Automatically Extracting Polarity-Bearing Topics for Cross-Domain Sentiment Classification

Yulan He **Chenghua Lin**[†] **Harith Alani**
Knowledge Media Institute, The Open University
Milton Keynes MK7 6AA, UK
{y.he,h.alani}@open.ac.uk
[†] School of Engineering, Computing and Mathematics
University of Exeter, Exeter EX4 4QF, UK
cl322@exeter.ac.uk

Abstract

Joint sentiment-topic (JST) model was previously proposed to detect sentiment and topic simultaneously from text. The only supervision required by JST model learning is domain-independent polarity word priors. In this paper, we modify the JST model by incorporating word polarity priors through modifying the topic-word Dirichlet priors. We study the polarity-bearing topics extracted by JST and show that by augmenting the original feature space with polarity-bearing topics, the in-domain supervised classifiers learned from augmented feature representation achieve the state-of-the-art performance of 95% on the movie review data and an average of 90% on the multi-domain sentiment dataset. Furthermore, using feature augmentation and selection according to the information gain criteria for cross-domain sentiment classification, our proposed approach performs either better or comparably compared to previous approaches. Nevertheless, our approach is much simpler and does not require difficult parameter tuning.

1 Introduction

Given a piece of text, sentiment classification aims to determine whether the semantic orientation of the text is positive, negative or neutral. Machine learning approaches to this problem (**?**; **?**; **?**; **?**; **?**; **?**) typically assume that classification models are trained and tested using data drawn from some fixed distribution. However, in many practical cases, we may have plentiful labeled examples in the *source* domain, but very few or no labeled examples in the *target* domain with a different distribution. For example, we may have many labeled books reviews, but we are interested in detecting the polarity of electronics reviews. Reviews for different produces might have widely different vocabularies, thus classifiers trained on one domain often fail to produce satisfactory results when shifting to another domain. This has motivated much research on sentiment transfer learning which transfers knowledge from a source task or domain to a different but related task or domain (**?**; **?**; **?**; **?**).

Joint sentiment-topic (JST) model (**?**; **?**) was extended from the latent Dirichlet allocation (LDA) model (**?**) to detect sentiment and topic simultaneously from text. The only supervision required by JST learning is domain-independent polarity word prior information. With prior polarity words extracted from both the MPQA subjectivity lexicon[1] and the appraisal lexicon[2], the JST model achieves a sentiment classification accuracy of 74% on the movie review data[3] and 71% on the multi-domain sentiment dataset[4]. Moreover, it is also able to extract coherent and informative topics grouped under different sentiment. The fact that the JST model does not required any labeled documents for training makes it desirable for domain adaptation in sentiment classification. Many existing approaches solve the sentiment transfer problem by associating words

[1] http://www.cs.pitt.edu/mpqa/

[2] http://lingcog.iit.edu/arc/appraisal_lexicon_2007b.tar.gz

[3] http://www.cs.cornell.edu/people/pabo/movie-review-data

[4] http://www.cs.jhu.edu/~mdredze/datasets/sentiment/index2.html

Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, pages 123–131,
Portland, Oregon, June 19-24, 2011.

from different domains which indicate the same sentiment (**?**; **?**). Such an association mapping problem can be naturally solved by the posterior inference in the JST model. Indeed, the polarity-bearing topics extracted by JST essentially capture sentiment associations among words from different domains which effectively overcome the data distribution difference between source and target domains.

The previously proposed JST model uses the sentiment prior information in the Gibbs sampling inference step that a sentiment label will only be sampled if the current word token has no prior sentiment as defined in a sentiment lexicon. This in fact implies a different generative process where many of the word prior sentiment labels are observed. The model is no longer "latent". We propose an alternative approach by incorporating word prior polarity information through modifying the topic-word Dirichlet priors. This essentially creates an informed prior distribution for the sentiment labels and would allow the model to actually be latent and would be consistent with the generative story.

We study the polarity-bearing topics extracted by the JST model and show that by augmenting the original feature space with polarity-bearing topics, the performance of in-domain supervised classifiers learned from augmented feature representation improves substantially, reaching the state-of-the-art results of 95% on the movie review data and an average of 90% on the multi-domain sentiment dataset. Furthermore, using simple feature augmentation, our proposed approach outperforms the structural correspondence learning (SCL) (**?**) algorithm and achieves comparable results to the recently proposed spectral feature alignment (SFA) method (**?**). Nevertheless, our approach is much simpler and does not require difficult parameter tuning.

