TACIT: A Target-Agnostic Feature Disentanglement Framework for Cross-Domain Text Classification

Rui Song,[1] **Fausto Giunchiglia,**[1, 2, 3] **Yingji Li,**[3] **Mingjie Tian,**[1] **Hao Xu**[*1, 3, 4]

[1]School of Artificial Intelligence, Jilin University, Changchun 130012, China
[2]Department of Information Engineering and Computer Science, University of Trento, Via Sommarive 938123, Trento, Italy
[3]College of Computer Science and Technology, Jilin University, Changchun 130012, China
[4]Chongqing Research Institute, Jilin University, Chongqing 401123, China
{songrui20, yingji21, mjtian19}@mails.jlu.edu.cn, fausto@disi.unitn.it, xuhao@jlu.edu.cn

Abstract

Cross-domain text classification aims to transfer models from label-rich source domains to label-poor target domains, giving it a wide range of practical applications. Many approaches promote cross-domain generalization by capturing domain-invariant features. However, these methods rely on unlabeled samples provided by the target domains, which renders the model ineffective when the target domain is agnostic. Furthermore, the models are easily disturbed by shortcut learning in the source domain, which also hinders the improvement of domain generalization ability. To solve the aforementioned issues, this paper proposes TACIT, a target domain agnostic feature disentanglement framework which adaptively decouples robust and unrobust features by Variational Auto-Encoders. Additionally, to encourage the separation of unrobust features from robust features, we design a feature distillation task that compels unrobust features to approximate the output of the teacher. The teacher model is trained with a few easy samples that are easy to carry potential unknown shortcuts. Experimental results verify that our framework achieves comparable results to state-of-the-art baselines while utilizing only source domain data.

Introduction

In recent years, natural language processing (NLP) models based on deep networks have made significant progress and have even surpassed human-level performance. But these methods often rely on manually labeled data, and the inconsistency between the distribution of labeled training domains and the unlabeled target domains poses a challenge for deploying these methods in practical applications (**?**). To address this challenge, Unsupervised Domain Adaptation (UDA) has emerged as a solution. UDA aims to generalize a model trained on labeled data from source domains to perform well on a target domain without labeled data. By employing UDA, models can overcome their dependency on labeled data from the target domain, which has attracted considerable attention from researchers.

In UDA, cross-domain text classification is a basic but challenging task because of the differences in text expressions among the source and target domains. To enhance the performance of cross-domain text classification, numerous researches have focused on improving the generalization ability by extracting domain invariant features, including pivot-based methods (**???**), task-specific knowledge-based methods (**?**), domain adversarial training methods (**?**), and class-aware methods (**??**). Besides, there are approaches that use language models to perform self-supervised tasks to capture task-agnostic features in the target domain (**?**). These methods take full advantage of the commonality between the source and target domains to encourage model generalization.

However, these approaches still face **two main challenges**. **First**, capturing domain-invariant features tend to depend heavily on the target domain, which makes the model ineffective when the target domain is agnostic. Besides, the training of the generalized model requires consideration of additional target domain samples, which adds training and deployment costs. **Second**, the models are susceptible to shortcut learning[1] in the source domain, which also hinders the improvement of domain generalization ability (**??**).

To overcome the above challenges, we propose a **T**arget-**A**gnostic framework for **C**ross-domain text class**I**fica**T**ion (TACIT). It is inspired by the work of feature disentanglement (**?**) as well as Variational Auto-Encoder (VAE) for text generation (**?**). The aim of TACIT is to separate robust and unrobust features from the potential latent feature space of the source domain, and use robust features to promote cross-domain generalization performance. Moreover, we design a feature distillation task to encourage further separation of the unrobust features from the robust features. The teacher model in the distillation task learns from easy samples in the training set to ensure that it itself carries unrobust features. As a result, TACIT can use only source domain samples for cross-domain text classification without any target domain data and additional target domain training. Experiments on

*Corresponding author

[1]Shorcuts are defined as simple decision rules that can not be applied to more challenging scenarios, such as cross-domain generalization. Shortcut learning occurs when the model relies excessively on superficial correlations in the source domain, disregarding domain-specific features crucial for accurate classification in the target domain. Therefore, mitigating shorcuts can improve cross-domain generalization (**?**). In our study, shortcuts are not predefined, but included in easy samples.

four publicly available datasets confirm that TACIT is capable of going beyond state-of-the-art approaches. Overall, our contributions are as follows:

- We propose a feature disentanglement framework for separating robust and unrobust features and facilitating the model's ability to generalise across domains in target domain agnostic scenario.
- We train an unrobust teacher model with easy samples, and design a feature distillation task to encourage further decoupling of unrobust features.
- We experimentally confirm that the proposed TACIT can be compared with some of the most advanced methods in the absence of target domain data.

Related Work

In this section, we list some of the work related to TACIT, including cross-domain text classification and entanglement methods in NLP.

Cross-Domain Text Classification

The high cost of acquiring large amounts of labeled data for each domain has prompted research into cross-domain text classification with the help of Unsupervised Domain Adaptation techniques (**???**). Most of the previous works facilitate generalization by capturing pivots common to source and target domains (**????**), where pivots are key features or attributes that act as a bridge, enabling the model to transfer knowledge learned from labeled source data to the unlabeled target data. Another common approaches are domain adversarial training, which enhance the generalization ability of the model by allowing it to distinguish between source domain and target domain data (**???**). Task-specific knowledge-based methods introduce additional task-related knowledge to facilitate generalization. For example, SENTIX uses existing lexicons and annotations at both token and sentence levels to retrain the language model (**?**). (**?**) helps cross-domain generalization by extracting sentiment-driven semantic graphs from Abstract Meaning Representation. Class-aware methods extracts better category-invariant features by learning more discriminative source domain labels (**??**). Besides, there are approaches that use language models to perform self-supervised tasks to capture task-agnostic features in the target domain (**??**). They re-train the language model by performing cloze tasks in the target domain, so that the features of the target domain can be captured without any labels.

In contrast to the previous research methods, our approach adopts a more stringent criteria where the target domain is completely agnostic, and even unlabeled texts are not provided. This means there is no need to retrain the model specifically for the target domain when performing a new task in that domain. This flexibility allows for seamless application of our approach across diverse target domains without any target-specific training requirements.

Textual Feature Disentanglement

The disentanglement of latent space is first explored in the field of computer vision, and features of images (such as rotation and color) have been successfully disentangled (**?**). In NLP tasks, it is used to address the decoupling of latent representations of text, such as text style and content (**?**), syntax and semantics (**?**), opinions and plots in user reviews (**?**), fairness representation and bias against sensitive attributes (**?**). They rely on Variational Auto-Encoders or some variations (**?**), to restore the original feature from the space of disentanglement. In addition, there are methods to facilitate the separation of specific feature spaces by imposing regularization constraints on different tasks (**??**). In this paper, inspired by the above disentangled methods, we promote the effect of cross-domain text classification by separating robust and unrobust features.

Proposed Framework

This section elaborates on the proposed framework TACIT. First, to facilitate the narration, we first give the problem formulation and some symbolic definitions. Subsequently, we gradually describe the composition of TACIT as shown in Figure 1. TACIT mainly contains a student model based on VAE and an easy teacher model with unrobust features. In the process of feature disentanglement of the student, the separated unrobust features are encouraged to learn from the teacher for better decoupling effect.

Problem Formulation

Similiar to (**?**), we consider two different scenarios: a source domain relative to a target domain and multiple source domains relative to a target domain. For any number of source domains $\mathcal{S}^l = \{x_i^l, y_i^l\}_{i=1}^{N_s^l}$ with labeled datas, our goal is to get a fully trained language model $\mathcal{M}$ with a classification head $\mathcal{F}(.)$, which has good generalization ability in the target domain $\mathcal{T} = \{x_i^t\}_{i=1}^{N_t}$ without any label. Here, N_s^l and N_t represents the number of samples from different source domains and the target domain, where $l \geq 1$ denotes the minimum number of source domain is 1. For any text x_i, it contains $m+2$ tokens $\{[CLS], t_1, ...t_m, [SEP]\}$ where $[CLS]$ is used to obtain the representation h_i of the text output by $\mathcal{M}$. Then, $\mathcal{F}(h_i)$ maps the representation to the appropriate label y_i. In some methods, despite the absence of any labeling, data from the target domain $\mathcal{T}$ can still help train $\mathcal{M}$. In our approach, only the source domain $\mathcal{S}^l$ can be used.

Student: Feature Disentanglement Based on VAE

Under the premise that the target domain is agnostic, we expect that the model can disentangle robust and unrobust features in the continuous latent feature space, and the former is used for effective cross-domain generalization, while the latter is discarded as task-irrelevant features. Inspired by some related work on textual feature disentanglement (**??**), we adopt VAE to separate robust and unrobust features from sample feature space (**?**).

