COLUMN: AFFECTIVE COMPUTING AND SENTIMENT ANALYSIS

Instance-based Domain Adaptation via Multiclustering Logistic Approximation

Feng Xu
Nanjing University of Science and Technology

Jianfei Yu
Singapore Management University

Rui Xia
Nanjing University of Science and Technology

Editor:
Erik Cambria, Nanyang Technological University

With the explosive growth in the number of online texts, we could easily collect a large amount of labeled training data from different source domains. However, a basic assumption in building statistical machine learning models for sentiment analysis is that the training and test data must be drawn from the same distribution. Otherwise, directly training a statistical model usually results in poor performance. Faced with the massive amount of labeled data from different domains, it is important to identify the source-domain training instances that are closely relevant to the target domain and make better use of them. In this work, we propose a new approach, called multiclustering logistic approximation (MLA), to address this problem. In MLA, we adapt the source-domain training data to the target domain via a framework of multiclustering logistic approximation. Experimental results demonstrate that MLA has significant advantages over the state-of-the-art instance adaptation methods, especially in the scenario of multidistributional training data.

With the growing volume of text available via the Internet, we can easily obtain huge amounts of labeled training texts from different domains. But only some of them might be beneficial for training a target-domain-desired classifier for sentiment analysis. A lot of research has been proposed to conduct domain adaptation for sentiment analysis,[1,2] open-domain opinion mining, and sentiment analysis.[3,4]

Published by the IEEE Computer Society
1541-1672/18/$33.00 ©2018 IEEE

As a special type of domain adaptation method for sentiment analysis, instance-based domain adaptation (instance adaptation for short) aims to identify the training samples that are most relevant to the target domain and make better use of them. Consider the following example. We want to learn a laptop sentiment classifier in the absence of labeled laptop reviews. Instead, we can obtain a large set of labeled e-product reviews, which covers reviews from several kinds of e-products including phones and digital cameras. In this case, the performance of a sentiment classifier simply trained with all labeled reviews might be unsatisfactory because only a few samples in the training data address topics closely related to laptops. Therefore, methods are needed to identify and adapt the labeled reviews from different source domains (or subdomains) to obtain a classifier that performs well on the laptop domain.

To address this problem, methods have been developed to conduct instance adaptation by reweighting the training instances and applying importance sampling. In the field of machine learning, instance adaptation was studied under the concepts of covariate shift, sample selection bias, and importance sampling. Most of the previous instance adaptation methods traditionally assume that the source-domain training data has just one underline distribution.

However, in many NLP applications including sentiment analysis, we are usually faced with a large amount of labeled data that might come from multiple source domains. Even if the data comes from one source domain, it also might contain many subdomains with different patterns in distribution. For example, the reviews collected from multiple domains (such as movie, book, or e-product) are multidomain training data. Even if the training data might come from one domain, there are still many subdomains or categories. For example, if the reviews come from the e-product domain, they might still cover several smaller subdomains (such as phones and laptops) and, therefore, have distinct differences in distribution.

An in-target-domain logistic approximation (ILA) approach has been proposed for single-domain instance adaptation.[5] On this basis, we propose a multiclustering logistic approximation (MLA) in this work, as an extension to ILA, to deal with domain adaption in case of multidistributional training data. Moreover, we infer the instance weighting learning criterion based on the multinomial event model, which leads to more profound insights of instance-based domain adaptation. To fully evaluate MLA, we conduct experiments on two tasks including cross-domain sentiment analysis and cross-domain text categorization. The experimental results show that our MLA approach can significantly outperform the state-of-the-art instance adaptation methods, especially in the case of multidistributional training data.

RELATED WORK

In general, domain adaptation methods include feature-based domain adaptation, parameter-based domain adaptation, and instance-based domain adaptation.[1,6] Note that different methods have different settings. In this work, we focus on instance-based domain adaptation.

Instance adaptation learns the importance of labeled data in the source domain by instance reweighting and importance sampling. The reweighted instances are then used for training a target-domain model. In the machine learning community, instance adaptation is also known as the covariate shift or instance selection bias. This concept was first introduced in the field of econometrics[7] and then brought into the field of machine learning.[8] The key problem in instance selection bias is density ratio estimation (DRE).

There was a line of kernel-based methods to solve the DRE problem, such as kernel density estimation,[9] maximum entropy density estimation,[10] kernel mean matching,[11] etc. A KLIEP algorithm was proposed to directly estimate the density ratio by using a linear model in a Gaussian kernel space.[12] Parameters were learned by minimizing the K-L divergence between the true and approximated distributions.[13,14]

A logistic regression model was used to learn the density ratio together with the classification parameters, under the multitask learning framework.[15,16] In the field of NLP, an instance selection and instance weighting approach via PU learning (PUIS and PUIW) was proposed for the task of cross-domain sentiment classification.[17] An instance-based domain adaptation in NLP via in-target-domain logistic approximation has also been proposed.[5]

As mentioned above, in many applications in NLP, the source-domain training data may come from many subdomains and have different distributions. While most of the previous work conducted domain adaptation in case of single-distributional training data. By contrast, in this work, we proposed a multiclustering logistic regression model to address this issue.