We proceed with a review of related work on sentiment domain adaptation. We then briefly describe the JST model and present another approach to incorporate word prior polarity information into JST learning. We subsequently show that words from different domains can indeed be grouped under the same polarity-bearing topic through an illustration of example topic words extracted by JST before proposing a domain adaptation approach based on JST. We verify our proposed approach by conducting experiments on both the movie review data and the multi-domain sentiment dataset. Finally, we conclude our work and outline future directions.

2 Related Work

There has been significant amount of work on algorithms for domain adaptation in NLP. Earlier work treats the source domain data as "prior knowledge" and uses maximum a posterior (MAP) estimation to learn a model for the target domain data under this prior distribution (**?**). Chelba and Acero (**?**) also uses the source domain data to estimate prior distribution but in the context of a maximum entropy (ME) model. The ME model has later been studied in (**?**) for domain adaptation where a mixture model is defined to learn differences between domains.

Other approaches rely on unlabeled data in the target domain to overcome feature distribution differences between domains. Motivated by the alternating structural optimization (ASO) algorithm (**?**) for multi-task learning, Blitzer et al. (**?**) proposed structural correspondence learning (SCL) for domain adaptation in sentiment classification. Given labeled data from a source domain and unlabeled data from target domain, SCL selects a set of pivot features to link the source and target domains where pivots are selected based on their common frequency in both domains and also their mutual information with the source labels.

There has also been research in exploring careful structuring of features for domain adaptation. Daumé (**?**) proposed a kernel-mapping function which maps both source and target domains data to a high-dimensional feature space so that data points from the same domain are twice as similar as those from different domains. Dai et al.(**?**) proposed translated learning which uses a language model to link the class labels to the features in the source spaces, which in turn is translated to the features in the target spaces. Dai et al. (**?**) further proposed using spectral learning theory to learn an eigen feature representation from a task graph representing features, instances and class labels. In a similar vein, Pan et al. (**?**) proposed the spectral feature alignment (SFA) algorithm where some domain-independent words are used as a bridge to construct a bipartite graph to model the co-occurrence relationship between domain-specific words and domain-independent words. Feature clusters are

generated by co-align domain-specific and domain-independent words.

Graph-based approach has also been studied in (**?**) where a graph is built with nodes denoting documents and edges denoting content similarity between documents. The sentiment score of each unlabeled documents is recursively calculated until convergence from its neighbors the actual labels of source domain documents and pseudo-labels of target document documents. This approach was later extended by simultaneously considering relations between documents and words from both source and target domains (**?**).

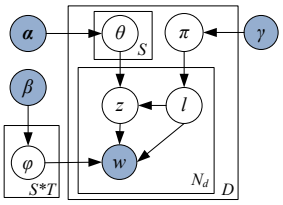
More recently, Seah et al. (**?**) addressed the issue when the predictive distribution of class label given input data of the domains differs and proposed Predictive Distribution Matching SVM learn a robust classifier in the target domain by leveraging the labeled data from only the relevant regions of multiple sources.

3 Joint Sentiment-Topic (JST) Model

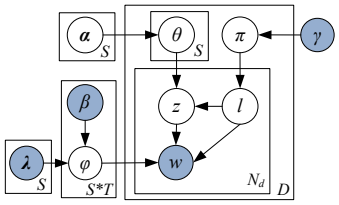
Assume that we have a corpus with a collection of D documents denoted by $C = \{d_1, d_2, ..., d_D\}$; each document in the corpus is a sequence of N_d words denoted by $d = (w_1, w_2, ..., w_{N_d})$, and each word in the document is an item from a vocabulary index with V distinct terms denoted by $\{1, 2, ..., V\}$. Also, let S be the number of distinct sentiment labels, and T be the total number of topics. The generative process in JST which corresponds to the graphical model shown in Figure **??**(a) is as follows:

- For each document d, choose a distribution $\pi_d \sim \text{Dir}(\gamma)$.
- For each sentiment label l under document d, choose a distribution $\theta_{d,l} \sim \text{Dir}(\alpha)$.
- For each word w_i in document d
  - choose a sentiment label $l_i \sim \text{Mult}(\pi_d)$,
  - choose a topic $z_i \sim \text{Mult}(\theta_{d,l_i})$,
  - choose a word w_i from $\varphi^{l_i}_{z_i}$, a Multinomial distribution over words conditioned on topic z_i and sentiment label l_i.

Gibbs sampling was used to estimate the posterior distribution by sequentially sampling each variable of interest, z_t and l_t here, from the distribution over



(a) JST model.



