Unsupervised Reinforcement Adaptation for Class-Imbalanced Text Classification

Anonymous ACL submission

Abstract

Unsupervised domain adaptation (UDA) augment model performance with only accessible annotations from the source domain and unlabeled data from the target domain. Existing state-of-the-art UDA models learn domain-invariant representations across domains and evaluate primarily on class-imbalanced data. In this work, we propose an unsupervised domain adaptation approach via reinforcement learning that jointly leverages both label prediction, domain, and imbalanced labels across domains. We experiment with the text classification task for its easily accessible datasets and compare the proposed method with five baselines. Experiments on three datasets prove that our proposed method can effectively learn robust domain-invariant representations and successfully adapt text classifiers over domains and imbalanced classes.

1 Introduction

Performance of text classifiers may get worse when training and test data come from different domains, where feature distributions are the difference between the source (training data) and target (test data) domains. However, annotated data from the target domain may not be available to train text classifiers causing the main challenge of adapting classifiers from the source domain to the target domain. This happens when obtaining annotations from the target domain is either expensive or requires domain expertise, such as classifying moral values in social psychology (Hoover et al., 2020). *Unsupervised domain adaptation* (UDA) is an essential approach to augment model performance on the target domain with only labeled data from the source domain and no access to any labeled data from the target domain.

The key idea of UDA is to find a shared feature space that is predictive across target and source domains (Ramponi and Plank, 2020). The shared space, *domain-independent* feature set, allows transferring of trained text classifiers from the source domain to the target domain. Methods to find the space have two major directions, pivot feature (Blitzer et al., 2006; Daumé III, 2007; Ziser and Reichart, 2018; Ben-David et al., 2020) and adversarial learning (Ganin and Lempitsky, 2015; Chen et al., 2020b; Du et al., 2020). The pivot-based method selects a subset of shared features, called pivots, which learn important cross-domain information to represent shared feature space. Adversarial learning approaches the shared feature space by reducing document features' capability to distinguish source and target domains. The common method to achieve this is Gradient Reversal Layer (GRL) (Ganin and Lempitsky, 2015) aiming to reduce domain-specific patterns. However, learning optimal predictive features across the source and target domains is still a challenge. Additionally, a wide evaluation benchmark of UDA for text classifiers is the Amazon review (Blitzer et al., 2006). The data has the same balanced-class distributions for both source and target domains. However, the class-imbalanced data (e.g., different label distributions across domains (Cui et al., 2017; Cheng et al., 2020)) is a different challenge. Under the class-imbalanced scenario, we assume that the label is imbalanced across domains, and the label distributions in source and target domains are not the same. For example, the source domain of Amazon Book reviews may have more positive reviews than negative reviews, and the target domain of Kitchen may have more negative reviews. However, evaluating unsupervised domain adaptation under the class-imbalanced scenario is under-examined than the ideal scenario of the class-balanced benchmark.

In this study, we proposed an unsupervised reinforcement adaptation model (URAM) for text classifiers under the UDA setting that only labeled source data and unlabeled target data are available. Specifically, we propose a neural mask mechanism to generate domain-dependent and -independent feature representations and a reward policy using a critic value network (Konda and Tsitsiklis, 2000) (CRN) to learn optimal domain-independent representations. The reward policy optimizes the URAM via three joint reward factors, label, domain, and domain distance. While the label reward aims to encourage text classification models on domain-independent features to predict correct document classes, the domain and domain distance rewards reduce domain variations of domain-dependent feature representations between source and target domains. We compare our reinforcement adaptation model with five baselines and experiment on four class-imbalanced data with both binary and non-binary labels. The results using the F1-score demonstrate the effectiveness of our reinforcement learning model that outperforms the baselines by 3.13 on average. The main contributions of this paper are as follows:

- We propose a reinforcement learning model for unsupervised domain adaptation that jointly leverages cross-domain variations and classification performance.
- We experiment unsupervised domain adaptation approaches on the class-imbalanced scenario that label distributions are different between source and target domains. The class-imbalanced scenario is under-explored for the UDA challenge.
- We conduct an extensive ablation analysis that demonstrates how the reinforcement model can coherently combines both pivot and adversarial directions of unsupervised domain adaptation.

2 Background

This section briefly recaps the concepts of unsupervised domain adaptation (UDA) and reinforcement learning.

2.1 UDA for Class-Imbalanced Data

UDA assumes a labeled dataset with $\mathcal{D}_S = \left\{\left(x_s^i, y_s^i\right)\right\}_{i=1}^{n_s}$ from source domain and a unlabeled data $\mathcal{D}_T = \left\{x_t^j\right\}_{j=1}^{n_t}$ from target domain, data distributions of the two domains are different, $p(x_s) \neq p(x_t)$, and the two domains share the same number of *unique* annotations. UDA is to find a common feature space aligning source and target domains so that $f(p(x_s)) \approx p(x_t)$ However, class-imbalanced data naturally exist in UDA tasks that may cause inefficient knowledge transfer (Ramponi and Plank, 2020). We assume both data and labels are not equally distributed in this work.