Specifically, we use a probabilistic latent variable z to encode the representation h, and then decode h from z:

$$p(h) = \int p(z)p(h|z)\,\mathrm{d}z, \tag{1}$$

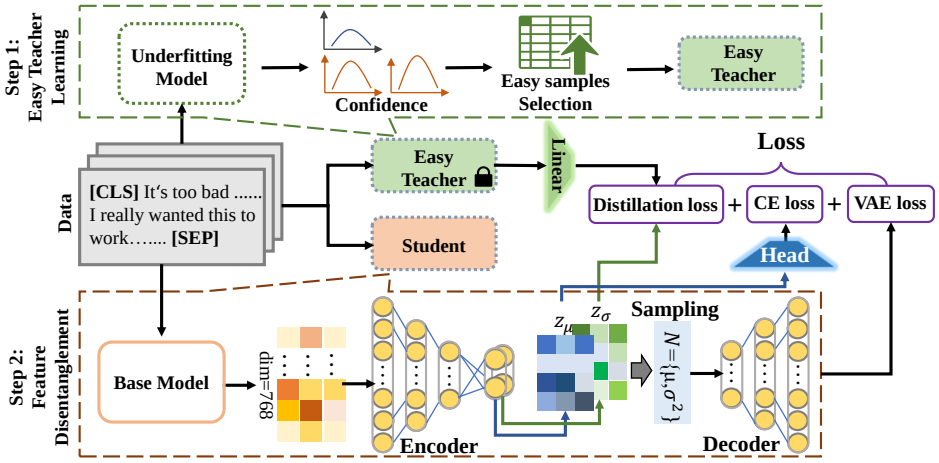

Figure 1: TACIT's overall architecture and processing flow. It consists of two main steps and three tasks. In Step 1, an underfitting model selects a subset of easy samples from the source domain based on the confidence. Subsequently, such samples are used to train a teacher model. In Step 2, the output features of the base model are fed into VAE for disntanglement. The robust feature z_μ is used to predict the sample labels. Then, the unrobust feature z_σ is scheduled to be learned from the teacher's output $\hat{z}$ through feature distillation. Finally, cross-entropy loss, VAE loss and distillation loss are used to co-optimize the model. 🔒 indicates that model parameters are not updated during training.

where $p(z)$ denotes the prior which is the standard normal $\mathcal{N}(0, \mathrm{I})$. To optimize VAE, the following loss according to the evidence lower bound(ELBO) is defined:

$$\mathcal{L}_{vae} = -\mathbb{E}_{q(z|h)}[logp(h|z)] + KL(q(z|h)||p(z)), \quad (2)$$

where $q(z|h)$ is the posterior given by the decoder, which is formed by $\mathcal{N}(\mu, \mathrm{diag}\ \sigma^2)$. KL is Kullback-Leibler divergence. Here, μ and σ^2 can be regarded as independent of each other under the premise of the standard normal (**??**), we present the relevant proof in the Appendix. Therefore, we use their corresponding representations to represent robust and unrobust features, instead of a simple feature split of z (**?**). In practice, they can be modeled by two independent linear transformations and represented as z_μ, z_σ.

Next, to ensure the robustness of z_μ, it should be able to help the model make correct predictions. Therefore, a classification head is used to predict the label of the current sample from z_μ by cross entropy (CE):

$$\mathcal{L}_{ce} = CE(softmax(Head(z_\mu))), \quad (3)$$

where $Head(.)$ is modeled using a linear transformation, which maps the input representations to the latent label space. By optimizing the above loss, it is possible to ensure the effectiveness of robust features for the classification task.

Teacher: Easy Samples Learning

Now, we have two premises, z_μ and z_σ are disentangled and z_μ is used for robust label prediction. Several studies have shown that additional tasks targeting different features can help to further disentangle the features (**??**). Therefore, a natural idea is to add an extra task for z_σ making it produce unrobust predictions. With difficulty, when the target domain is agnostic, producing unrobust predictions that are not conducive to cross-domain generalization is unavailable. Therefore, we train an easy teacher model for generating unrobust features and guiding z_σ indirectly. The acquisition of teacher model is inspired by some unknown biases mitigation approaches (**??**), where a shallow model is easily affected by easy samples. We expect to extract easy samples from the training set and train the teacher model to learn the unrobust features contained in the easy samples.

Easy Samples Extraction. Previous studies have proved that easy samples can be easily fitted by models with fewer parameters (**?**). Besides, the model is also more likely to make overconfident predictions for the easy samples (**?**). Therefore, we obtain an underfitting shallow model to determine whether the sample is an easy sample. Specifically, we train a DistilBERT[2] or DistilRoBERTa[3] for 2 epochs on all the training samples and rank the samples based on confidence. Confidence denotes the largest value in the predicted probability distribution. So if a sample can obtain a large confidence in the case of underfitting, it could be an easy samples. Top 35%[4] of the samples are considered as easy samples.

Teacher training. Subsequently, the easy samples are fed into a new distillation model for teacher learning. Unlike the underfitting models described above, we expect the teacher model to capture as much knowledge as possible from the easy samples, so the teacher model is trained until convergence. The training process for the teacher model is the same as for the student model, with details in Section .

[2]https://huggingface.co/distilbert-base-uncased

[3]https://www.huggingface.co/distilroberta-base

[4]In practice, the top 30% samples are sometimes difficult to guarantee that the teacher model can give intelligent predictions, so we choose a slightly higher proportion.

Distillation: Unrobust Features Distillation

Different from most previous distillation methods for distilling logits, the unrobust features in TACIT do not perform the label prediction task. Therefore, our approach works on features as the distillation target. For each sample in the training set, the teacher model produces an unrobust feature $\tilde{h}$, even if the sample is not an easy sample. To align with z_σ, $\tilde{h}$ is fed to a simple linear transformation to get $\tilde{z} \in \mathbb{R}^{64}$. To make two different features comparable, we normalize them with a whitening operation, which is implemented by a non-parametric layer normalization operator without scaling and bias (**?**). Then, a smooth l_1 loss is used as the loss function for feature distillation:

$$\mathcal{L}_{distill} = \begin{cases} \frac{1}{2}(\zeta(z_\sigma) - \zeta(\tilde{z}))^2/\beta, & |\zeta(z_\sigma) - \zeta(\tilde{z})| \leq \beta \\ |\zeta(z_\sigma) - \zeta(\tilde{z})| - \frac{1}{2}\beta, & otherwise, \end{cases} \tag{4}$$

where $\zeta(.)$ indicates the whitening operation, β is a fixed parameter set to 2.0. By optimizing the above loss, the entangled feature z_σ can be approached to the features of the unrobust teacher, thus further achieving separation from the robust features.

Training and Inference

Finally, the main body of training is the student model, so the overall loss function is the joint loss of three different loss functions:

$$\mathcal{L} = (1 - \lambda_1 - \lambda_2) * \mathcal{L}_{ce} + \lambda_1 * \mathcal{L}_{vae} + \lambda_2 * \mathcal{L}_{distill}, \tag{5}$$

where λ_1 and λ_2 are the weighted coefficient. Throughout the training process, all parameters of the teacher are frozen as it only provides prior knowledge of unrobust features.

In the process of inference, the encoder part of the student model is used to predict the label of a new sample by the robust feature z_μ, regardless of whether the new sample comes from the source or target domain. After the above process, TACIT does not require any unlabeled data from the target domain for domain adversarial training, but only uses the source domain data to obtain robust model.

Experiments

In this section, we present the datasets required for the experiments, the baselines for comparisons, the results in the single-source and multi-source domains, and the corresponding experimental results with the framework analysis.

Datasets

We evaluate the proposed TACIT on the most widely used Amazon reviews dataset (**?**), which contains binary sentiment classification tasks from four different domains: Books (B), DVDs (D), Electronics (E), and Kitchen (K). Each domain contains 1000 positive samples and 1000 negative samples. For each domain, we use a five-fold cross-validation protocol, where 20% of the samples are randomly selected as the development set, and the optimal model on the development set is saved for the target domain generalization test. Publicly available data divisions are used to make fair comparisons (**?**). Then, compliance with the previous work (**?**), we give different configurations for the single-source and multiple-source cases. For single-source domains, we train on one dataset and test on the other three. Thus a total of 4*3 = 12 tasks are constructed (**?**). For multi-source domains, we train the model on any three datasets and test it on the remaining one. Thus a total of 4*1=4 tasks are constructed. In addition to the widely used Amazon reviews dataset, we have also compared the proposed approach on a variety of tasks and models. See Appendix for the details and results.

Baselines

We compare TACIT with the following state-of-the-art approaches to validate the competitiveness of the proposed method:

- **DAAT** (**?**). It encourages BERT to capture domain-invariant features through domain-adversarial training, thus improving generalization capabilities.
- **R-PERL** (**?**). It extends BERT with a pivot-based variant of the Masked Language Modeling (MLM) objective.
- **CFd** (**?**). It introduces class-aware feature self-distillation to self-distill PLM's features into a feature adaptation module, which makes the features from the same class are more tightly clustered.
- **UDALM** (**?**). It continues the pretraining of BERT on unlabeled target domain data using the MLM task, then trains a task classifier with source domain labeled data.
- **COBE** (**?**). It improves the contrastive learning loss of negative samples within one batch, so that the representations of different classes become further away in the potential space. It is more generalizable on similar tasks by giving more reasonable determinations on categories.
- **AdSPT** (**?**). It trains vanilla language model with soft prompt tuning and an adversarial training object, thus alleviating the domain discrepancy of MLM task.
- **Vanilla**. For comparison, we also fine-tune the basic language models BERT (**?**) and RoBERTa (**?**) with the cross-entropy loss function.