(b) Modified JST model.

Figure 1: JST model and its modified version.

that variable given the current values of all other variables and data. Letting the superscript $-t$ denote a quantity that excludes data from t^{th} position, the conditional posterior for z_t and l_t by marginalizing out the random variables φ, θ, and π is

$$P(z_t = j, l_t = k | \mathbf{w}, \mathbf{z^{-t}}, \mathbf{l^{-t}}, \alpha, \beta, \gamma) \propto \frac{N^{-t}_{w_t,j,k} + \beta}{N^{-t}_{j,k} + V\beta} \cdot \frac{N^{-t}_{j,k,d} + \alpha_{j,k}}{N^{-t}_{k,d} + \sum_j \alpha_{j,k}} \cdot \frac{N^{-t}_{k,d} + \gamma}{N^{-t}_{d} + S\gamma}. \quad (1)$$

where $N_{w_t,j,k}$ is the number of times word w_t appeared in topic j and with sentiment label k, $N_{j,k}$ is the number of times words assigned to topic j and sentiment label k, $N_{j,k,d}$ is the number of times a word from document d has been associated with topic j and sentiment label k, $N_{k,d}$ is the number of times sentiment label k has been assigned to some word tokens in document d, and N_d is the total number of words in the document collection.

In the modified JST model as shown in Figure **??**(b), we add an additional dependency link of $\boldsymbol{\varphi}$ on the matrix $\boldsymbol{\lambda}$ of size $S \times V$ which we use to encode word prior sentiment information into the JST model. For each word $w \in \{1, ..., V\}$, if w is found in the sentiment lexicon, for each $l \in \{1, ..., S\}$, the element λ_{lw} is updated as follows

$$\lambda_{lw} = \begin{cases} 1 & \text{if } S(w) = l \\ 0 & \text{otherwise} \end{cases}, \quad (2)$$

where the function $S(w)$ returns the prior sentiment label of w in a sentiment lexicon, i.e. neutral, posi-

	Book	DVD	Book	Elec.	Book	Kitch.	DVD	Elec.	DVD	Kitch.	Elec.	Kitch.
Pos.	recommend	funni	interest	pictur	interest	qualiti	concert	sound	movi	recommend	sound	pleas
	highli	cool	topic	clear	success	easili	rock	listen	stori	highli	excel	look
	easi	entertain	knowledg	paper	polit	servic	favorit	bass	classic	perfect	satisfi	worth
	depth	awesom	follow	color	clearli	stainless	sing	amaz	fun	great	perform	materi
	strong	worth	easi	accur	popular	safe	talent	acoust	charact	qulati	comfort	profession
Neg.	mysteri	cop	abus	problem	bore	return	bore	poorli	horror	cabinet	tomtom	elimin
	fbi	shock	question	poor	tediou	heavi	plot	low	alien	break	region	regardless
	investig	prison	mislead	design	cheat	stick	stupid	replac	scari	install	error	cheapli
	death	escap	point	case	crazi	defect	stori	avoid	evil	drop	code	plain
	report	dirti	disagre	flaw	hell	mess	terribl	crap	dead	gap	dumb	incorrect

Table 1: Extracted polarity words by JST on the combined data sets.

tive or negative.

The matrix $\boldsymbol{\lambda}$ can be considered as a transformation matrix which modifies the Dirichlet priors $\boldsymbol{\beta}$ of size $S \times T \times V$, so that the word prior polarity can be captured. For example, the word "*excellent*" with index i in the vocabulary has a positive polarity. The corresponding row vector in $\boldsymbol{\lambda}$ is $[0, 1, 0]$ with its elements representing neutral, positive, and negative. For each topic j, multiplying λ_{li} with β_{lji}, only the value of $\beta_{l_{pos}ji}$ is retained, and $\beta_{l_{neu}ji}$ and $\beta_{l_{neg}ji}$ are set to 0. Thus, the word "*excellent*" can only be drawn from the positive topic word distributions generated from a Dirichlet distribution with parameter $\boldsymbol{\beta}_{l_{pos}}$.

4 Polarity Words Extracted by JST

The JST model allows clustering different terms which share similar sentiment. In this section, we study the polarity-bearing topics extracted by JST. We combined reviews from the source and target domains and discarded document labels in both domains. There are a total of six different combinations. We then run JST on the combined data sets and listed some of the topic words extracted as shown in Table **??**. Words in each cell are grouped under one topic and the upper half of the table shows topic words under the positive sentiment label while the lower half shows topic words under the negative sentiment label.