2.2 Reinforcement Learning

Actor-Critic (Konda and Tsitsiklis, 2000) is an RL algorithm that combines Actor and Critic networks. Critic, a value network (denote as V_{θ_c}), estimates rewards at state s_t and is optimized by state difference error as follows

$$\mathcal{L}(\theta_c) = \left\|V_{\theta_c}\left(s_t\right) - r\left(\mathrm{s}_t, a_t\right) - V_{\theta_c}\left(\mathrm{s}_{t+1}\right)\right\|^2 \quad (1)$$

where $r(s_t, a_t)$ is a target reward and tells us the reward for taking action a in state s. The actor is a policy function that gives us the probability of taking action a in the state s. The actor decides which action should be taken, and the critic evaluates how good the action is and how it should adjust. The learning of the actor (θ_a) is based on policy gradient approach as the following

$$\mathcal{L}^A(\theta_a) = \sum_t \log \pi_{\theta_a}\left(a_t, s_t\right) A\left(s_t, a_t\right) \quad (2)$$

, where $A\left(s_t, a_t\right) = r\left(s_t, a_t\right) + \gamma V\left(s_{t+1}\right) - V\left(s_t\right)$. γ is a decay factor that discounts rewards backward over steps. To encourage the actor to explore more actions, the algorithm adds an entropy penalty,

$$\mathcal{L}^S(\theta_a) = -\sum_a \pi_\theta(a \mid s) \log \pi_{\theta_a}(a \mid s) \quad (3)$$

The overall objective is as following,

$$\mathcal{L} = \mathcal{L}(\theta_c) - (\mathcal{L}^A(\theta_a) + \mathcal{L}^s(\theta_a)) \quad (4)$$

3 Unsupervised Reinforcement Adaptation Model (URAM)

In this section, we present details of the URAM, shown in Figure 1. The URAM trains classifiers on the labeled data from the source domain and unlabeled data from the target domain. The model contains three major modules: 1) a base model; 2) adversarial learning; 3) reinforcement learning.

3.1 Based Model

Our based model consists of an encoder and a classifier. The encoder extracts features from input documents, and the classifier predicts document labels. The based model takes a regular in-domain

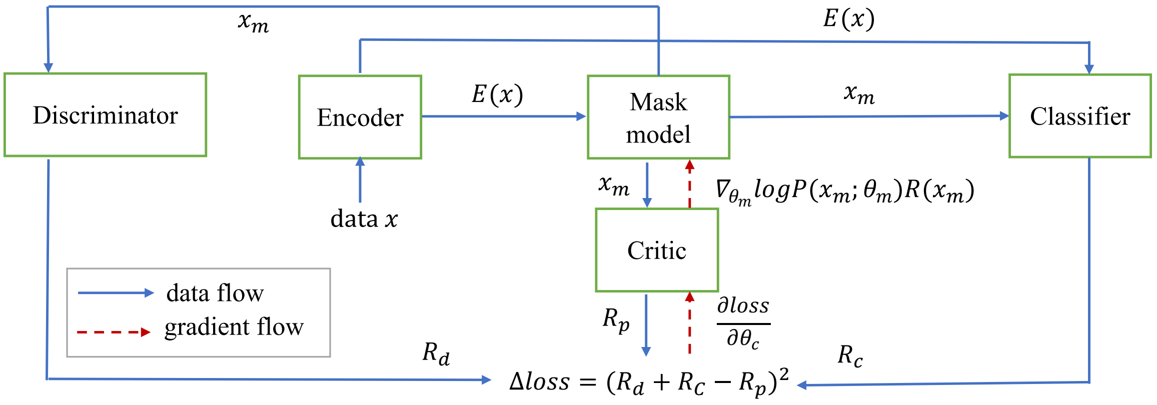
x_m
$E(x)$
Discriminator
Encoder
$E(x)$
Mask model
x_m
Classifier
data x
x_m
$\nabla_{\theta_m} logP(x_m;\theta_m)R(x_m)$
Critic
data flow
gradient flow
R_p
$\frac{\partial loss}{\partial \theta_c}$
R_d
R_c
$\Delta loss = (R_d + R_c - R_p)^2$

Figure 1: Illustration of proposed URAM.

training method with n_s labeled samples from the source domain

$$\min_{\theta_e,\theta_{cla}} \sum_{i}^{n_s} (\mathcal{L}(C(E(x_s^i,\theta_e),\theta_{cla}),y_s^i) \qquad (5)$$

, where θ_e, θ_{cla} are the parameters of the encoder and classifier respectively. $\mathcal{L}(\cdot)$ is the cross-entropy loss.

3.2 Adversarial Learning Strategy

In this part, we propose an adversarial learning strategy to train our discriminator and the mask model. Source and target domains have a feature distribution discrepancy, and therefore representations ($E(x)$) from the encoder will capture domain-specific patterns. With these domain-specific features, a discriminator (θ_d) can distinguish a sample's domain by minimizing the classification error

$$\min_{\theta_d}(\mathcal{L}(D(E(X_s),\theta_d),\mathbb{1})+\mathcal{L}(D(E(X_t),\theta_d),\mathbb{0})). \qquad (6)$$

Next, we propose a mask model to confuse the discriminator. The mask model is based on the idea that the features extracted by the encoder contain domain-specific features and domain-independent features. The domain-independent features are transferable knowledge cross domains. Learning the domain-independent features can improve the generability of text classifiers, while overfitting domain-specific features degrade the cross-domain performance. Based on this idea, the mask model captures common knowledge cross domains and generates domain-independent representations. We design a training objective for the mask model. Intuitively, if the discriminator uses the generated features from the mask model and fails to recognize domains of input data, then this indicates the features generated by the mask model are domain-independent. Therefore, our first goal is to maximize the loss of the discriminator as the following formulation:

$$R_d = \max_{\theta_m}(\mathcal{L}(D(M(E(X_s),\theta_m),\theta_d),\mathbb{1})+ \\ \mathcal{L}(D(M(E(X_t),\theta_m),\theta_d),\mathbb{0})) \qquad (7)$$

Our second goal is to minimize the classification loss using the domain-independent features (x_m) on the classifier (C). The mask model yields the domain-independent features by learning how to mask out domain-dependent features. To reduce the relevance of the features to the classification task, we want to obtain the masked features that can keep an identical prediction from the original features while removing the domain-dependent patterns. Therefore, our second goal is to make a consistent prediction between $C(M(E(X)))$ and $C(M(E(X)))$. Here, we follow the work (Saito et al., 2018b) and employ L1-distance to measure the representation discrepancy loss between $C(M(E(X)))$ and $C(M(E(X)))$ as the following:

$$R_c = min(\mathcal{L}_{dis}(|C(M(E(X))) - C(E(X))|)) \qquad (8)$$

, where R_c measures cross-domain variations.