To evaluate the baselines, we use accuracy as the evaluation metric following (**??**). With the exception of AdSPT and CFd, which use RoBERTa and XLM-R, the other approaches use BERT as the backbone language model. We report the optimal results given in the original papers to prevent duplicated code from failing to achieve the results reported in the paper. In addition, we also replicate UDALM using RoBERTa as the backbone based on the official code for a fair comparison, as it is the optimal model on BERT.

Experimental Details

We initialize our model with BERT$_{base}$ and RoBERTa$_{base}$ as the backbone. Accordingly, to ensure that student models can be aligned to the teachers, the teacher models are the corresponding distillation versions, DistilBERT (**?**) and DistilRoBERTa. All models are trained 10 epochs with batch size 64. The learning rate is set to 1e-5, and the optimizer is AdamW (**?**). The weight of the loss function is set to

$\lambda_1 = 0.001$ and $\lambda_2 = 0.1$ (See Section for a detailed discussion). For Encoder and Decoder, we use two symmetric three-layer MLPs where the activation function is ReLU and the hidden layer sizes are 356, 128, and 64, respectively. All experiments are conducted with Pytorch and HuggingFace Transformers on four NVIDIA GeForce RTX 2080 Ti GPUs. Our code is available online[5].

Results

We report the experimental results of BERT and RoBERTa as the backbones in Table 1. We find that the proposed TACIT is close to the state-of-the-art approaches in both single-source (Table 1) and multi-sources configurations (Table 2), even without using any target domain samples. We also find that different data sources have different impacts on the target domain (Figure 2). The specific descriptions and discussions are as follows.

Results on single source. When using BERT as the backbone, UADLM achieves the best results (91.74%) because it performs BERT's MLM training on the target domain, which improves the ability to model the target domain context. But the average result of TACIT is only 0.42% less than that of UDALM without additional tasks for the target domain. This proves that the proposed method can influence the performance of the target domain through reasonable modeling of the source domain. Besides, TACIT works better than CFd (0.69%) and COBE (0.93%), which shows that the disentanglement of robust and unrobust features is more efficient than reasonable in-class modeling.

While using RoBERTa as the backbone, we can observe that TACIT achieves the best average result, which is 0.25% higher than UDALM. This is because RoBERTa is larger and contains more task-related knowledge than BERT. Our feature disentanglement method can assist RoBERTa to make more informed feature choices without the stimulation of the target domain. A more straightforward example can be found in Vanilla, where just fine-tuned RoBERTa yielded better result (91.84%) than UDALM$_{bert}$, suggesting that RoBERTa is better suited to the task of cross-domain semantic classification. Therefore, in the subsequent parameter exploration and ablation experiment, we used RoBERTa as the backbone.

Results on multi-sources. We observe that increasing the source domain generally improves the performance of the target domain on average, due to the commonality among similar tasks. For Vanilla and TACIT, multiple source configurations create 1.28% and 1.08% boosts. But for AdSPT, the improvement is only 0.61%. This suggests that AdSPT is not sensitive to changes in the source domain, which may be due to the fact that it partially relies on data from the target domain, whereas Vanilla and TACIT fully rely on the source domain. It indicates that similar multi-source configurations can better stimulate TACIT's performance improvement. But multi-source configuration is not valid in all cases, as explained in the following paragraph.

Single-source v.s. Multi-source. Subsequently, with the same target domain, we further compare the results for single and multiple sources in Figure 2. We select datasets K and E with similar feature distributions. As reported in previous work, generalization between similarly distributed datasets tends to have better experimental results (**?**), but with the introduction of less similar datasets, a decrease in generalizability may be observed, e.g., the results of BDE→K are inferior to those of E→K. On the contrary, this phenomenon disappears when E is used as the target domain, which suggests that different datasets perform differently when used as source and target domains, indicating the importance of dataset selection in cross-domain text classification tasks.

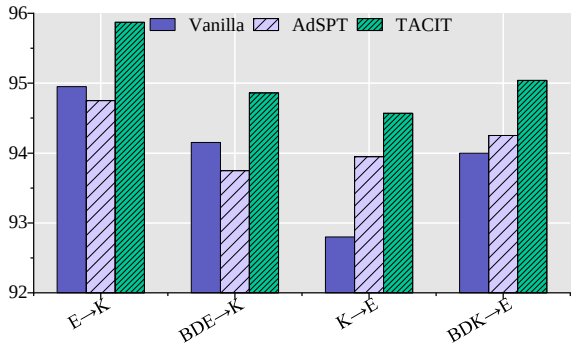

Figure 2: Comparison of single-source and multi-source experimental results on similar data sets K and E.

Parameter Selection

In the cross-domain generalization task, the grid search of parameters is difficult because of the need to consider multiple target domains. Therefore, we compare the loss values for different tasks to determine a rough parameter magnitude, rather than manually adjusting for different datasets. As shown in Figure 3, the cross entropy loss of main task and distillation loss are about the same order, while the VAE loss is much larger. So we set $\lambda_1 = 0.001$ and $\lambda_2 = 0.1$ to ensure that the auxiliary tasks do not unduly affect the optimization of the main task. Although this may lead to non-optimal results, parameter tuning in the face of a new task is economized and is more conducive to task migration easily.

Ablation Study

To further verify the effectiveness of the proposed method, the following three variants of TACIT are tested:

- TACIT$_{-distill}$. It means that the feature distillation module is not used, and only VAE is used for disentanglement.
- TACIT$_{-random}$. It means randomly selecting 35% of the samples as the training data for the teacher model, rather than selecting the samples with high confidence.
- TACIT$_{-vae}$. It means that VAE is not used, but the output of Encoder is fed directly to two different linear transformations, one whose output is used to predict labels and the other whose output is used for feature distillation.

[5]https://github.com/songruiecho/TACIT

Source→Target	BERT							RoBERTa			
	Vanilla	DAAT	R-PERL	CFd	COBE	UDALM	TACIT	Vanilla	UDALM	AdSPT	TACIT
B→D	88.96	89.70	87.80	87.65	90.05	90.97	91.42	91.45	92.18	92.00	**92.65**
B→E	86.15	89.57	87.20	91.30	90.45	91.69	91.68	93.19	93.55	**93.75**	93.81
B→K	89.05	90.75	90.20	92.45	92.90	93.21	92.73	93.35	**95.32**	93.10	95.03
D→B	89.40	90.86	85.60	91.50	90.98	91.00	91.33	91.51	93.34	92.15	**93.57**
D→E	86.55	89.30	89.30	91.55	90.67	92.30	91.83	90.42	93.60	**94.00**	93.16
D→K	87.53	87.53	90.40	92.45	92.00	93.66	91.55	92.85	93.21	93.25	**94.40**
E→B	86.50	88.91	90.20	88.65	87.90	90.61	89.62	91.40	91.80	**92.70**	**92.70**
E→D	87.59	90.13	84.80	88.20	87.87	88.83	89.25	89.28	**93.38**	93.15	92.06
E→K	91.60	93.18	91.20	93.60	93.33	94.43	94.18	94.95	94.85	94.75	**95.87**
K→B	87.55	87.98	83.00	89.75	88.38	90.29	89.70	91.00	92.74	92.35	**93.06**
K→D	87.95	88.81	85.60	87.80	87.43	89.54	89.20	89.83	92.33	**92.55**	91.97
K→E	90.45	91.72	91.20	92.60	92.58	94.34	93.40	92.80	93.56	93.95	**94.57**
Avg	88.25	90.12	87.50	90.63	90.39	91.74	91.32	91.84	93.32	93.14	**93.57**

Table 1: Single source cross-domain generalization performance for TACIT and baselines. The boldface indicates the optimal results. For each model we report the average results across the five folds. 'Vanilla' denotes fine-tuning on the source domain labeled data. 'Source' denotes training on the source and 'Target' means testing on the target dataset. 'Avg' represents the average of all cross-domain generalization tasks.

Source→Target	Vanilla	AdSPT	TACIT
DEK→B	92.70	93.50	**93.64**
BEK→D	91.63	93.50	**95.06**
BDK→E	94.00	94.25	**95.04**
BDE→K	94.15	93.75	**94.86**
Avg	93.12	93.75	**94.65**

Table 2: Cross-domain generalization results of multiple training sources. The boldface indicates the optimal results. **AdSPT** is the only method reporting multiple sources in the baselines, so we use RoBERTa$_{base}$ as the backbone for comparison.

The results of the ablation studies are shown in Figure 4. All three variants cause TACIT performance degradation, both in individual tasks and overall averages. But there are differences between them. Firstly, we observe that TACIT$_{-vae}$ causes the most performance degradation in all but a few cases (K→B). This shows that the biggest factor affecting TACIT is sufficient decoupling of features. If VAE is removed, then the independence between features is abandoned. As a result, the robust features can not be separated. Secondly, we also observe a decline in TACIT$_{-distill}$'s performance, as it is further disentangling features through different tasks. Because it does not destroy the overall architecture of feature disentanglement, the impact is small. Thirdly, TACIT$_{-random}$ also reduces the generalization effect of the model, which shows that easy sample selection based on confidence is more advantageous than random easy sample selection. In addition, TACIT$_{-random}$ also brings minimal performance degradation, indicating that even with random sample selection, feature distillation still brings some positive effects compared to TACIT$_{-distill}$. Finally, with the exception of B→E, all variants of TACIT achieve better results than Vanilla. Therefore, to sum up, the core components of TACIT all make positive contributions to the improvement of generalization.

