Table 4: The Test Classification Accuracy on The Data Sets Generated from 20Newsgroups and Reuters-21578

Data Set	LG	SVM	SFA	TPLSA	CDPLSA	$\text{TCA}^{\text{share}}$	TCA
Comp vs Rec	0.906	0.895	**0.939**	0.910	0.914	0.867	**0.940**
Comp vs Sci	0.759	0.719	0.830	0.802	0.877	0.792	**0.891**
Comp vs Talk	0.911	0.898	**0.971**	0.938	0.955	0.912	0.967
Rec vs Sci	0.719	0.696	0.885	**0.928**	0.872	0.735	0.879
Rec vs Talk	0.848	0.827	0.935	0.849	0.912	0.828	**0.962**
Sci vs Talk	0.780	0.747	0.854	0.890	0.862	0.785	**0.940**
Orgs vs People	0.681	0.670	0.671	0.746	**0.808**	0.731	0.792
Orgs vs Places	0.692	0.669	0.683	0.719	0.714	0.660	**0.730**
People vs Places	0.513	0.520	0.506	0.623	0.548	0.614	**0.626**
Average	0.757	0.738	0.808	0.823	0.829	0.769	**0.859**

Table 5: Example of Domain-Specific Topics Extracted by Our Method on Data Set *Comp vs Sci*

Source Domain $\mathcal{D}^s$				Target Domain $\mathcal{D}^t$			
Topic 1	Topic 2	Topic 3	Topic 4	Topic 1	Topic 2	Topic 3	Topic 4
windows	drive	edu	space	jpeg	mac	encryption	medical
dos	scsi	space	nasa	image	apple	government	disease
file	mb	henry	launch	file	db	clipper	health
com	ide	writes	earth	bit	edu	chip	cancer
edu	controller	nasa	spacecraft	gif	drive	law	patients
mouse	disk	toronto	orbit	color	scsi	key	hiv
os	bus	article	satellite	images	lc	security	treatment
ms	drives	pat	system	files	mb	privacy	vitamin
microsoft	hard	shuttle	solar	format	quadra	escrow	aids
win	dx	cost	data	quality	writes	nsa	infection

Implementation Details and Parameter Settings

For data preprocessing, we convert all words to lower cases and remove stop words. Besides, we filter out the words with document frequencies less than 3. For TCA, we set the total number of topics (i.e., $K + K^{\ell}$) in each domain to 12 and 20 for the experiments on 20Newsgroups and Reuters-21578, respectively. The topic numbers are tuned on some documents from data sets *Sci vs Talk* and *Orgs vs Places*. After tuning, the topic numbers are fixed and respectively used for the experiments on 20Newsgroups and Reuters-21578. The tuned documents are then put back to the original data sets. Without any prior knowledge, we simply set the proportion of the shared topics in each domain to 0.5, that is $K = K^{\ell}$. Since topic modeling methods can be sensitive to initial parameter values, we use the output of PLSA trained on individual domains to initialize the domain-specific topics, and the output of PLSA trained on the merged domain to initialize the shared topics. We set the hyperparameter α for Beta distribution to 20 in all experiments. The number of EM iterations is set to 200. LIBLINEAR LG (Fan et al. 2008) is used as the base classifier, and all parameters are set to their default values. For SFA, TPLSA and CDPLSA, we adopt the same parameter settings as the original papers.

Experimental Results

In the first experiment, we compare our method with all baselines. Table 4 summarizes the classification performance on each data set. The last row of the table shows the average accuracy over all data sets. From the table, we can observe that our proposed method outperforms all baselines on six data sets. Table 5 presents the examples of extracted domain-specific topics on data set *Comp vs Sci*. We sort the words with the learned topic-word probability. The associated topic mapping matrix $\mathbf{U}$ is:

$$\mathbf{U} = \begin{bmatrix} 0.8046 & 0.4330 & -0.7232 & -0.3768 \\ 0.4348 & 0.9665 & -0.5876 & -0.4789 \\ -0.1364 & -0.3183 & 0.5664 & 0.1692 \\ -0.4292 & -0.4613 & 0.1722 & 0.1312 \end{bmatrix}$$

By examining the topical words and the topic mapping matrix, we can observe that the domain-specific topics having positive correlation scores are always semantically relevant. For example, Topic 1 in the source domain is about the Windows OS, and its positively correlated topics in the target domain (i.e., Topic 1 and Topic 2) are about the computer graphics and the Mac hardware, respectively. The result shows that our method can effectively identify the correlations between domain-specific features from different domains.

In the second experiment, we study the parameter sensitivity for the proportion of shared topics in each domain. In this experiment, we fix the total number of topics in each domain and vary the proportion of shared topics. Figure 2a shows the average classification accuracy of TCA under varying proportions of shared topics. We can observe that TCA performs well and steadily when the proportion of shared topics ranges from 0.4 to 0.6, which verifies the effectiveness of jointly modeling both the shared and the domain-specific topics.