}_{\text{TD}}} \mathcal{L}_{\text{SP}} - \lambda \mathcal{L}_{\text{TD}} \tag{7}$$

where λ is a trade-off parameter. We implement this procedure through a gradient reversal operation during backpropagation [5], a widely used technique in transfer learning-based models [3, 12].

Table 2: Summary statistics of evaluation datasets.

Target	# Unlabeled	# Labeled (Favor/Against/None)
Feminist Movement	6,070	949 (268/511/170)
Legalization of Abortion	3,173	933 (167/544/222)
Hillary Clinton	9,690	984 (163/565/256)
Donald Trump	11,301	707 (148/299/260)

3.4 Optimization

We adopt an end-to-end training manner to optimize our VTN. Note that it consists of two parts: the NVI network trained using $\mathcal{D}_e$, and the memory network with discriminator (MN-D) trained using $\mathcal{D}_s^l \cup \mathcal{D}_d$. Specifically, we iteratively train two parts: we first train two epochs for NVI, and use the trained topic embeddings to initialize the memory of MN-D and further train two epochs for MN-D. We then update the topic embeddings of NVI using the trained topic memory of MN-D, and train two epochs for NVI. The above procedure iterates until convergence.

4 EXPERIMENTS

4.1 Experimental Setup

4.1.1 Evaluation datasets and metric. We evaluate our method on SemEval-2016 Task 6 dataset [10]. We use four targets: Feminist Movement (FM), Legalization of Abortion (LA), Hillary Clinton (HC) and Donald Trump (DT). They are categorized into four source→destination settings for cross-target evaluation: FM→LA, LA→FM, HC→DT and DT→HC. Table 2 describes the statistics of datasets. When a target is used as destination target, we split its labeled data to obtain validation set and test set with the proportion being 3:7. We use the mean value of F_1 score for 'favor' and F_1 score for 'against' as the evaluation metric [10].

4.1.2 Baseline methods for comparison. We choose the methods from the related work of cross-target stance detection as our baselines: BiCond [1] and CrossNet [18]. For fair comparison, we further extend them to form two other baselines by adding a discriminator.

- BiCond [1]. This model learns target-dependent text representation by conditional encoding, using target representation to initialize the cell state of LSTM for encoding texts.
- CrossNet [18]. This model improves conditional encoding by incorporating a self-attention layer to capture important words in texts, achieving the state-of-the-art performance.
- BiCond w/ D. We extend BiCond to be a transfer learning-based model by adding a target discriminator.
- CrossNet w/ D. We extend CrossNet similarly.

Moreover, we compare VTN with its variant that removes the discriminator, named VTN w/o D. To further test the effect of the NVI network, we introduce two other VTN variants (Section 4.2.2).

4.1.3 Implementation details. We pretrain 100-dimensional word embeddings by Skip-gram and fix them during training. Following [9, 14], the number of topics K is 50. The output size of bidirectional LSTM is 100 (i.e., 50 for each direction). We employ 3-hop memory network. For training NVI (MN-D), the batch size is 256 (8), and the learning rate is 0.002 (0.001). During updating the discriminator ($\textsc{Mlp}_{\text{TD}}$), the learning rate is multiplied by 0.1 for more stable update. The trade-off parameter λ is 0.005.

Table 3: Results of cross-target stance detection. ▲ represents that VTN is statistically significantly better than the best baseline (Wilcoxon signed rank test, $p < 0.05$).

Model	Cross-target setting			
	FM→LA	LA→FM	HC→DT	DT→HC
BiCond	0.450	0.416	0.297	0.358
BiCond w/ D	0.461	0.401	0.386	0.353
CrossNet	0.454	0.433	0.400	0.362
CrossNet w/ D	0.404	0.418	0.462	0.363
VTN w/o D	0.468	**0.480**▲	0.430	0.363
VTN	**0.473**	0.478	**0.479**▲	**0.364**

Table 4: Results of methods for building memory.

Variant	Cross-target setting			
	FM→LA	LA→FM	HC→DT	DT→HC
VTN (rand. init)	0.440	0.379	0.422	0.349
VTN (pipeline)	0.469	0.451	0.399	**0.398**
VTN	**0.473**	**0.478**	**0.479**	0.364

4.2 Experimental Results and Analysis

4.2.1 Experimental results. The comparison results with baselines are shown in Table 3. In most cases, VTN w/ D shows poor performance compared with VTN. Similarly, BiCond and CrossNet generally performs poorly compared with BiCond w/ D and CrossNet w/ D, respectively. Thus, introducing a target discriminator is able to improve the adaptation ability of stance detection models.

The overall performance of our method outperforms baselines by a large margin, which indicates that VTN's adaptation ability is much higher than the previous state-of-the-art methods. This benefits from the knowledge acquisition mechanism (i.e., the NVI network for topic modeling) by exploiting data from both targets, and verifies that multi-hop memory can integrate external knowledge into feature representation effectively.