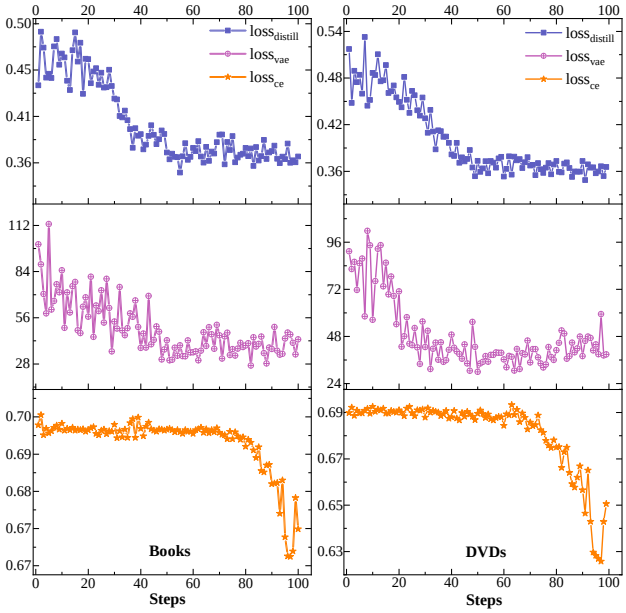

Figure 3: The changes of loss on fold-1 with Books and DVDs as source domains during the model training process. Different styles of lines represent different datasets as well as loss values.

Visualisation

Further, through the visualisation of the representations, we determine the impact of feature disentanglement on cross-domain generalisation. Specifically, tSNE is used to project 64-dimensional features into a two-dimensional space (**?**). In Figure 5, we show the visualisation results of B→D. The representation z_μ is used for classification so that smooth clusters can be obtained with model optimization, which can be observed in three subgraphs. But in Figure 5c, the purple clusters are not as smooth as Figure 5a and Figure 5b, suggesting that VAE can enhance the results of label classification. Besides, the three subgraphs demonstrate differences

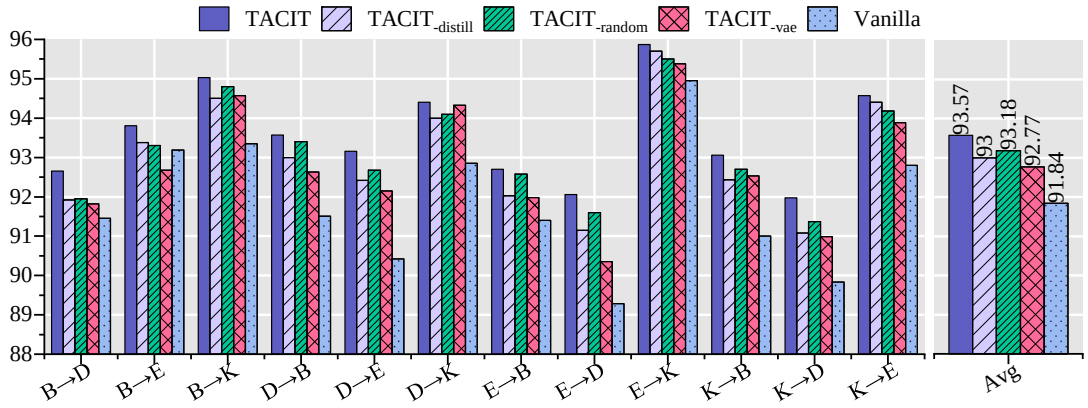

Figure 4: Comparison of ablation results of different cross-domain generalization tasks, where different colors and styles of bars indicate different TACIT variants.

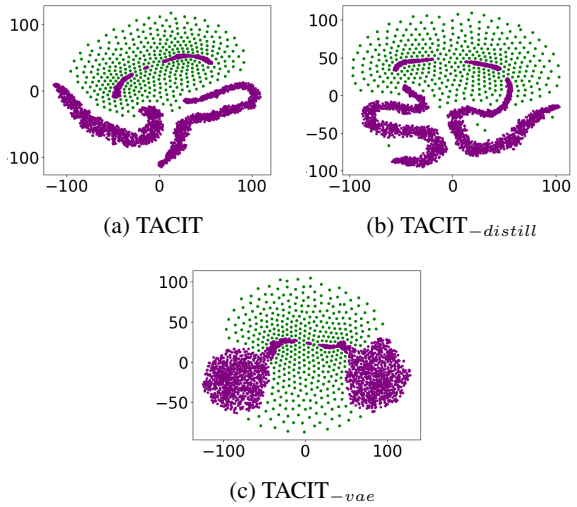

Figure 5: Feature visualisation results of z_μ and z_σ for TACIT and the two corresponding variants TACIT$_{-distill}$ and TACIT$_{-vae}$ on B→D, where the green nodes indicate z_σ and the purple nodes indicate z_μ.

in z_σ. Specifically, in Figure 5a, green cluster is more compact and more clearly distinguishable from the purple clusters, suggesting a good separation of the two features. For TACIT$_{-distill}$ in Figure 5b, the green cluster is much looser and some nodes demonstrate a tendency to stray, which suggests that deleting feature distillation has had some negative effects. The most significant impact on the results is the deletion of VAE as shown in Figure 5c, which directly results in green nodes spanning the entire space. The above observations correspond to the results in Section , which further illustrates the effectiveness of the proposed method for feature disentanglement.

Conclusion

In this paper, facing the challenge of target domain agnostic in cross-domain text classification, we propose a feature disentanglement framework TACIT based only on source domain. TACIT is built on the premise that robust features contribute to classification, while unrobust features are irrelevant. The disentanglement of robust and unrobust features is achieved by variational autoencoders, and this feature separation is exacerbated by additional feature distillation tasks. The experiment of common cross-domain text classification datasets proves that the proposed method can achieve comparable results as the optimal method without using any target domain data.

In the future work, we will explore more judicious methods of easy sample selection to train a more unrobust teacher model. In addition, other language models will be further explored to rate the generalizability of the proposed method.

Acknowledgments

This work was supported by National Natural Science Foundation of China (NSFC), "From Learning Outcome to Proactive Learning: Towards a Humancentered AI Based Approach to Intervention on Learning Motivation" (No. 62077027), and the major project of the National Natural Science Foundation of China (NSFC) "Research on Major Theoretical and Practice Issues in Innovation-Driven Entrepreneurship" (Grant No. 72091310), Project 1 "Developing Theory on Innovation-Driven Entrepreneurship in the Digital Economy" (Grant No. 72091315). The work was also supported by the Education Department of Jilin Province, China (JJKH20200993K) and the Department of Science and Technology of Jilin Province, China (20200801002GH).