3.3 Actor-Critic Learning

We apply actor-critic algorithm (Konda and Tsitsiklis, 2000) to optimize our model since searching domain-independent features is a non-differentiable process. The optimization process of the mask model is shown in Fig 1.

First, we introduce a value estimation network, critic. The critic helps to estimate an action's reward by giving a state. Our critic is a 2-layer feed-forward network, with the input of $M(E(X))$ and

$E(X)$. The predictive reward R_p is formulated as follow:

$$R_p = f_{w_2}(f_{w_1}(E(X)) - f_{w_1}(M(E(X)))) \tag{9}$$

, where $E(x)$ and $M(E(x))$ represent the state s and s_{t+1} respectively. The loss function is as the following,

$$\mathcal{L}(\theta_c) = \left(R_d + R_c - R_p\right)^2 \tag{10}$$

The critic is trained with Adam on a mean squared error $\mathcal{L}(\theta_c)$.

The mask model generates a mask matrix $\mathcal{M}_a$ and is an actor model by a fully connected neural network and a sigmoid unit. It accepts inputs from the encoder and calculates a masked probability of each features $\mathcal{M}_p$. Then we adopt Bernoulli sampling and obtain a logical matrix $\mathcal{M}_a$. The elements in $\mathcal{M}_a$ belongs to $\{0, 1\}$. We denote the output of the mask model as $x_m = \mathcal{M}_a * E(x)$. The mask model's training objective is to maximum the total reward R_d and R_c defined in e.q. 7 and e.q. 8

$$J(\mathcal{M}_a \mid E(X)) = \mathbb{E}_{\mathcal{M}_a \sim \pi(\mathcal{M}_p | E(X))}\{R_d - R_c + R_{reg}\}, \tag{11}$$

, where π is a policy function and R_{reg} is a regularization term, controlling the number of masked features. We set $R_{reg} = (\sum \mathcal{M}_a)$. The optimization of e.q. 11 follows e.q. 2 and e.q. 3. Since we only take one action, the optimization in e.q. 2 and e.q. 3 can be simplified as the following

$$\mathcal{L}(\theta_m) = -\log \pi_{\theta_m}(a, s)\, A(s, a) + \pi_{\theta_m}(a \mid s) \log \pi_{\theta_m}(a \mid s) \tag{12}$$

, where $A(s, a) = R_d + R_c - R_p$. We update θ_m by maximizing $\mathcal{L}(\theta_m)$.

Algorithm 1 Optimization Process of Our Model.

```
Input: The source data $D_s = (X_s, Y_s)$ and target data $D_t = (X_t)$, maximum iteration $I$;
Output: The network parameter $\theta_e, \theta_{cla}, \theta_d, \theta_m, \theta_c$;
for $i = 1; i < I; i++$ do
    Samples a batch from $D_s$ and $D_t$;
    Update $\theta_e, \theta_{cla}$ via e.q.(5);
    Update $\theta_d$ via e.q.(6)
    Update $\theta_m$, $\theta_c$ via section (3.3)
end for
return $\theta_e, \theta_{cla}, \theta_d, \theta_m, \theta_c$;
```

3.4 Training Procedure

Our training procedure includes three steps: 1) **step A** trains the encoder and classifier as e.q. 5; 2) **step B** trains the discriminator by e.q. 6; 3) **step C** training the mask model by the reinforcement learning. We summarize the optimization process in Algorithm 1.

4 Experiment

4.1 Datasets

We assembled four datasets, three online reviews and one Twitter data. The reviews are binary labels, and the Twitter data has 11 unique labels. We summarize data statistics in Table 2.

Amazon Review (Ni et al., 2019) consists of four different product genres: Books (B), DVDs (D), Electronics (E) and Kitchen (K). We treat each genre as a domain, where each domain contains 4,000 samples and two classes (positive and negative). We name cross-domain evaluations by the source-target format. For example, Books-Kictchen means that Books is the source data and Kictchen is the target data. In this task, we randomly select 2000 samples from each domain that follows the standard benchmark (Blitzer et al., 2006) for the UDA evaluations, while label distributions are not the same cross domains. We split 1600 as a training set and 400 as a test set for each domain.

Yelp and **IMDB** are two online review datasets from torchtext.[1] The binary label distributions are balanced. Therefore, to create imbalanced datasets, we first randomly produce a label ratio and then sample data depending on the label ratio. Following the Amazon review, we randomly select 2000 samples from Yelp Review Polarity and IMDb training set, separately. We treat Yelp and IMDB as domains and split the training and test sets into 1600 and 400 samples.

MFTC (Hoover et al., 2020) is a multi-label classification Twitter data with 35,108 tweets. These tweets are drawn from seven different discourse domains with moral sentiment across seven social movements, including MeToo, Black Lives Matter (BLM), Sandy, Davidson, Baltimore, All Lives Matter (ALM), and US Presidential Election (Election). We treat social movements as domains. These domains share the same set of 11 moral senti-

[1] https://pytorch.org/text/stable/datasets.html

Method	MeToo - Davidson	Davidson-MeToo	Book-Kitchen	Kitchen-Book	Yelp-IMDB	IMDB-Yelp
			LSTM			
DANN	45.00	23.17	83.33	93.55	45.16	61.79
MCD	40.25	23.61	83.85	94.17	48.27	61.54
JUMBOT	46.94	23.26	81.79	93.66	42.57	56.78
ALDA	38.20	23.31	84.14	93.88	42.30	52.46
URAM	47.06	24.00	85.09	94.49	50.58	62.50
			BERT			
DANN	78.20	23.50	73.23	69.64	54.36	43.44
MCD	79.51	23.39	74.33	69.54	43.67	42.37
JUMBOT	73.74	23.23	80.57	75.00	53.37	43.08
ALDA	77.26	24.42	77.21	70.54	47.01	39.84
URAM	81.93	27.09	86.24	76.97	57.70	45.16

Table 1: Cross-domain performance of UDA models using F1 score. Each UDA model testifies over two popular neural feature extractor, LSTM. We list extensive evaluations in the Appendix.