4.2.2 Effects of different methods for building topic memory. The topic memory in VTN is built using an iterative training manner (Section 3.4). To show its effect, we compare it with two other VTN variants. The first one is inspired by [17], named VTN (rand. init): we randomly initialize the memory in MN-D and update it during training, without using any knowledge from topic modeling. It is equal to removing the NVI network. The second one is a pipeline manner, named VTN (pipeline): we first train the NVI network until convergence to obtain topic knowledge, and then we use it to initialize the memory of MN-D and train the MN-D subsequently.

Table 4 shows that VTN (pipeline) and VTN perform much better than VTN (rand. init), and VTN also achieves better performance than VTN (pipeline). Therefore, topic knowledge acquired by NVI indeed improves model's adaptation ability across targets.

4.2.3 Case study. To qualitatively show the effect of topic knowledge acquired by NVI, we provide a case study. The sentence's stance in Table 5 is correctly predicted by VTN whereas incorrectly predicted by CrossNet. We visualize two topics (top-5 words for each one) with highest match scores (α_k) produced by VTN, and the attention weights produced by CrossNet. Clearly, our VTN notices crucial topics for knowledge transfer and enhancing the sentence representation. Although CrossNet computes word-level scores, it is hard to give correct prediction only based on them. Hence, incorporating topic knowledge can overcome the limitation of previous methods and improve the model's adaptation ability.

Table 5: Case study: effectiveness of topic knowledge.

Setting: Feminist Movement → Legalization of Abortion
Text: U know what isn't funny? Male politicians deciding what women should do with their body's. **Stance**: In favor

VTN prediction: In favor
law,choice,life,equality,social, right,equality,support,equal,unborn

CrossNet prediction: Against
U know what is n't funny ? Male politicians deciding what women should do with their body 's .

5 CONCLUSION

We propose an effective method VTN for cross-target stance detection, which leverages shared latent topics across targets as transferable knowledge to improve model's adaptation ability. Our method integrates topic knowledge acquisition and target-invariant representation learning in a unified end-to-end framework. Experimental results demonstrate the superiority of our proposed method.

ACKNOWLEDGMENTS

This work was supported in part by the National Key R&D Program of China under Grant #2016QY02D0305, NSFC Grants #71621002, #11832001 and #71702181, and CAS Key Grant #ZDRW-XH-2017-3. We thank all the anonymous reviewers for the valuable comments.

REFERENCES

[1] Isabelle Augenstein, Tim Rocktäschel, Andreas Vlachos, and Kalina Bontcheva. 2016. Stance detection with bidirectional conditional encoding. In *EMNLP*.
[2] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research* 3, Jan (2003).
[3] Daniel Cohen, Bhaskar Mitra, Katja Hofmann, and W Bruce Croft. 2018. Cross domain regularization for neural ranking models using adversarial learning. In *SIGIR*.
[4] William Ferreira and Andreas Vlachos. 2016. Emergent: A novel data-set for stance classification. In *NAACL-HLT*.
[5] Yaroslav Ganin and Victor Lempitsky. 2015. Unsupervised domain adaptation by backpropagation. In *ICML*.
[6] Myungha Jang and James Allan. 2018. Explaining controversy on social media via stance summarization. In *SIGIR*.
[7] Diederik P Kingma and Max Welling. 2014. Auto-encoding variational Bayes. In *ICLR*.
[8] Cheng Li, Xiaoxiao Guo, and Qiaozhu Mei. 2017. Deep memory networks for attitude identification. In *WSDM*.
[9] Yishu Miao, Edward Grefenstette, and Phil Blunsom. 2017. Discovering discrete latent topics with neural variational inference. In *ICML*.
[10] Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiao-Dan Zhu, and Colin Cherry. 2016. SemEval-2016 task 6: Detecting stance in tweets. In *SemEval*.
[11] Mitra Mohtarami, Ramy Baly, James Glass, Preslav Nakov, Lluís Màrquez, and Alessandro Moschitti. 2018. Automatic stance detection using end-to-end memory networks. In *NAACL-HLT*.
[12] Darsh J Shah, Tao Lei, Alessandro Moschitti, Salvatore Romeo, and Preslav Nakov. 2018. Adversarial domain adaptation for duplicate question detection. In *EMNLP*.
[13] Swapna Somasundaran and Janyce Wiebe. 2010. Recognizing stances in ideological on-line debates. In *NAACL-HLT Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*.
[14] Akash Srivastava and Charles Sutton. 2017. Autoencoding variational inference for topic models. In *ICLR*.
[15] Sainbayar Sukhbaatar, Arthur Szlam, Jason Weston, and Rob Fergus. 2015. End-to-end memory networks. In *NIPS*.
[16] James Thorne, Andreas Vlachos, Christos Christodoulopoulos, and Arpit Mittal. 2018. FEVER: A Large-scale dataset for Fact Extraction and VERification. In *NAACL-HLT*.
[17] Penghui Wei, Junjie Lin, and Wenji Mao. 2018. Multi-target stance detection via a dynamic memory-augmented network. In *SIGIR*.
[18] Chang Xu, Cecile Paris, Surya Nepal, and Ross Sparks. 2018. Cross-target stance classification with self-attention networks. In *ACL*.
[19] Jichuan Zeng, Jing Li, Yan Song, Cuiyun Gao, Michael R Lyu, and Irwin King. 2018. Topic memory networks for short text classification. In *EMNLP*.