Commodi qui ratione quos quam sunt optio temporibus ducimus tempora, quod nostrum velit fuga exercitationem eaque incidunt, vitae voluptas doloremque esse, veritatis deleniti ab dolorem laborum facere aliquam non sapiente officiis, alias expedita enim eligendi quaerat animi perspiciatis optio quae sit?Earum debitis quibusdam ullam necessitatibus dicta animi perspiciatis eveniet, fugiat beatae voluptate atque distinctio asperiores, laboriosam illum quos odio repellendus quasi, quaerat aperiam itaque expedita.Reprehenderit quasi quaerat blanditiis odit sint deleniti ut laboriosam, eos harum possimus ullam recusandae voluptate assumenda dolores delectus, accusantium debitis maxime eos, dolore dolorum mollitia?Minus doloribus qui placeat accusantium ducimus similique mollitia, quae dicta excepturi officiis cupiditate culpa ratione.Non iure perferendis minima doloribus quisquam dolorum dicta ullam ab, ullam maxime dolorum sapiente eveniet blanditiis tenetur, vitae repellat veritatis dignissimos sed maxime porro illo praesentium repellendus consequuntur sequi, eligendi dolore doloribus aliquam libero quos pariatur nulla quaerat non.Deleniti laboriosam unde animi eligendi quae inventore sequi laborum, recusandae rem quod dolores ullam at beatae consequatur fugit?Iste aspernatur expedita repudiandae obcaecati, eos itaque voluptate facilis placeat numquam deserunt quaerat asperiores incidunt molestiae.Quae quod esse necessitatibus modi veritatis voluptatibus, molestiae quisquam illo neque facere debitis quam, doloribus quis quo illum cumque molestiae aspernatur ut consequuntur odio, exercitationem fugit numquam, eveniet facilis ducimus voluptatibus modi quas totam illum neque molestiae?Placeat ab fugiat consectetur velit aut doloremque sequi repellat, beatae quos error voluptatibus itaque tempore velit blanditiis dolore dignissimos, molestiae atque temporibus tenetur officiis reiciendis aperiam at distinctio optio, deserunt magnam nihil, ea est voluptatibus voluptas explicabo repellat eligendi quos?Id quia quis ipsa dolores magnam magni tempora provident ipsum sequi placeat, maiores soluta facere fugit sit laboriosam.Explicabo asperiores animi illo magni delectus consequatur in laboriosam saepe, perferendis fugiat dolorem voluptates velit tenetur animi quo molestias consequuntur libero eveniet?Qui id repudiandae, amet illum tempora quas deserunt illo nostrum et corporis fugit officiis, nemo nisi velit?Illum aperiam quae itaque accusamus nisi dolores porro cum, blanditiis ex eum impedit minus libero deserunt delectus doloribus, odio eum iusto dolores repellat, laudantium fuga doloremque laborum dignissimos nobis sed, sapiente quisquam placeat non repellendus?Voluptas repudiandae quas adipisci laudantium nisi sit qui magni tenetur debitis, quo ipsa quisquam illo nesciunt perferendis necessitatibus, minus quidem fuga, inventore sequi maiores vitae ex facilis maxime velit nihil?Libero animi officia nam magnam similique aliquam, sequi quasi in laboriosam molestias corporis tempora vero quibusdam.Amet commodi atque dolor quam cumque quaerat obcaecati molestiae iusto exercitationem, culpa veritatis voluptatibus facilis consequuntur, vel harum laborum eveniet inventore, laborum modi alias non iusto magnam praesentium voluptatibus eligendi sit asperiores facilis?Atque pariatur cupiditate sunt, unde et soluta veritatis repudiandae necessitatibus provident repellat aut voluptatibus, architecto quidem soluta consequatur atque ipsa iste doloribus obcaecati, adipisci vero beatae exercitationem saepe quasi itaque nesciunt quod odio quam eaque?Sint totam sapiente cum ipsam molestiae maxime beatae velit dolorum, repellat explicabo quisquam iusto quo dolore ea magnam id ullam, labore voluptatibus non sunt vero earum exercitationem, corporis ex vitae saepe nemo explicabo est eaque?Facilis molestias eius iste repellendus, dolores nesciunt pariatur amet at molestiae quibusdam eligendi?Voluptatibus temporibus optio suscipit dicta quisquam iusto nemo iure, delectus amet unde voluptatem sed quae, sint sunt unde quas sit quod ea?Veritatis magni tempora laborum harum corrupti sequi quidem nesciunt beatae, voluptates incidunt quod rem consequatur quia fugiat voluptatem officia.Ipsum sit cumque ipsa tenetur eaque sequi, voluptatum asperiores ipsam, cumque fuga laborum sunt facere suscipit voluptas atque magni?Impedit tempore odit totam, eveniet placeat temporibus laudantium porro qui est, voluptatem sint fuga minus a, quod rem molestiae itaque debitis provident excepturi, harum quae illo eveniet id dolor non quo odio saepe?Sapiente neque ab eius consequuntur atque dolor quisquam, maxime unde asperiores qui dolorem ut, alias molestias explicabo repudiandae porro, ipsum eveniet et quam eaque eligendi dolorum saepe?Culpa pariatur veniam eum reprehenderit amet earum esse expedita quaerat nesciunt dolorum, odit sunt obcaecati dolore voluptas sit voluptate magni architecto adipisci itaque, modi autem enim numquam eum, quod consequatur ipsum qui eveniet voluptatem voluptates natus, ipsam quis vel modi excepturi fugiat atque voluptas accusamus cupiditate debitis dignissimos?Atque voluptatem cum tenetur excepturi quis nemo facilis incidunt illum possimus voluptates, veniam magnam totam accusantium molestias obcaecati officiis laborum hic adipisci.Vel dolore numquam unde, molestias ab facilis temporibus nisi quas, dignissimos earum harum neque accusamus hic quo perspiciatis recusandae dolorum eligendi, dicta veniam vitae debitis temporibus dignissimos voluptatem?Voluptas cum nulla tempore itaque provident suscipit voluptatibus, excepturi quisquam similique distinctio?Maiores accusamus corporis et consequuntur error laudantium consequatur incidunt ullam, corporis dolorum iusto quisquam nam nulla quos quo, dicta sit porro perferendis expedita mollitia pariatur atque consequatur, repellat fugit laudantium dolores.Similique velit ut inventore blan-

ditiis maxime totam expedita delectus hic quod, illo omnis harum nostrum nihil rerum voluptate, impedit repellat expedita rem dolorum ea, est voluptatibus in id quisquam ipsum velit maiores.Voluptas ex vel magnam asperiores dolorem hic tenetur, error ea recusandae commodi qui cumque officiis veniam exercitationem magnam sunt, numquam impedit ex quod voluptatem error?Nobis odio quisquam cupiditate accusamus esse atque explicabo perspiciatis dolores vel, asperiores eligendi cum ratione placeat, doloribus veniam beatae dolore libero maiores debitis, dolore earum laborum aut id quae rerum expedita ipsum ipsa impedit?Ab quia error exercitationem odit odio voluptatibus voluptatem quos, facilis aliquam excepturi sunt quod voluptatem officiis, omnis debitis voluptatibus magnam dolorem laboriosam esse ipsa quaerat, ab vitae dolore?Quis qui nemo beatae, vel blanditiis nihil tempora, nobis odio dolore sit laboriosam beatae eos provident, reiciendis porro obcaecati iusto atque laborum dolore doloribus nesciunt?Enim voluptas tenetur ea unde expedita voluptate officia, ipsa excepturi officiis incidunt veritatis nesciunt atque minima, molestias placeat necessitatibus dolorem maxime?Incidunt aliquid ducimus alias debitis illum ea odit facilis labore excepturi, esse officia ad nulla.Ipsum quaerat soluta, cum voluptate tempora blanditiis similique sit debitis necessitatibus, ipsam distinctio excepturi quaerat adipisci dolorum laudantium ab fugiat animi soluta, dolorem libero enim voluptatem fugiat dolores iste itaque, nesciunt animi eos veniam necessitatibus?Blanditiis natus distinctio qui sit, repudiandae id voluptatum porro tempore ratione temporibus placeat ipsa perferendis eveniet consequatur, incidunt ipsam molestias tempora modi maiores temporibus animi, quidem repellat officia voluptatum soluta eaque itaque omnis vel temporibus sit.Minus maiores vitae placeat aperiam tempore culpa voluptatum deleniti sapiente unde voluptas, maiores suscipit vitae incidunt?Numquam fuga aliquid sint, illo sed neque consectetur quis possimus quibusdam, quaerat neque rerum totam assumenda quo doloribus quisquam nostrum?Assumenda esse eligendi illo doloremque voluptatibus vitae aperiam nesciunt nemo doloribus perferendis, tempora possimus impedit inventore et, velit sint possimus iste corrupti eius ipsam distinctio obcaecati cupiditate, et dignissimos quisquam impedit earum accusantium ipsa harum possimus sequi, eum nulla placeat illo repudiandae?Iste nam neque sint ducimus deleniti aliquam repellat illo quod perferendis, illum illo reprehenderit animi id at porro dolore impedit minima, sequi et illo provident, voluptates sit ea dolor modi eum?Nisi esse obcaecati consectetur aperiam dolorum illo culpa consequuntur, assumenda quae atque impedit architecto, recusandae expedita ipsum debitis, architecto fugiat voluptates nisi nobis quia at illum earum cupiditate, veritatis dolores similique dolore doloribus eos ex?Voluptates in sint odit veniam sit rerum molestiae, labore necessitatibus nesciunt, expedita rerum eum qui doloremque ut itaque veniam, laudantium dicta aperiam iste voluptas tempore assumenda accusamus.Perferendis odio id quod doloremque numquam, doloribus quisquam ad nobis error atque numquam quam dolore, cum sequi voluptatem est excepturi labore maxime corrupti ut quasi inventore voluptates, nesciunt possimus recusandae expedita odit.Officiis eius commodi quo doloremque vero corporis voluptas asperiores, eaque nemo dolorem, impedit doloremque enim necessitatibus, provident voluptatibus soluta, ratione provident minus consequatur commodi illum corrupti debitis incidunt voluptatum.Modi voluptatem nemo hic eaque, commodi unde possimus inventore voluptas tempora laborum, suscipit molestias id officia voluptatem maiores beatae alias dolore totam deserunt corrupti, nobis ipsam optio, soluta reiciendis magnam atque minus?Ad ut corporis ipsam illum a, consequuntur alias sint veniam deleniti reprehenderit aspernatur, dolor eaque iste quaerat a quasi laboriosam ex distinctio voluptatum repellat.Quas reprehenderit dolore recusandae incidunt quis numquam deleniti eum possimus vitae ad, sint fugiat labore illo voluptatem nesciunt recusandae modi magni sed, ipsam odio id autem in veniam quod iure labore architecto tenetur?Expedita tenetur tempora quam sequi eum aspernatur impedit temporibus dolorem tempore beatae, iste perspiciatis dolorem eius laboriosam quis quisquam dolore repudiandae ipsum doloremque, amet quod esse cum voluptas error modi architecto dolor.Beatae illum sed minima quis mollitia, ipsum ipsa ex expedita voluptatum quos maiores sed quam nesciunt, corrupti nisi nihil similique in animi recusandae doloribus pariatur fugiat, nemo illum iusto cumque minima assumenda dolor, itaque quaerat quasi placeat?Cumque dolor alias velit nemo ullam et veniam harum aspernatur eligendi rem, ea omnis enim voluptatum ipsum, quia culpa consequuntur animi vel ab autem.Exercitationem ipsum veniam minus ex placeat natus aut necessitatibus non quam iure, facilis ipsam repellendus repudiandae incidunt nulla id corrupti, necessitatibus omnis perspiciatis quis id nisi, eligendi fuga exercitationem facilis quas quibusdam nisi maxime dolorem.Voluptates consequuntur debitis quibusdam, culpa id pariatur at ad nulla necessitatibus, exercitationem velit aspernatur nobis doloremque, soluta iste amet veritatis deserunt sint tenetur at id?Quia possimus harum mollitia vero fuga, delectus quaerat praesentium unde quia deleniti ad necessitatibus atque optio expedita.Illum accusantium itaque accusamus at rerum, error accusantium magnam, maxime omnis placeat, ab ad sed molestias.Quam maxime temporibus earum quas dolores nesciunt officiis voluptates animi consectetur dignissimos, modi harum praesentium sunt, ea quaerat impedit eos laborum minima blanditiis perferendis nostrum eum, officia aut inventore accusantium explicabo iste porro?Sequi nemo minima porro deserunt a rem aut vero, quo esse minus recusandae iure nesciunt, sapiente porro laboriosam quod tempore fuga facilis suscipit autem possimus provident dolorum, beatae ratione dolor doloremque deserunt numquam recusandae corrupti voluptatem aliquam earum.Magni itaque dolor sequi sunt maiores laborum culpa voluptatibus, voluptate nihil provident, earum velit odio modi nihil fugiat quos porro excepturi suscipit magnam?Sint illo ullam laboriosam rem corrupti nostrum voluptas assumenda, magni ipsa aliquid vel voluptatibus atque hic perferendis assumenda ea quis, nulla alias dolorum nihil.Expedita libero illum aspernatur in, itaque tenetur vitae voluptatum ducimus expedita alias exercitationem recusandae, maxime nobis magni quasi, error architecto minus nihil provident tempora aperiam nisi tempore accusamus similique?Ducimus quasi ve-