We can see that JST appears to better capture sentiment association distribution in the source and target domains. For example, in the DVD+Elec. set, words from the DVD domain describe a rock concert DVD while words from the Electronics domain are likely relevant to stereo amplifiers and receivers, and yet they are grouped under the same topic by the JST model. Checking the word coverage in each domain reveals that for example "bass" seldom appears in the DVD domain, but appears more often in the Electronics domain. Likewise, in the Book+Kitch. set, "stainless" rarely appears in the Book domain and "interest" does not occur often in the Kitchen domain and they are grouped under the same topic. These observations motivate us to explore polarity-bearing topics extracted by JST for cross-domain sentiment classification since grouping words from different domains but bearing similar sentiment has the effect of overcoming the data distribution difference of two domains.

5 Domain Adaptation using JST

Given input data x and a class label y, labeled patterns of one domain can be drawn from the joint distribution $P(x, y) = P(y|x)P(x)$. Domain adaptation usually assume that data distribution are different in source and target domains, i.e., $P_s(x) \neq P_t(x)$. The task of domain adaptation is to predict the label y_i^t corresponding to x_i^t in the target domain.

We assume that we are given two sets of training data, $\mathcal{D}^s$ and $\mathcal{D}^t$, the *source domain* and *target domain* data sets, respectively. In the multiclass classification problem, the source domain data consist of labeled instances, $\mathcal{D}^s = \{(x_n^s; y_n^s) \in \mathcal{X} \times \mathcal{Y} : 1 \leq n \leq N^s\}$, where $\mathcal{X}$ is the input space and $\mathcal{Y}$ is a finite set of class labels. No class label is given in the target domain, $\mathcal{D}^t = \{x_n^t \in \mathcal{X} : 1 \leq n \leq N^t, N^t \gg N^s\}$. Algorithm **??** shows how to perform domain adaptation using the JST model. The source and target domain data are first merged with document labels discarded. A JST model is then

learned from the merged corpus to generate polarity-bearing topics for each document. The original documents in the source domain are augmented with those polarity-bearing topics as shown in Step 4 of Algorithm **??**, where $l_i\_z_i$ denotes a combination of sentiment label l_i and topic z_i for word w_i. Finally, feature selection is performed according to the information gain criteria and a classifier is then trained from the source domain using the new document representations. The target domain documents are also encoded in a similar way with polarity-bearing topics added into their feature representations.

Algorithm 1 Domain adaptation using JST.

Input: The source domain data $\mathcal{D}^s = \{(x_n^s; y_n^s) \in \mathcal{X} \times \mathcal{Y} : 1 \leq n \leq N^s\}$, the target domain data, $\mathcal{D}^t = \{x_n^t \in \mathcal{X} : 1 \leq n \leq N^t, N^t \gg N^s\}$

Output: A sentiment classifier for the target domain $\mathcal{D}^t$

1: Merge $\mathcal{D}^s$ and $\mathcal{D}^t$ with document labels discarded, $\mathcal{D} = \{(x_n^s, 1 \leq n \leq N^s; x_n^t, 1 \leq n \leq N^t\}$
2: Train a JST model on $\mathcal{D}$
3: **for** each document $x_n^s = (w_1, w_2, ..., w_m) \in \mathcal{D}^s$ **do**
4: Augment document with polarity-bearing topics generated from JST, $x_n^{s'} = (w_1, w_2, ..., w_m, l_1\_z_1, l_2\_z_2, ..., l_m\_z_m)$
5: Add $\{x_n^{s'}; y_n^s\}$ into a document pool $\mathcal{B}$
6: **end for**
7: Perform feature selection using IG on $\mathcal{B}$
8: Return a classifier, trained on $\mathcal{B}$

As discussed in Section **??** that the JST model directly models $P(\mathbf{l}|d)$, the probability of sentiment label given document, and hence document polarity can be classified accordingly. Since JST model learning does not require the availability of document labels, it is possible to augment the source domain data by adding most confident pseudo-labeled documents from the target domain by the JST model as shown in Algorithm **??**.

6 Experiments

We evaluate our proposed approach on the two datasets, the movie review (MR) data and the multi-domain sentiment (MDS) dataset. The movie review data consist of 1000 positive and 1000 negative movie reviews drawn from the IMDB movie archive while the multi-domain sentiment dataset contains four different types of product reviews extracted from Amazon.com including Book, DVD, Electronics and Kitchen appliances. Each category

Algorithm 2 Adding pseudo-labeled documents.