	Docs	Tokens	pos/neg
M-MeToo	4480	13.86	-
M-Davidson	4480	19.13	-
A-Book	2000	25.65	0.65
A-Kitchen	2000	29.73	4.78
Yelp	2000	231.57	0.26
IMDB	2000	146.01	0.67

Table 2: Data statistics summary of Morality and three review data, Amazon, Yelp and IMDB. We include multi-label distributions of the Morality data in appendix, Table 7.

ment types: Subversion, Authority, Cheating, Fairness, Harm, Care, Betrayal, Loyalty, Purity, Degradation, Non-moral. The rates of each of the virtues and vices vary substantially across the domain. For example, only approximately 2% of the ALM data were labeled as degradation while approximately 14% of the Sandy data were labeled as degradation. In this task, we randomly split 3584 samples as training and 896 samples for testing for each domain.

We conduct an exploratory analysis of domain variations. The analysis follows the name format as source-target. We use KL-divergence of the class distribution to measure the category-wise distribution and Euclidean distance to measure the domain-wise distribution. The domain-wise discrepancy refers to the euclidean distance of the encoder's output between the training and test sets. The category-wise is the KL-divergence of labels' distribution between the training and test sets. We extract feature vectors using LSTMs trained over the domains. We show cross-domain discrepancy in Table 3. We can find that the multi-label Twitter data has more variations in both domain and label distributions.

4.2 Baselines

We compare our models with four recent methods.

- DANN (Ganin and Lempitsky, 2015) maps source and target domains to a common subspace through shared parameters. This approach introduces a gradient reversal layer to confuse domain prediction to improve classification robustness across domains with the adversarial train.
- MCD (Saito et al., 2018a) proposes to maximize the discrepancy between two classifiers' outputs to detect target samples that are far from the support of the source. Then, A feature generator learns to generate target features near the support to minimize the discrepancy.
- JUMBOT (Fatras et al., 2021) proposes a new formulation of the mini-batch optimal transport strategy coupled with an unbalanced optimal transport program to calculate optimal transport distance.
- ALDA (Chen et al., 2020b) constructs a new loss function by introducing a confusion matrix. The confusion matrix reduces the gap and aligns the feature distributions in an adversarial manner.

4.3 Implementation Details

In this study, we evaluate the UDA methods using two standard neural models as feature extractors, LSTM (Hochreiter and Schmidhuber, 1997)

Table 3: Domain discrepancy summary.

discrepancy	MeToo-Davidson	Davidson-MeToo	Book-Kitchen	Kitchen-Book	Yelp-IMDB	IMDB-Yelp
domain-wise	10889	661	15986	11680	1.523	1.692
category-wise	0.1197	0.1933	2.0×10^{-4}	1.0×10^{-4}	0.044	0.050

and BERT (Devlin et al., 2019). For the LSTM-based encoder, we use pre-trained word vectors GloVe (Pennington et al., 2014) by torchtext [2] to train word embedding. The learning rate is set to 1×10^{-3} and batch size set to 64. We utilize a Bidirectional LSTM as our encoder and set the LSTM hidden number as 256. For the BERT-based encoder, we load the pre-trained BERT model (`bert-base-uncased`) from the transformer toolkit (Wolf et al., 2020). We set the learning rate as 1×10^{-5} and batch size as 16.

In all the above experiments, we used Adam (Kingma and Ba, 2015) to optimize our model and maximum iteration set to 50 in all experiments. We run each experiment five times and average F1 as the final performance.

4.4 Result

Table 4: The domain-wise discrepancy based on domain adaptation methods.

	DANN	MCD	JUMBOT	ALDA	URAM
MeToo - Davidson	3.937	5.806	0.072	7.902	0.401
Davidson-MeToo	0.016	10.862	0.121	0.016	0.044
Book-Kitchen	0.950	1.651	0.046	3.922	0.233
Kitchen-Book	0.649	1.749	0.073	2.984	0.196
Yelp-IMDB	3.376	3.029	0.492	8.106	0.586
IMDB-Yelp	2.951	6.184	0.733	31.469	0.665

In this section, we present model performance on the cross-domain adaptation task and conduct an ablation analysis to examine the effects of the two reward factors, R_d and R_c. We include extensive evaluation results in the appendix (Table B).

Overall Performance. The table 1 reports the overall performance. Our method achieves the best result in the datasets with a significant discrepancy both in domain and category. We obtain a significant improvement on Amazon datasets, Book-Kitchen (1.12%-17.7%) and Kitchen-Book (2.62%-10.68%), respectively. Amazon datasets follow the traditional assumption that different domains have significant feature discrepancies but have similar label distributions. Our improvement on Amazon datasets verifies our model effectiveness of learning transferable knowledge. On the other hand, our method also can release the category discrepancy problem. As shown in the table 1, our method outperforms existing methods remarkably on the MFTC dataset (Metoo-Davidson) with the significant discrepancy in domain and category since we can align the distribution both in-text features and labels. We notice some latest methods fail to compete with DANN. We infer the reasons behind this are that some methods do not consider category discrepancy. For example, the performance of ALDA is lower than DANN on Metoo-Davidson since ALDA tries to align category discrepancy by narrowing domain discrepancy, which causes negative knowledge transfer. The other reason is due to poor robustness. Some methods may ascribe samples' feature discrepancy to domain discrepancy, and aligning these sample's specific features lead to a lower distinguished ability among different samples (e.g., ALDA on Yelp-IMDB). All methods have similar performance on Davidson-Metoo since Davidson datasets have an extreme label distribution. Most samples focus on the same category, which causes models not to access enough samples to learn the features in other classes.