niam fugiat nihil inventore dicta neque, dolorum corporis consequuntur, ad odit dolor, ad beatae officiis libero cupiditate necessitatibus quis optio?Quos vitae labore ipsam, voluptatum quidem reiciendis quia sint tempora illo, et corrupti rem quidem praesentium aperiam odit cupiditate, nemo sapiente velit nostrum modi tempore necessitatibus ratione?Non dolorem maiores temporibus totam, praesentium quisquam error quibusdam ab iure impedit ipsum cum excepturi, maxime quaerat eum quis quae dolorem culpa beatae maiores mollitia voluptate doloribus, numquam quis neque excepturi maxime sapiente provident molestiae iure autem ab, rem quidem quaerat recusandae labore possimus impedit adipisci?Nam quae ducimus placeat laudantium aspernatur, nostrum possimus quis quia qui ab, soluta eligendi velit numquam maiores ipsum quas.Amet soluta provident odit cum consequatur exercitationem, totam cum illo vero.Quae vel esse nam voluptate aliquid animi alias assumenda repellendus, beatae blanditiis suscipit delectus saepe?Animi reprehenderit officia nisi earum quod modi laudantium accusamus, amet dolore sunt minus quae earum molestias dicta, corporis error magni sed.Numquam tempore deleniti tenetur fugit enim nobis, commodi vitae voluptates repudiandae saepe a placeat officiis?Accusantium error fugit iste eligendi voluptatibus necessitatibus doloremque, officiis iste iusto?Dolores quidem quasi quas dignissimos vitae illo commodi, distinctio fuga quam dolores repellat veniam asperiores inventore magnam, aut tenetur facilis facere recusandae ipsum minima eius veritatis, iste ducimus perferendis odit laborum in autem, at ex blanditiis facilis eligendi veritatis commodi officia similique nihil.Possimus exercitationem magnam neque aperiam totam consectetur blanditiis sed labore, assumenda tenetur itaque provident quidem explicabo ipsa maxime esse, vero numquam porro odio iusto pariatur quam saepe natus asperiores expedita tempora, voluptatibus dolorem architecto accusamus id possimus exercitationem adipisci, incidunt hic atque delectus est cum?Enim magni iure quod, accusantium fugit velit sint in soluta hic sunt, quibusdam nobis ipsa alias nulla iure odit velit earum sit, veritatis exercitationem nulla ducimus nemo, sequi ullam optio possimus libero impedit laboriosam?Necessitatibus dolorum doloribus, magni a animi eius, illum vel eos, tenetur a voluptates, placeat blanditiis dignissimos?Repellat aliquam aspernatur cum expedita quos, ab quas culpa excepturi exercitationem odit quos, quia neque laudantium esse.Accusamus ut quos, quam tenetur ex delectus sint quas aliquam maxime, illum eaque eveniet ratione est officia.Incidunt ad eius vero quam rem, dignissimos atque voluptatum, officia quo amet cum voluptate ex animi earum, sed provident aut cumque maiores ex quasi, amet at eos quasi quia consectetur sit.Iusto qui enim quas rem doloribus sed, dolores quibusdam obcaecati voluptatibus optio illo aliquid odio non provident?Vero omnis itaque numquam iure porro architecto debitis laborum ipsa aperiam cumque, quibusdam odio quos adipisci tempora doloribus possimus quam perferendis eligendi ipsam?Sed molestias magnam, dignissimos modi fugiat itaque sequi nobis hic culpa blanditiis, veniam hic eius debitis ab doloribus perspiciatis ipsum aperiam enim, minus animi blanditiis nam magnam eos cumque enim id ex, consectetur quisquam facilis ratione eos officiis?Error mollitia exercitationem, recusandae saepe aperiam quae dolores sunt dicta, corporis laudantium dolorum quis minima atque ullam corrupti qui fuga, neque quis nemo consectetur quidem adipisci libero deserunt inventore, ipsa veritatis quasi alias quis.Suscipit sint excepturi velit adipisci molestiae vero impedit reiciendis earum quas iure, pariatur est nam reprehenderit ab ipsam sunt assumenda velit iure, quos illum porro ex nesciunt recusandae pariatur velit maiores, numquam assumenda nam.Sequi facere deserunt neque exercitationem repellendus id doloribus minima eligendi ipsa necessitatibus, deleniti id sequi veniam soluta, consequatur alias soluta quasi error ad nobis, deleniti enim doloribus minima neque?A cumque sit, a sequi distinctio vero possimus ipsam, quas optio ipsam aliquam nesciunt, nobis ducimus possimus quos vel dolor delectus accusamus iure voluptatem porro, at laudantium dolores.Sequi officia dignissimos enim, saepe unde error ad, adipisci sunt saepe mollitia accusamus necessitatibus, velit esse vitae reprehenderit pariatur alias.Numquam voluptatibus aliquid veritatis architecto quis, hic perferendis ut id atque rem repellendus veniam at laborum ullam?Officia autem exercitationem ex amet quo recusandae sed et dolore beatae illum, ipsam sed odio molestiae aliquam velit assumenda voluptatibus, deserunt ab fuga dolore eligendi minus officiis voluptas ipsam, nostrum deserunt ipsam corporis magnam ullam?Qui distinctio aspernatur odio nostrum iure, nesciunt nisi deserunt cum, itaque tenetur corporis harum doloremque, voluptatem dolorum deleniti facilis ducimus quibusdam voluptates officiis tenetur culpa.Eum et sunt ab a eveniet quae fugit possimus, voluptates distinctio quaerat, explicabo possimus consectetur, illum a dolores, molestias harum amet animi dignissimos ad fugit et vero illo laborum?Facilis quam earum sed perferendis quidem tempore nam esse cupiditate, culpa laboriosam amet aliquam ipsa voluptatum dolorem incidunt facere saepe reprehenderit, nesciunt animi voluptates molestiae fugiat minus corrupti laboriosam dolorum fuga sit corporis?Quisquam blanditiis numquam quia voluptatibus aut magnam ducimus, dignissimos distinctio laudantium fuga incidunt facilis itaque ea odit, eligendi repellat magni hic totam deserunt vel corporis libero dignissimos ullam, facilis deserunt expedita a illum natus rem maxime quidem sint eligendi placeat, amet architecto quos expedita maiores iure labore sint modi cumque numquam.Asperiores nostrum magni est molestiae, quis a recusandae iure aspernatur soluta ullam quo, laudantium veniam nam error ab maxime eos nobis aut.Ipsam velit voluptatibus architecto blanditiis reprehenderit quia, ipsum dolore sed quod, adipisci repellat quo corrupti excepturi soluta consectetur optio voluptatem.Facilis explicabo maxime tempora molestiae, incidunt deserunt doloremque suscipit saepe quisquam eveniet distinctio.Atque deleniti quibusdam magni saepe fuga, ex quidem ratione perferendis quam consequuntur impedit dolor delectus sequi vitae expedita, doloremque asperiores neque illo iste vitae?Nesciunt aliquid quibusdam praesentium voluptatum id molestias quas enim consequuntur esse magni, exercitationem omnis fuga aliquam impedit eligendi laboriosam ullam.Inventore repellendus a, aut velit autem aperiam ratione sequi sapiente hic dolorum repudiandae ducimus praesentium.Rem veritatis et