Input: The target domain data, $\mathcal{D}^t = \{x_n^t \in \mathcal{X} : 1 \leq n \leq N^t, N^t \gg N^s\}$, document sentiment classification threshold τ

Output: A labeled document pool $\mathcal{B}$

1: Train a JST model parameterized by Λ on $\mathcal{D}^t$
2: **for** each document $x_n^t \in \mathcal{D}^t$ **do**
3: Infer its sentiment class label from JST as $l_n = \arg\max_s P(l|x_n^t; \Lambda)$
4: **if** $P(l_n|x_n^t; \Lambda) > \tau$ **then**
5: Add labeled sample (x_n^t, l_n) into a document pool $\mathcal{B}$
6: **end if**
7: **end for**

of product reviews comprises of 1000 positive and 1000 negative reviews and is considered as a domain. Preprocessing was performed on both of the datasets by removing punctuation, numbers, non-alphabet characters and stopwords. The MPQA subjectivity lexicon is used as a sentiment lexicon in our experiments.

6.1 Experimental Setup

While the original JST model can produce reasonable results with a simple symmetric Dirichlet prior, here we use asymmetric prior $\boldsymbol{\alpha}$ over the topic proportions which is learned directly from data using a fixed-point iteration method (**?**).

In our experiment, $\boldsymbol{\alpha}$ was updated every 25 iterations during the Gibbs sampling procedure. In terms of other priors, we set symmetric prior $\beta = 0.01$ and $\gamma = (0.05 \times L)/S$, where L is the average document length, and the value of 0.05 on average allocates 5% of probability mass for mixing.

6.2 Supervised Sentiment Classification

We performed 5-fold cross validation for the performance evaluation of supervised sentiment classification. Results reported in this section are averaged over 10 such runs. We have tested several classifiers including Naïve Bayes (NB) and support vector machines (SVMs) from WEKA[5], and maximum entropy (ME) from MALLET[6]. All parameters are set to their default values except the Gaussian

[5] http://www.cs.waikato.ac.nz/ml/weka/
[6] http://mallet.cs.umass.edu/

prior variance is set to 0.1 for the ME model training. The results show that ME consistently outperforms NB and SVM on average. Thus, we only report results from ME trained on document vectors with each term weighted according to its frequency.

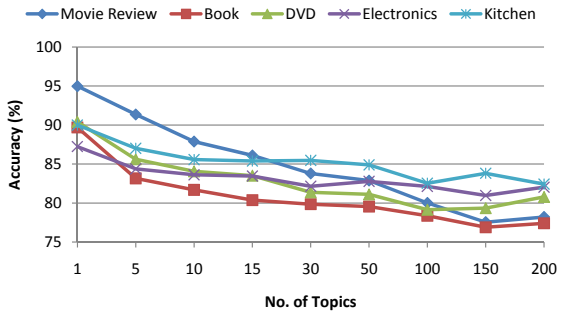


Figure 2: Classification accuracy vs. no. of topics.

The only parameter we need to set is the number of topics T. It has to be noted that the actual number of feature clusters is $3 \times T$. For example, when T is set to 5, there are 5 topic groups under each of the positive, negative, or neutral sentiment labels and hence there are altogether 15 feature clusters. The generated topics for each document from the JST model were simply added into its bag-of-words (BOW) feature representation prior to model training. Figure **??** shows the classification results on the five different domains by varying the number of topics from 1 to 200. It can be observed that the best classification accuracy is obtained when the number of topics is set to 1 (or 3 feature clusters). Increasing the number of topics results in the decrease of accuracy though it stabilizes after 15 topics. Nevertheless, when the number of topics is set to 15, using JST feature augmentation still outperforms ME without feature augmentation (the baseline model) in all of the domains. It is worth pointing out that the JST model with single topic becomes the standard LDA model with only three sentiment topics. Nevertheless, we have proposed an effective way to incorporate domain-independent word polarity prior information into model learning. As will be shown later in Table **??** that the JST model with word polarity priors incorporated performs significantly better than the LDA model without incorporating such prior information.

For comparison purpose, we also run the LDA model and augmented the BOW features with the generated topics in a similar way. The best accuracy was obtained when the number of topics is set to 15 in the LDA model. Table **??** shows the classification accuracy results with or without feature augmentation. We have performed significance test and found that LDA performs statistically significant better than Baseline according to a paired t-test with $p < 0.005$ for the Kitchen domain and with $p < 0.001$ for all the other domains. JST performs statistically significant better than both Baseline and LDA with $p < 0.001$.