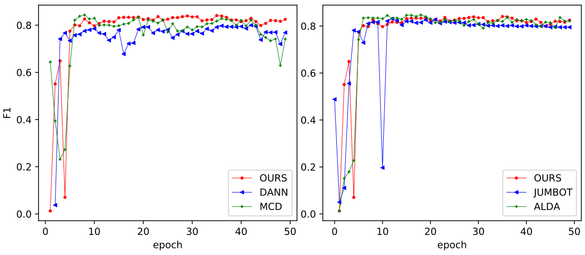

Figure 2: The convergence comparison between our model and baselines on Book-Kitchen.

Convergence Investigation The convergence curves of our model and baselines are respectively depicted in Fig. (2). We conduct a convergence experiment on Book-Kitchen datasets based on LSTM to verify the training stability during knowledge transfer. This task focuses on evaluating the ability to align domain-wise discrepancy since the feature's center of Book and Kitchen have a remarkable difference (up to 15986), but their categories are similar. Specifically, we observe that our model significantly outperforms DANN and MCD dur-

[2] https://pytorch.org/text/stable/index.html

ing training. DANN has relatively low stability since it only aligns different domain features without considering task-specific features. Compared with ALDA, our model achieves similar stability. Our model can achieve efficient convergence after iterating 15 epochs, which proves our model's robustness.

Knowledge Transfer. We measure the feature center distance between the training set in the source data and the test set in the target data to evaluate models' ability to transfer knowledge. Generally, the domain-wise discrepancy is significantly narrowed after applying domain adaptation methods. Our model achieves relatively significant improvements, but there are some exceptions. For example, ALDA has a lower domain-wise discrepancy on Davidson-MeToo than ours. However, ALDA's performance is unsatisfactory, especially when the datasets have similar domains (e.g., Yelp-IMDB and IMDB-Yelp). A similar situation also happens on DANN and MCD. These methods enlarge domain-wise discrepancy when the domains have similar feature distribution. Compared with JUMBOT, our model has a slightly large domain-wise discrepancy. However, our model is more efficient on knowledge transfer when the domain has huge category-wise discrepancies. For example, the distance of our model is .0438 on Davidson-MeToo, while the corresponding figure is .1207 on JUMBOT.

4.5 Ablation Analysis

In this subsection, we investigate the importance of different rewards in RL learning by conducting variant experiments, as shown in the Table 5.

$-R_c$ means we delete reward R_c in our R_{adv}. R_c is a unsupervised reward. Instead of aligning features, R_c aims to search subspace features, ensuing the consistent prediction between completed features $E(X)$ and sub-spaced features $M(E(X))$. This method is efficient since removing R_c is significantly detrimental to cross-domain performance. Especially, we find that R_c plays a more critical role Book-Kitchen and Kitchen-Book tasks by comparing the R_d since removing R_c lower the performance than R_d.

R_d is proposed to align domain features by fooling the discriminator. $-R_d$ means we do not need to train the discriminator and R_{adv} only combines with R_c and R_{reg}. $-R_d$ achieves a better performance than our completed model on Book-Kitchen. We infer the reason behind this is because R_d only focuses on feature shift rather than considering the discrepancy among different classes, which causes class-specific features to be weakened, and the model fails to distinguish the boundaries of other classes. However, removing R_d decreases the performance in most of the situations, which proves feature shift is efficient in domain adaptation.

Generally, R_d and R_c work together to guide critical knowledge transfer and removing any one of them degrade the performance badly. Which reward dominates an improvement depends on the datasets' property. When the domains have significant discrepancy both in features and label distribution, R_d and R_c work in an adversarial way to ensure shifting features as well as keeping class-specific features.

5 Related work

Unsupervised Domain Adaptation for text classification has several major approaches (Ramponi and Plank, 2020), such as distribution adaptation, feature selection and subspace learning. Distribution adaptation reduce the difference in the marginal distribution (Gretton et al., 2007), conditional distribution (Satpal and Sarawagi, 2007) or joint distribution (Long et al., 2013) by explicitly minimizing predefined distance measures. For example, Zhang et al. adopts the Margin Disparity Discrepancy (Zhang et al., 2019) to solve the cross-lingual text classification problems. Zellinger et al. proposes Central Moment Discrepancy (CMD), which explicitly minimizes differences of higher-order central moments for each moment order by matching the domain-specific hidden representations. Feature selection minimizes the difference between the domains by finding commonality in features or pivots. For example, SCL (Blitzer et al., 2006) uses unlabeled data and frequently-occurring pivot features from both source and target domains to find correspondences among features from these domains. PBLM (Ziser and Reichart, 2018) combines SCL with a neural language model based on long short-term memory (LSTM) networks which predict the presence of pivots and non-pivots. FSDA (Sun et al., 2019) finds informative features to reduce the domain discrepancy and eliminate noisy features by developing a cutting-plane algorithm. Subspace learning aligns the features in the different domain into the same space and then build a unified model for these domains.

Table 5: Ablation studies of our model on LSTM

Method	MeToo - Davidson	Davidson-MeToo	Book-Kitchen	Kitchen-Book	Yelp-IMDB	IMDB-Yelp
$-R_d$	31.97	23.14	86.09	44.15	43.28	61.40
$-R_c$	35.84	23.19	84.51	40.87	43.71	60.87

One of the key methods in this work focus on adversarial learning. Du et al. design a post-training procedure to distill the domain-specific features in a self-supervised way and then conduct the adversarial training to derive the enhanced domain-invariant features. Qu et al. propose an adversarial category alignment network (ACAN) to enforce the category-level alignment under a prior condition of global marginal alignment.