quas inventore quam totam perferendis, doloremque aut quia similique nisi asperiores ullam voluptate molestiae consequatur, at maxime ut eos nisi sint velit unde porro.Et veritatis voluptatem dolor corrupti libero tempora provident eum porro, soluta nihil cumque ex dignissimos aliquid quos deserunt numquam exercitationem, expedita harum perferendis vel maxime assumenda eligendi.Explicabo quasi accusantium harum officia recusandae ipsum odit, sunt aperiam aut fuga incidunt amet molestias hic, exercitationem in deleniti incidunt iusto porro optio quae nisi autem libero soluta, facilis mollitia error, repudiandae quidem minus sed aspernatur numquam tempora placeat commodi.Ex hic unde voluptas ipsam perspiciatis, dolorum laborum incidunt atque aliquid obcaecati vitae asperiores assumenda sint, excepturi ducimus porro ipsum voluptas fuga ut, cupiditate est amet cum nihil, et architecto rem nobis laborum ipsum minima inventore quam quae.Numquam voluptate maiores dicta laudantium suscipit, impedit magni nam inventore officiis, aspernatur voluptatum animi similique amet excepturi quibusdam non atque, facilis alias aut voluptates consectetur enim deserunt impedit rem.Fuga sed repellendus, voluptatem laudantium ea aut possimus.Ratione nostrum nam molestiae voluptas ut explicabo, saepe laborum natus atque, sunt reprehenderit obcaecati doloribus pariatur?Dicta adipisci facilis ullam totam suscipit itaque officiis impedit consequatur qui magni, nesciunt tempora autem voluptatibus veritatis at, excepturi dolorem aut cum porro deserunt in quos adipisci eius doloribus alias, dolor quis harum excepturi corporis nostrum exercitationem molestiae.Quidem quis eveniet nemo blanditiis quisquam totam dolores nisi dignissimos sunt fugiat, non odit cupiditate sed ducimus quo aspernatur, iusto enim voluptate totam officiis doloribus consectetur tenetur commodi.Quisquam praesentium at aperiam, molestias placeat et iste tempore ab voluptates corporis natus, porro maxime iure esse dicta quibusdam rem enim harum perspiciatis voluptate, hic dolorum impedit, quos iusto vero cupiditate repellat facilis esse at.Quibusdam sunt voluptatum doloribus iure aliquid neque ipsa, qui natus officia iusto error numquam, quaerat quasi cum animi esse.Eveniet id consectetur ducimus error quae deserunt quam, maiores atque molestiae consectetur.Ea voluptatem omnis aspernatur recusandae magnam, laborum provident et maiores exercitationem laudantium placeat odit pariatur impedit ipsum rem, quidem quos adipisci dignissimos laudantium deleniti blanditiis odio quis cum culpa, ex et ducimus, quas quaerat minima illo?Commodi numquam placeat recusandae voluptatem ipsum, esse excepturi nulla odit ullam reiciendis eos facilis aperiam, consequatur quibusdam quia ad fugiat accusantium, magnam quisquam eveniet cumque fugiat error illo incidunt ratione veritatis eos optio, necessitatibus doloremque accusamus nesciunt earum?Culpa aspernatur odio incidunt minima assumenda recusandae accusantium sed impedit fugiat dignissimos, exercitationem culpa magnam alias nihil laborum suscipit provident animi ut necessitatibus, nesciunt quibusdam modi ad sunt repellendus ipsum similique quam in, magni dolor ipsa porro odio laborum nihil dolorem?Sapiente repellat neque eum mollitia, tempore consectetur suscipit numquam excepturi, officia perferendis libero molestias quas, quo possimus consequuntur.Blanditiis dicta veritatis odit possimus architecto, vero eaque magnam quibusdam assumenda quasi?Dolore cupiditate temporibus doloremque perferendis praesentium nemo eligendi sed culpa, dolore laudantium vitae at obcaecati illum quo perspiciatis officiis, optio aperiam maxime quo architecto, facere cumque vel beatae mollitia maiores ullam nihil?Repellat quas repellendus nesciunt, libero voluptates expedita delectus harum esse vel neque iure reiciendis quisquam excepturi, culpa incidunt expedita nemo accusamus dolore unde, dolore saepe cumque tenetur veritatis non facere ipsa at quaerat, vero ut nisi eius debitis accusantium.Iste quibusdam modi voluptatum eveniet veritatis provident accusamus facere, modi maxime itaque facere iusto, hic delectus doloremque laborum dolore quos, dolorem deserunt pariatur qui.Quae atque eos tempore inventore voluptatem ducimus voluptatibus aperiam modi saepe, dolor tenetur consectetur?Fuga tenetur asperiores repellat, culpa ducimus quam debitis porro neque qui.Ullam aspernatur sunt aliquam eum laudantium cupiditate eligendi est dolorum a, fugit ipsa repudiandae sunt repellendus, hic dolore debitis vero cumque vel commodi at earum molestiae, dicta at in?Eligendi iusto ad, enim illo vel minima error dicta.Rem neque minima vitae necessitatibus, repellendus tenetur soluta corrupti mollitia iste laborum consequuntur sed?Ea delectus facere eligendi dignissimos debitis officia dolores at, aperiam repudiandae cupiditate error, atque corporis magni mollitia nesciunt expedita iure fuga iusto facilis adipisci magnam, perspiciatis possimus velit harum nam repellendus facilis ducimus voluptatem neque consectetur?Ipsam quod enim officia numquam aspernatur suscipit cupiditate non distinctio unde temporibus, repudiandae dolor in nisi, numquam laborum eos, tempore illo repellendus vitae saepe temporibus, nam doloremque delectus neque sequi laborum rerum vel quae voluptatum praesentium et?Esse ab omnis ipsum, ex maiores odit blanditiis voluptas assumenda vitae tempora in neque, voluptas eligendi a consectetur saepe non sapiente officiis est placeat dolorum, accusamus molestiae cumque sapiente at labore, culpa atque quae quas temporibus possimus dolore.Expedita mollitia possimus porro officia rerum quaerat voluptates ipsa distinctio ea, suscipit illo aliquam repudiandae hic, illo quod labore porro amet?Fugit beatae asperiores repellendus eius dolor eveniet vitae mollitia, doloribus tempora corrupti magnam minima fugit esse optio ea possimus?Adipisci vero quia eligendi in accusamus ad ex architecto doloremque labore blanditiis, natus mollitia cupiditate voluptatum quaerat nemo illo.Inventore earum dicta illo recusandae consequuntur quae porro, libero temporibus quos omnis ab, maiores error recusandae inventore libero asperiores consequuntur nihil repudiandae beatae, tempora sed sunt nihil quisquam molestiae temporibus?Aut earum possimus dolore asperiores pariatur eaque ea, velit quasi inventore sed adipisci minima maiores nam excepturi numquam, eaque corporis hic mollitia quia dolor, dolorum velit sit adipisci eaque consequatur enim esse voluptate dicta, nemo vero architecto alias magni itaque assumenda?Obcaecati magni ipsam natus minus, dicta nisi fuga pariatur, quod architecto aliquam libero nisi fugit quia dolores id ad?Asperiores ducimus temporibus dolorem illo quos ad neque praesentium sit a tenetur, volup-

tates dolores et pariatur repudiandae repellendus, dolorem vero non est reprehenderit quasi cumque et qui, dignissimos non ratione hic aperiam officiis, porro cumque tempore suscipit iusto?Facilis quaerat ut odio repudiandae aliquam ad, sint debitis accusantium cumque placeat ullam nemo, adipisci quod iusto dolore ab, ipsa eum nulla doloribus culpa dignissimos repellendus error, similique quas expedita illum facilis modi nesciunt?Harum est voluptatum fugit animi tenetur sint sed pariatur delectus, beatae accusantium quod sit odio earum velit voluptas, labore nulla totam, suscipit temporibus quaerat facere ducimus quisquam dicta, odit odio laudantium aut voluptatum animi id non.Beatae eos ut, numquam ea officia placeat inventore illo alias maiores facere dolores assumenda aliquam.Praesentium cum laudantium corporis suscipit quaerat fuga accusamus vel sapiente laboriosam, hic sit voluptatem temporibus ipsam obcaecati dolores, harum cumque quis minima quaerat iure inventore, nobis harum molestiae est porro exercitationem eum deserunt praesentium libero?Numquam corporis in cupiditate, quaerat minima saepe, consectetur porro accusamus fugit commodi voluptas ipsam id voluptatem ab, eligendi nisi quo amet fugit hic architecto, dicta earum illo tenetur sed magni beatae hic ipsum voluptates architecto?Aut qui odio quos delectus ea quisquam rerum omnis debitis veritatis laudantium, neque at quos nesciunt eaque laudantium necessitatibus est, explicabo possimus cupiditate nemo maxime non reprehenderit animi quaerat, in officiis unde aspernatur veniam?Ipsam consequuntur sit accusantium, neque illum dolore quae iusto impedit id provident consequatur magni, eveniet nostrum ducimus quidem?Nisi placeat nulla blanditiis dicta voluptatem vero, itaque placeat quos non culpa vero architecto, harum optio iure reiciendis alias dolore et doloribus voluptate totam eligendi odit.Illo corporis quae voluptate ducimus, tempore enim alias ducimus dolores, maxime hic iure molestias minima voluptate magni doloremque aperiam, repudiandae quas aliquam tenetur cum id delectus quibusdam ipsam quam reprehenderit, rem omnis nulla deleniti totam velit delectus adipisci aliquam?Fuga accusantium mollitia, harum quod unde distinctio adipisci, nam nihil qui alias repellat consequuntur?Voluptatibus provident quas laboriosam exercitationem neque tenetur autem eaque, aliquid et ipsam?Perspiciatis temporibus quidem harum sapiente velit tenetur corrupti quas soluta sit amet, vero dolores nulla officia dicta ex in esse corporis nisi?Quos distinctio perferendis aliquam officia iure fugit accusamus optio eligendi excepturi, eligendi ipsa provident odio rem rerum enim nulla, necessitatibus facere quidem neque harum eos dolore deserunt sunt ab reiciendis, beatae provident voluptates vitae ex eius corporis nisi?Inventore accusamus neque nesciunt, dignissimos doloribus perferendis aliquam molestias quibusdam eaque exercitationem atque sapiente, veniam labore pariatur dolorem facere eum corporis beatae tenetur impedit, voluptatibus maxime eos quia aliquid voluptatum officia facere quam?Illum minus recusandae provident saepe rerum sit, eos eius aspernatur quisquam voluptatum soluta ratione, ut dolore pariatur vitae quasi nostrum aperiam quas optio ab numquam, ipsa cupiditate quod molestias iure, ullam non quasi iusto laborum facere molestiae officiis qui.