Method	*MR*	*MDS*			
		Book	*DVD*	*Elec.*	*Kitch.*
Baseline	82.53	79.96	81.32	83.61	85.82
LDA	83.76	84.32	85.62	85.4	87.68
JST	**94.98**	**89.95**	**91.7**	**88.25**	**89.85**
[YE10]	91.78	82.75	82.85	84.55	87.9
[LI10]	-	79.49	81.65	83.64	85.65

Table 2: Supervised sentiment classification accuracy.

We also compare our method with other recently proposed approaches. Yessenalina et al. (**?**) explored different methods to automatically generate annotator rationales to improve sentiment classification accuracy. Our method using JST feature augmentation consistently performs better than their approach (denoted as [YE10] in Table **??**). They further proposed a two-level structured model (**?**) for document-level sentiment classification. The best accuracy obtained on the MR data is 93.22% with the model being initialized with sentence-level human annotations, which is still worse than ours. Li et al. (**?**) adopted a two-stage process by first classifying sentences as personal views and impersonal views and then using an ensemble method to perform sentiment classification. Their method (denoted as [LI10] in Table **??**) performs worse than either LDA or JST feature augmentation. To the best of our knowledge, the results achieved using JST feature augmentation are the state-of-the-art for both the MR and the MDS datasets.

6.3 Domain Adaptation

We conducted domain adaptation experiments on the MDS dataset comprising of four different domains, Book (B), DVD (D), Electronics (E), and Kitchen appliances (K). We randomly split each do-

main data into a training set of 1,600 instances and a test set of 400 instances. A classifier trained on the training set of one domain is tested on the test set of a different domain. We preformed 5 random splits and report the results averaged over 5 such runs.

Comparison with Baseline Models

We compare our proposed approaches with two baseline models. The first one (denoted as "Base" in Table **??**) is an ME classifier trained without adaptation. LDA results were generated from an ME classifier trained on document vectors augmented with topics generated from the LDA model. The number of topics was set to 15. JST results were obtained in a similar way except that we used the polarity-bearing topics generated from the JST model. We also tested with adding pseudo-labeled examples from the JST model into the source domain for ME classifier training (following Algorithm **??**), denoted as "JST-PL" in Table **??**. The document sentiment classification probability threshold τ was set to 0.8. Finally, we performed feature selection by selecting the top 2000 features according to the information gain criteria ("JST-IG")[7].

There are altogether 12 cross-domain sentiment classification tasks. We showed the adaptation loss results in Table **??** where the result for each domain and for each method is averaged over all three possible adaptation tasks by varying the source domain. The adaptation loss is calculated with respect to the in-domain gold standard classification result. For example, the in-domain goal standard for the Book domain is 79.96%. For adapting from DVD to Book, baseline achieves 72.25% and JST gives 76.45%. The adaptation loss is 7.71 for baseline and 3.51 for JST.

It can be observed from Table **??** that LDA only improves slightly compared to the baseline with an error reduction of 11%. JST further reduces the error due to transfer by 27%. Adding pseudo-labeled examples gives a slightly better performance compared to JST with an error reduction of 36%. With feature selection, JST-IG outperforms all the other approaches with a relative error reduction of 53%.

[7] Both values of 0.8 and 2000 were set arbitrarily after an initial run on some held-out data; they were not tuned to optimize test performance.

Domain	Base	LDA	JST	JST-PL	JST-IG
Book	10.8	9.4	7.2	6.3	**5.2**
DVD	8.3	6.1	4.8	4.4	**2.9**
Electr.	7.9	7.7	6.3	5.4	**3.9**
Kitch.	7.6	7.6	6.9	6.1	**4.4**
Average	8.6	7.7	6.3	5.5	**4.1**

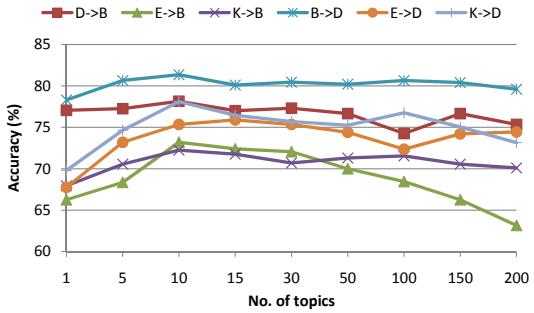
Table 3: Adaptation loss with respect to the in-domain gold standard. The last row shows the average loss over all the four domains.