5.1 Reinforcement Learning

With the robustness in learning sophisticated policies, recent works introduce Reinforcement learning (RL) into the domain adaptation task (Chen et al., 2020a; Dong et al., 2020; Zhang et al., 2021). DARL (Chen et al., 2020a) employs deep Q-learning in partial domain adaptation. The DARL framework designs a reward for the agent-based on how relevant the selected source instances are to the target domain. With the action-value function optimizer, DARL can automatically select source instances in the shared classes for circumventing negative transfer as well as to simultaneously learn transferable features between domains by reducing the domain shift. However, DARL does not generalize to unsupervised domain adaptation. Highly relying on the rich labels in the source domain will cause failure when insufficient labels are in the source domain. To address this problem, Zhang et al. develop a new reward across both source and target domains. This reward can guide the agent to learn the best policy and select the closest feature pair for both domains. However, these works only focus on computer vision. To our best knowledge, we are the first work introducing RL for the UDA under the class-imbalanced text classification.

5.2 Imbalanced-class

Increasing works study the class-imbalanced domain adaptation (Tan et al., 2020; Lee et al., 2020; Bose et al., 2021; Li et al., 2020). COAL (Tan et al., 2020) deals with feature shift and label shift in a unified way. With the idea of prototype-based conditional distribution alignment and class-balanced self-training, COAL tackles feature shift in the context of label shift. However, present works only focus on computer vision, and the imbalanced class domain adaptation in NLP is unexplored. The other similar works is category-level feature alignment (Qu et al., 2019; Luo et al., 2019; Li et al., 2021, 2019; Yang et al., 2020). These works usually focus on domain shifts and propose domain-level aligned strategies while ignoring the local category-level distributions, reducing cross-domain text classifiers' effectiveness. A popular strategy for category-level alignment is aligning the same class features among different domains respectively by resorting to pseudo labels (Dong et al., 2020; Yang et al., 2020).

6 Conclusion

In this study, we have proposed an unsupervised reinforcement adaptation model (URAM) for the novel cross-domain adaptation challenge where the source and target domains are class-imbalanced. We demonstrate the effectiveness of our reinforcement approach with the other four state-of-art baselines on the task of text classification. The URAM learns domain-independent representations by leveraging three reward factors, label, domain, and domain distance, which coherently combines pivot and adversarial approaches in UDA. Extensive experiments and ablation analysis show that the URAM can obtain robust domain-invariant representations and effectively adapt text classifiers over both domains and imbalanced data.

6.1 Limitation and Future Work

Our work opens several future directions on the limitations of this study. First, class-imbalanced data naturally exist in NLP tasks, such as discourse inference (Spangher et al., 2021), text generation (Nishino et al., 2020), and question answering (Li et al., 2020). Our next step will examine the effectiveness of our model over the NLP tasks. Second, we only validate the URAM on English datasets, and additional multilingual settings will be verified in future work, such as multilingual text classification (Schwenk and Li, 2018).

References

Eyal Ben-David, Carmel Rabinovitz, and Roi Reichart. 2020. PERL: Pivot-based Domain Adaptation for Pre-trained Deep Contextualized Embedding Models. *Transactions of the Association for Computational Linguistics*, 8:504–521.

John Blitzer, Ryan McDonald, and Fernando Pereira. 2006. Domain adaptation with structural correspondence learning. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 120–128, Sydney, Australia. Association for Computational Linguistics.

Tulika Bose, Irina Illina, and Dominique Fohr. 2021. Unsupervised domain adaptation in cross-corpora abusive language detection. In *Proceedings of the Ninth International Workshop on Natural Language Processing for Social Media*, pages 113–122, Online. Association for Computational Linguistics.

Jin Chen, Xinxiao Wu, Lixin Duan, and Shenghua Gao. 2020a. Domain adversarial reinforcement learning for partial domain adaptation. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–15.

Minghao Chen, Shuai Zhao, Haifeng Liu, and Deng Cai. 2020b. Adversarial-learned loss for domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3521–3528.

Lu Cheng, Ruocheng Guo, K Selçuk Candan, and Huan Liu. 2020. Representation learning for imbalanced cross-domain classification. In *Proceedings of the 2020 SIAM international conference on data mining*, pages 478–486. SIAM.

Xia Cui, Frans Coenen, and Danushka Bollegala. 2017. Effect of data imbalance on unsupervised domain adaptation of part-of-speech tagging and pivot selection strategies. In *First International Workshop on Learning with Imbalanced Domains: Theory and Applications*, pages 103–115. PMLR.

Hal Daumé III. 2007. Frustratingly easy domain adaptation. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 256–263.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Jiahua Dong, Yang Cong, Gan Sun, Yuyang Liu, and Xiaowei Xu. 2020. Cscl: Critical semantic-consistent learning for unsupervised domain adaptation. In *Computer Vision – ECCV 2020*, pages 745–762, Cham. Springer International Publishing.

Chunning Du, Haifeng Sun, Jingyu Wang, Qi Qi, and Jianxin Liao. 2020. Adversarial and domain-aware BERT for cross-domain sentiment analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4019–4028, Online. Association for Computational Linguistics.

Kilian Fatras, Thibault Séjourné, Nicolas Courty, and Rémi Flamary. 2021. Unbalanced minibatch optimal transport; applications to domain adaptation. *CoRR*, abs/2103.03606.

Yaroslav Ganin and Victor Lempitsky. 2015. Unsupervised domain adaptation by backpropagation. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, page 1180–1189. JMLR.org.

Arthur Gretton, Karsten Borgwardt, Malte Rasch, Bernhard Schölkopf, and Alex Smola. 2007. A kernel method for the two-sample-problem. In *Advances in Neural Information Processing Systems*, volume 19. MIT Press.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Comput.*, 9(8):1735–1780.