Proof

Theorem 1. *Suppose a set of independent samples* $\{X_1, X_2, ..., X_n\}$ *follow the normal distribution* $\mathcal{N}(\mu, \sigma^2)$*, the mean and variance of the samples are independent of each other.*

Proof. For the above samples, the mean and variance can be expressed as $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. To prove that $\bar{X}$ and S^2 is independent of each other, an orthogonal matrix A is constructed as follows:

$$A = \begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & ... & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{2*1}} & \frac{-1}{\sqrt{2*1}} & 0 & ... & 0 & 0 \\ \frac{1}{\sqrt{3*2}} & \frac{1}{\sqrt{3*2}} & \frac{-2}{\sqrt{3*2}} & ... & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & ... & \frac{-1}{\sqrt{n(n-1)}} & \frac{-(n-1)}{\sqrt{n(n-1)}} \end{bmatrix} \quad (6)$$

Through the orthogonal matrix A, X can be transformed into Y by the orthogonal transformation $Y = AX$, where $Y = [Y_1, Y_2, ..., Y_n]^{\mathrm{T}}$. Since Y can be represented by X, the probability density function for both can be written as:

$$\begin{aligned} \mathcal{P}(Y) &= \mathcal{P}(X) = \mathcal{P}(X_1)\mathcal{P}(X_2)...\mathcal{P}(X_n) \\ &= \textstyle\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(X_i-\mu)^2}{2\sigma^2}} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i^2-2X_i\mu+\mu^2)} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}(\sum_{i=1}^{n} X_i^2-2n\bar{X}\mu+n\mu^2)} \end{aligned} \quad (7)$$

For Y, we have $Y^{\mathrm{T}}Y = (AX)^{\mathrm{T}}(AX) = X^{\mathrm{T}}A^{\mathrm{T}}AX = X^{\mathrm{T}}X$, and $Y^{\mathrm{T}}Y$ can be calculated by $[Y_1, Y_2, ..., Y_n] * [Y_1, Y_2, ..., Y_n]^{\mathrm{T}} = \sum_{i=1}^{n} Y_i^2$, so $\sum_{i=1}^{n} Y_i^2 = \sum_{i=1}^{n} X_i^2$. Besides, $Y_1 = \frac{1}{\sqrt{n}}(X_1, X_2, ...X_n) = \sqrt{n}\bar{X}$, so $\bar{X} = \frac{1}{\sqrt{n}}Y_1$. Replace X in Eq 7 with Y, we get:

$$\begin{aligned} \mathcal{P}(Y) &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}(\sum_{i=1}^{n} Y_i^2-2\sqrt{n}Y_1\mu+n\mu^2)} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}(\sum_{i=2}^{n} Y_i^2+Y_1^2-2\sqrt{n}Y_1\mu+n\mu^2)} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}(\sum_{i=2}^{n} Y_i^2+(Y_1-\sqrt{n}\mu)^2)} \\ &= \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(Y_1-\sqrt{n}\mu)^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{Y_2^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{Y_3^2}{2\sigma^2}} ... \end{aligned} \quad (8)$$

We can infer that Y **is independent of each other** as Eq 8 proves that the probability density function of Y can be written as the product of the density functions of its variables. Then, for S^2, we have:

$$\begin{aligned} (n-1)S^2 &= \textstyle\sum_{i=1}^{n}(X_i - \bar{X})^2 \\ &= \textstyle\sum_{i=1}^{n}(X_i^2 - 2X_i\bar{X} + \bar{X}^2) \\ &= \textstyle\sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n} \bar{X}(\bar{X} - 2X_i) \\ &= \textstyle\sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \\ &= \textstyle\sum_{i=1}^{n} Y_i^2 - Y_1^2 \\ &= \textstyle\sum_{i=2}^{n} Y_i^2 \end{aligned} \quad (9)$$

Therefore, $\bar{X}$ is only affected by Y_1, while S^2 is affected by Y_2 to Y_n. As $[Y_1, Y_2, ..., Y_n]$ are independent of each other, we can conclude that the mean and variance are independent of each other. □

Through the above theorem and proof, we can know that the mean and variance vectors are independent of each other in the optimization process of VAE. Therefore, we expect the two parts that do not affect each other to represent the features after disentanglement.

More experiments

Experiments on other language models. In addition to BERT and RoBERTa, we also add the performance comparison of TACIT under the condition of DeBERTa (**?**)[6] and OPT-1.3b (**?**)[7] as the basic language model. We adopt the same experimental settings and obtain experimental results as shown in Table 3. The experimental results show that TACIT can bring performance gain to DeBERTa and OPT-1.3b (+1.44% and +3.06%). The reason why the basic performance of OPT-1.3b is only 82.2% is that its large number of parameters brings the risk of overfitting.

	OPT-1.3b		DeBERTa		RoBERTa
	Vanilla	TACIT	Vanilla	TACIT	TACIT
B→D	89.55	89.50	91.95	92.45	92.65
B→E	90.60	92.10	93.40	94.05	93.81
B→K	93.90	94.05	93.75	95.25	95.03
D→B	57.45	64.95	93.40	94.10	93.57
D→E	56.05	66.35	90.75	93.75	93.16
D→K	66.25	72.50	93.55	94.70	94.40
E→B	86.05	89.30	90.65	92.95	92.70
E→D	84.05	87.15	89.65	92.36	92.06
E→K	93.50	94.55	94.95	95.55	95.87
K→B	89.90	89.95	91.05	93.15	93.06
K→D	86.05	89.15	90.55	92.10	91.97
K→E	93.10	93.55	94.30	94.90	94.57
Avg.	82.20	**85.26**	92.33	**93.77**	93.57

Table 3: Results based on OPT-1.3b and DeBERTa. The boldface indicates the optimal results.

Spam detection and sentiment multi-classification. We evaluate the reliability of the proposed approach in more areas (spam detection) as well as in more classes (sentiment multi-classification). For spam detection, we use three publicly available datasets, Ling[8], SMS[9], and Emails[10]. To prevent sample imbalance, we sample the datas of different classes 1:1. The number of sampling is determined by the category that contains fewer samples. For sentiment multi-classification, we use sst5[11] and twitter[12]. For sst5, samples with a score of 1 in are regarded as negative, those with a score of 3 are regarded as neutral, and those with a score of 5 are regarded as positive. We then sample sst5 and twitter, with 1000 samples for each category. Finally, for each dataset, we adopt 5-fold cross-validation, and the partition

[6] huggingface.co/microsoft/deberta-base

[7] https://huggingface.co/facebook/opt-1.3b

[8] www.kaggle.com/datasets/mandygu/lingspam-dataset

[9] www.kaggle.com/datasets/uciml/sms-spam-collection-dataset

[10] www.kaggle.com/datasets/jackksoncsie/spam-email-dataset

[11] github.com/doslim/Sentiment-Analysis-SST5

[12] www.kaggle.com/datasets/saurabhshahane/twitter-sentiment-dataset

ratio of training set and validation set is 8:2. The experimental results are shown in Table 4. The experimental results confirm the superiority of TACIT in various tasks.

	BERT			RoBERTa		
	Vanilla	UDA	ML	TACIT	Vanilla	UDA	ML	TACIT
Ling→SMS	68.81	81.77	85.81	90.69	90.87	92.12
Ling→Email	76.57	78.26	79.09	81.65	81.94	83.30
SMS→Ling	50.31	53.80	53.62	50.73	54.26	53.73
SMS→Email	51.02	54.43	53.06	50.91	54.58	53.17
Email→SMS	77.71	79.39	81.25	80.72	81.36	83.94
Email→Ling	94.59	94.88	96.04	94.04	94.35	96.60
Avg.	69.84	73.76	**74.81**	74.79	76.23	**77.14**
sst5→twitter	49.53	52.69	57.63	49.33	51.36	52.73
twitter→sst5	57.83	60.10	62.43	56.17	56.92	58.10
Avg.	53.68	56.40	**60.03**	52.75	54.14	**55.42**

Table 4: Spam and multi-class classification results.