Parameter Sensitivity

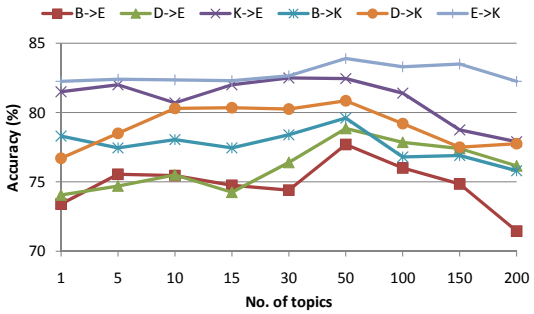
There is only one parameters to be set in the JST-IG approach, the number of topics. We plot the classification accuracy versus different topic numbers in Figure **??** with the number of topics varying between 1 and 200, corresponding to feature clusters varying between 3 and 600. It can be observed that for the relatively larger Book and DVD data sets, the accuracies peaked at topic number 10, whereas for the relatively smaller Electronics and Kitchen data sets, the best performance was obtained at topic number 50. Increasing topic numbers results in the decrease of classification accuracy. Manually examining the extracted polarity topics from JST reveals that when the topic number is small, each topic cluster contains well-mixed words from different domains. However, when the topic number is large, words under each topic cluster tend to be dominated by a single domain.

Comparison with Existing Approaches

We compare in Figure **??** our proposed approach with two other domain adaptation algorithms for sentiment classification, SCL and SFA. Each set of bars represent a cross-domain sentiment classification task. The thick horizontal lines are in-domain sentiment classification accuracies. It is worth noting that our in-domain results are slightly different from those reported in (**?**; **?**) due to different random splits. Our proposed JST-IG approach outperforms SCL in average and achieves comparable results to SFA. While SCL requires the construction of a reasonable number of auxiliary tasks that are useful to model "pivots" and "non-pivots", SFA relies on a good selection of domain-independent features for the construction of bipartite feature graph before running spectral clustering to derive feature clusters.

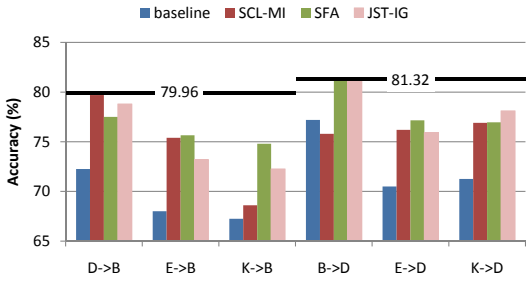


(a) Adapted to Book and DVD data sets.

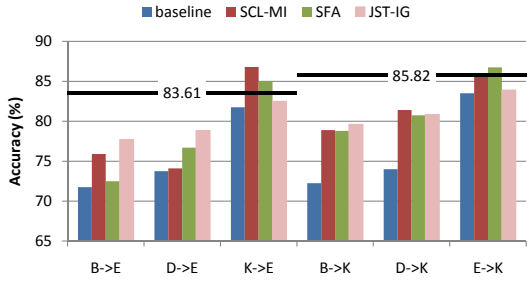


(b) Adapted to Electronics and Kitchen data sets.

Figure 3: Classification accuracy vs. no. of topics.



(a) Adapted to Book and DVD data sets.



(b) Adapted to Electronics and Kitchen data sets.

Figure 4: Comparison with existing approaches.

On the contrary, our proposed approach based on the JST model is much simpler and yet still achieves comparable results.

7 Conclusions

In this paper, we have studied polarity-bearing topics generated from the JST model and shown that by augmenting the original feature space with polarity-bearing topics, the in-domain supervised classifiers learned from augmented feature representation achieve the state-of-the-art performance on both the movie review data and the multi-domain sentiment dataset. Furthermore, using feature augmentation and selection according to the information gain criteria for cross-domain sentiment classification, our proposed approach outperforms SCL and gives similar results as SFA. Nevertheless, our approach is much simpler and does not require difficult parameter tuning.

There are several directions we would like to explore in the future. First, polarity-bearing topics generated by the JST model were simply added into the original feature space of documents, it is worth investigating attaching different weight to each topic maybe in proportional to the posterior probability of sentiment label and topic given a word estimated by the JST model. Second, it might be interesting to study the effect of introducing a tradeoff parameter to balance the effect of original and new features. Finally, our experimental results show that adding pseudo-labeled examples by the JST model does not appear to be effective. We could possibly explore instance weight strategies (**?**) on both pseudo-labeled examples and source domain training examples in order to improve the adaptation performance.