Joe Hoover, Gwenyth Portillo-Wightman, Leigh Yeh, Shreya Havaldar, Aida Mostafazadeh Davani, Ying Lin, Brendan Kennedy, Mohammad Atari, Zahra Kamel, Madelyn Mendlen, Gabriela Moreno, Christina Park, Tingyee E. Chang, Jenna Chin, Christian Leong, Jun Yen Leung, Arineh Mirinjian, and Morteza Dehghani. 2020. Moral foundations twitter corpus: A collection of 35k tweets annotated for moral sentiment. *Social Psychological and Personality Science*, 11(8):1057–1071.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Vijay Konda and John Tsitsiklis. 2000. Actor-critic algorithms. In *Advances in Neural Information Processing Systems*, volume 12. MIT Press.

Suhyeon Lee, Junhyuk Hyun, Hongje Seong, and Euntai Kim. 2020. Unsupervised domain adaptation for semantic segmentation by content transfer. *CoRR*, abs/2012.12545.

Lusi Li, Haibo He, Jie Li, and Guang Yang. 2019. Adversarial domain adaptation via category transfer. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8.

Shuai Li, Jianqiang Huang, Xiansheng Hua, and Lei Zhang. 2021. Category dictionary guided unsupervised domain adaptation for object detection. In *AAAI*.

Xiaoya Li, Xiaofei Sun, Yuxian Meng, Junjun Liang, Fei Wu, and Jiwei Li. 2020. Dice loss for data-imbalanced NLP tasks. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 465–476, Online. Association for Computational Linguistics.

M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu. 2013. Transfer feature learning with joint distribution adaptation. In *2013 IEEE International Conference on Computer Vision (ICCV)*, pages 2200–2207, Los Alamitos, CA, USA. IEEE Computer Society.

Yawei Luo, Liang Zheng, Tao Guan, Junqing Yu, and Yi Yang. 2019. Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2502–2511.

Jianmo Ni, Jiacheng Li, and Julian McAuley. 2019. Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 188–197, Hong Kong, China. Association for Computational Linguistics.

Toru Nishino, Ryota Ozaki, Yohei Momoki, Tomoki Taniguchi, Ryuji Kano, Norihisa Nakano, Yuki Tagawa, Motoki Taniguchi, Tomoko Ohkuma, and Keigo Nakamura. 2020. Reinforcement learning with imbalanced dataset for data-to-text medical report generation. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2223–2236, Online. Association for Computational Linguistics.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.

Xiaoye Qu, Zhikang Zou, Yu Cheng, Yang Yang, and Pan Zhou. 2019. Adversarial category alignment network for cross-domain sentiment classification. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2496–2508, Minneapolis, Minnesota. Association for Computational Linguistics.

Alan Ramponi and Barbara Plank. 2020. Neural unsupervised domain adaptation in NLP—A survey. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6838–6855, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Kuniaki Saito, Kohei Watanabe, Y. Ushiku, and Tatsuya Harada. 2018a. Maximum classifier discrepancy for unsupervised domain adaptation. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3723–3732.

Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. 2018b. Maximum classifier discrepancy for unsupervised domain adaptation. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3723–3732.

Sandeepkumar Satpal and Sunita Sarawagi. 2007. Domain adaptation of conditional probability models via feature subsetting. In *Proceedings of the 11th European Conference on Principles and Practice of Knowledge Discovery in Databases*, ECMLPKDD'07, page 224–235, Berlin, Heidelberg. Springer-Verlag.

Holger Schwenk and Xian Li. 2018. A corpus for multilingual document classification in eight languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Alexander Spangher, Jonathan May, Sz-Rung Shiang, and Lingjia Deng. 2021. Multitask semi-supervised learning for class-imbalanced discourse classification. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 498–517, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Feng Sun, Hanrui Wu, Zhihang Luo, Wenwen Gu, Yuguang Yan, and Qing Du. 2019. Informative feature selection for domain adaptation. *IEEE Access*, 7:142551–142563.

Shuhan Tan, Xingchao Peng, and Kate Saenko. 2020. Class-imbalanced domain adaptation: An empirical odyssey. In *Computer Vision – ECCV 2020 Workshops*, pages 585–602, Cham. Springer International Publishing.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

Guanglei Yang, Haifeng Xia, Mingli Ding, and Zhengming Ding. 2020. Bi-directional generation for unsupervised domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 6615–6622.

Werner Zellinger, Thomas Grubinger, Edwin Lughofer, Thomas Natschläger, and Susanne Saminger-Platz. 2017. Central moment discrepancy (CMD) for domain-invariant representation learning. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net.

Dejiao Zhang, Ramesh Nallapati, Henghui Zhu, Feng Nan, Cicero Nogueira dos Santos, Kathleen McKeown, and Bing Xiang. 2020. Margin-aware unsupervised domain adaptation for cross-lingual text labeling. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 3527–3536, Online. Association for Computational Linguistics.

Youshan Zhang, Hui Ye, and Brian D. Davison. 2021. Adversarial reinforcement learning for unsupervised domain adaptation. In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 635–644.

Yuchen Zhang, Tianle Liu, Mingsheng Long, and Michael Jordan. 2019. Bridging theory and algorithm for domain adaptation. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 7404–7413. PMLR.

Yftah Ziser and Roi Reichart. 2018. Pivot based language modeling for improved neural domain adaptation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1241–1251, New Orleans, Louisiana. Association for Computational Linguistics.

A Imbalanced Data Label Distribution

	Docs	Tokens	pos/neg
D-DVD	2000	30.51	2.52
E-Electronic	2000	27.65	2.26

Table 6: Stats of the Amazon review data. We present the average number of tokens and the imbalanced-class ratio.