Acknowledgements

This work was supported in part by the EC-FP7 projects ROBUST (grant number 257859).

References

R.K. Ando and T. Zhang. 2005. A framework for learning predictive structures from multiple tasks and unlabeled data. *The Journal of Machine Learning Research*, 6:1817–1853.

A. Aue and M. Gamon. 2005. Customizing sentiment classifiers to new domains: a case study. In *Proceedings of Recent Advances in Natural Language Processing (RANLP)*.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan.

2003. Latent Dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022.

J. Blitzer, M. Dredze, and F. Pereira. 2007. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *ACL*, page 440–447.

C. Chelba and A. Acero. 2004. Adaptation of maximum entropy classifier: Little data can help a lot. In *EMNLP*.

W. Dai, Y. Chen, G.R. Xue, Q. Yang, and Y. Yu. 2008. Translated learning: Transfer learning across different feature spaces. In *NIPS*, pages 353–360.

W. Dai, O. Jin, G.R. Xue, Q. Yang, and Y. Yu. 2009. Eigentransfer: a unified framework for transfer learning. In *ICML*, pages 193–200.

H. Daumé III and D. Marcu. 2006. Domain adaptation for statistical classifiers. *Journal of Artificial Intelligence Research*, 26(1):101–126.

H. Daumé. 2007. Frustratingly easy domain adaptation. In *ACL*, pages 256–263.

J. Jiang and C.X. Zhai. 2007. Instance weighting for domain adaptation in NLP. In *ACL*, pages 264–271.

A. Kennedy and D. Inkpen. 2006. Sentiment classification of movie reviews using contextual valence shifters. *Computational Intelligence*, 22(2):110–125.

S. Li, C.R. Huang, G. Zhou, and S.Y.M. Lee. 2010. Employing personal/impersonal views in supervised and semi-supervised sentiment classification. In *ACL*, pages 414–423.

C. Lin and Y. He. 2009. Joint sentiment/topic model for sentiment analysis. In *Proceedings of the 18th ACM international conference on Information and knowledge management (CIKM)*, pages 375–384.

C. Lin, Y. He, and R. Everson. 2010. A Comparative Study of Bayesian Models for Unsupervised Sentiment Detection. In *Proceedings of the 14th Conference on Computational Natural Language Learning (CoNLL)*, pages 144–152.

Ryan McDonald, Kerry Hannan, Tyler Neylon, Mike Wells, and Jeff Reynar. 2007. Structured models for fine-to-coarse sentiment analysis. In *ACL*, pages 432–439.

T. Minka. 2003. Estimating a Dirichlet distribution. Technical report.

S.J. Pan, X. Ni, J.T. Sun, Q. Yang, and Z. Chen. 2010. Cross-domain sentiment classification via spectral feature alignment. In *Proceedings of the 19th international conference on World Wide Web (WWW)*, pages 751–760.

Bo Pang and Lillian Lee. 2004. A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In *ACL*, page 271–278.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *EMNLP*, pages 79–86.

B. Roark and M. Bacchiani. 2003. Supervised and unsupervised PCFG adaptation to novel domains. In *NAACL-HLT*, pages 126–133.

C.W. Seah, I. Tsang, Y.S. Ong, and K.K. Lee. 2010. Predictive Distribution Matching SVM for Multi-domain Learning. In *ECML-PKDD*, pages 231–247.

Casey Whitelaw, Navendu Garg, and Shlomo Argamon. 2005. Using appraisal groups for sentiment analysis. In *Proceedings of the ACM international conference on Information and Knowledge Management (CIKM)*, pages 625–631.

Q. Wu, S. Tan, and X. Cheng. 2009. Graph ranking for sentiment transfer. In *ACL-IJCNLP*, pages 317–320.

Q. Wu, S. Tan, X. Cheng, and M. Duan. 2010. MIEA: a Mutual Iterative Enhancement Approach for Cross-Domain Sentiment Classification. In *COLING*, page 1327-1335.

A. Yessenalina, Y. Choi, and C. Cardie. 2010a. Automatically generating annotator rationales to improve sentiment classification. In *ACL*, pages 336–341.

A. Yessenalina, Y. Yue, and C. Cardie. 2010b. Multi-Level Structured Models for Document-Level Sentiment Classification. In *EMNLP*, pages 1046–1056.

Jun Zhao, Kang Liu, and Gen Wang. 2008. Adding redundant features for CRFs-based sentence sentiment classification. In *EMNLP*, pages 117–126.