B Cross-domain Evaluations

dataset	Non-moral	Degradation	Harm	Fairness	Subversion	Care	Cheating	Purity	Betrayal	Authority	Loyalty
MeToo	21.40	15,30	6.86	6.30	14.70	3.40	11.00	2.98	5.83	6.93	5.29
BLM	23.59	4.23	19.36	8.58	5.74	5.93	13.84	2.76	2.71	5.40	7.83
Sandy	13.68	1.94	1.69	3.82	9.63	21.30	9.80	1.45	3.12	9.46	8.86
Davidson	92.13	1.34	2.76	0.08	0.14	0.18	1.24	0.10	0.82	0.40	0.82
Baltimore	54.93	0.55	4.86	2.60	5.34	3.26	9.38	0.69	11.18	0.40	6.83
ALM	20.98	3.18	19.15	13.42	2.37	11.88	13.16	2.11	1.04	6.36	6.36
Election	47.70	2.13	9.09	8.66	2.55	6.15	9.59	6.32	1.98	2.61	3.20

Table 7: Label distributions of the multi-class morality dataset (Hoover et al., 2020)

No-adapt	MeToo	BLM	Sandy	Davidson	Baltimore	ALM	Election
MeToo	47.16	18.09	6.28	35.61	29.58	14.61	16.95
BLM	16.23	76.32	17.22	26.27	25.28	16.16	26.40
Sandy	8.81	14.46	58.50	19.27	7.49	15.68	9.04
Davidson	23.12	31.98	8.09	99.17	66.96	24.93	58.49
Baltimore	23.32	32.42	10.07	99.17	66.54	25.00	59.09
ALM	12.11	17.60	14.27	24.88	25.12	43.71	20.33
Election	23.18	32.59	15.24	99.11	66.57	24.95	58.87

MCD	MeToo	BLM	Sandy	Davidson	Baltimore	ALM	Election
MeToo	48.14	25.86	13.77	40.25	38.86	22.81	32.41
BLM	16.48	78.42	17.27	29.17	55.27	23.40	34.51
Sandy	24.37	16.68	60.17	15.74	32.50	16.52	12.58
Davidson	23.62	31.99	13.94	99.17	66.96	25.73	58.49
Baltimore	23.12	32.44	14.80	99.17	66.21	24.93	59.09
ALM	16.88	23.37	15.48	37.11	34.33	63.18	25.22
Election	23.12	32.53	14.10	99.17	66.54	24.93	63.91

DANN	MeToo	BLM	Sandy	Davidson	Baltimore	ALM	Election
MeToo	40.03	17.98	9.74	45.00	20.65	13.69	24.30
BLM	16.33	75.40	15.48	35.68	22.94	17.82	24.39
Sandy	8.37	14.55	56.84	6.78	6.47	14.65	9.34
Davidson	23.17	31.98	8.17	99.17	66.96	24.93	58.49
Baltimore	23.17	32.42	9.82	99.17	66.24	24.95	59.03
ALM	12.63	16.78	14.93	19.18	20.87	60.88	17.26
Election	23.14	32.57	14.23	99.17	66.57	24.93	64.01

JUMBOT	MeToo	BLM	Sandy	Davidson	Baltimore	ALM	Election
MeToo	43.12	28.32	10.47	46.94	42.33	21.08	36.11
BLM	24.37	72.57	16.02	52.20	48.92	32.18	48.91
Sandy	19.34	33.17	57.60	10.86	41.23	30.86	39.59
Davidson	23.26	32.99	8.35	99.17	66.96	26.64	58.49
Baltimore	23.48	32.66	12.16	99.17	66.18	25.03	59.09
ALM	23.30	39.82	17.04	66.60	61.70	42.01	46.50
Election	23.12	32.49	15.20	99.17	66.42	24.93	60.41

ALDA	MeToo	BLM	Sandy	Davidson	Baltimore	ALM	Election
MeToo	21.50	25.89	14.17	38.21	1.12	9.84	58.75
BLM	14.82	56.82	13.97	51.90	39.98	16.53	23.39
Sandy	23.36	14.23	34.84	33.81	6.01	22.06	28.03
Davidson	23.31	31.99	26.59	99.17	66.96	32.31	58.49
Baltimore	23.03	31.63	8.77	42.12	65.33	25.50	28.77
ALM	22.43	14.83	5.94	31.16	58.96	38.50	37.35
Election	25.44	39.70	19.16	98.32	66.54	23.17	58.87

URAM	MeToo	BLM	Sandy	Davidson	Baltimore	ALM	Election
MeToo	45.54	19.34	10.48	47.07	38.14	16.97	34.80
BLM	16.03	79.12	15.86	50.31	30.57	18.56	26.74
Sandy	9.28	14.65	60.44	10.50	10.28	15.28	8.86
Davidson	24.00	32.53	11.59	99.17	66.96	25.02	58.49
Baltimore	23.10	28.57	12.09	98.96	63.52	24.93	53.43
ALM	12.58	16.51	15.70	34.43	27.88	63.11	17.29
Election	22.54	31.92	12.38	99.06	58.10	24.88	65.23

Table 8: Cross-domain performance evaluation over the Morality dataset (Hoover et al., 2020) using F1. Each subtable presents results of one UDA model.

	book-dvd	dvd-book	book-eletronic	eletronic-book	kitchen-eletronic	eletronic-kitchen	dvd-kitchen	kitchen-dvd	dvd-eletroic	eletronic-dvd
DANN	83.16	94.00	86.87	92.15	95.24	91.21	94.24	94.29	94.63	92.57
MCD	84.39	94.34	85.06	93.36	94.08	91.61	94.14	94.99	94.22	92.54
JUMBOT	82.27	91.51	77.34	84.83	92.91	85.58	92.49	94.01	91.64	92.23
ALDA	84.49	93.52	84.14	94.49	93.93	92.39	92.70	94.21	94.00	90.91
URAM	86.56	94.58	87.90	93.51	94.96	92.87	94.81	95.15	95.03	93.02

Table 9: Cross-domain performance evaluation over the Amazon review dataset (Blitzer et al., 2006).