MULTICLUSTERING LOGISTIC APPROXIMATION

Logistic Approximation for Single Source-Domain Instance Adaptation

In-target-domain logistic approximation (ILA) for instance adaptation from a single source domain has been introduced.[5] This assumed that the target-domain data is generated as follows:

- an instance x is first drawn from the source domain distribution $p_s(x)$; and
- an in-target-domain selector $p(d=1 \mid x) = 1/(1+e^{(-\beta^T x)})$ then adapts x from the source to the target domain, where d denotes a domain indicator ($d = 1$ represents the target domain, and $d = 0$ represents the source domain).

The approximated target-domain distribution was then formulated as $q_t(x) = p(d=1|x)p_s(x) = a/(1+e^{(-\beta^T x)} ?)p_s(x)$.

The normalized in-target-domain probability $w(x) = a/(1+e^{(-\beta^T x)})$ was used as the instance weight for training a weighted classification model after instance adaptation. Different instance weights can yield different target-domain approximated distributions. The instance weight can be learned by minimizing the statistical distance (such as K-L distance) between the target-domain true distribution $p_t(x)$ and the approximated distribution $q_t(x)$.

However, when the training data contains many subdomains and has distinct distributions, ILA might lead to poor adaptation performance, especially when the gap between subdomains is large.

Let us use an artificial example to illustrate the motivation. In Figure 1, the red dots denote the target-domain test data. The blue and black crosses denote the training data, which are drawn from two different distributions. In instance adaptation, we learn an in-target-domain selector and assign an in-target-domain probability as the weight to each training instance. The size of the cross denotes the weight. The crosses that are closer to the separating line will have larger weights. Figure 1 illustrates how ILA conducts instance adaptation. It treats the multidistributional training data as a single domain, and learns a single in-target-domain separating line. The crosses that are closer to the separating line will have larger weights.

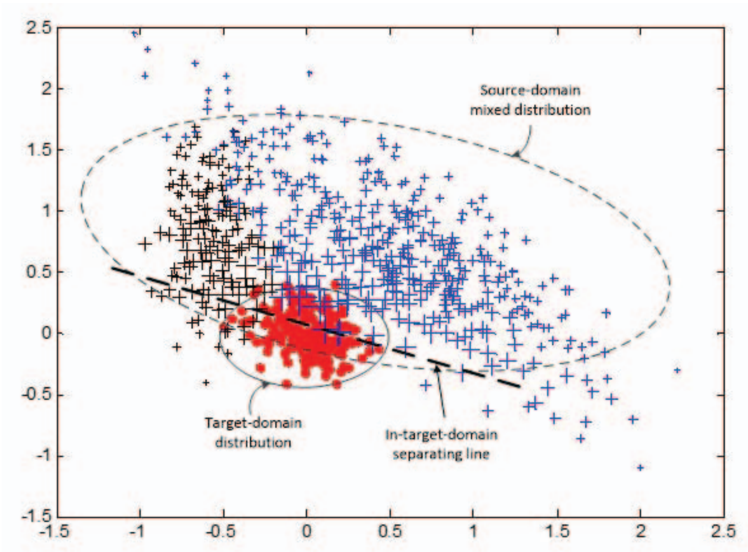

Figure 1. An illustration of ILA instance adaptation for multidistributional training data.

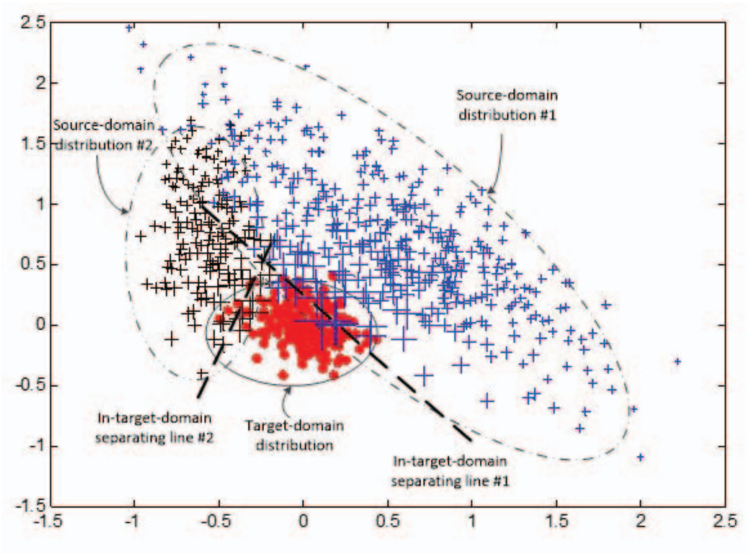

Figure 2. An illustration of MLA instance adaptation for multidistributional training data.

Multiclustering Logistic Approximation

Unlike the standard ILA, in this work we propose a multiclustering logistic approximation (MLA) to address this issue. Figure 2 shows how MLA works in this case. It first clusters the training data into several categories and treats each category as a single domain. Then, it learns the in-target-domain separating lines for each cluster, and conducts instance adaptation respectively. A weighted combination of the instance weights learned in each cluster is finally utilized as the overall instance weight. In comparison with ILA, MLA can capture various patterns in the massive training data. Therefore, it would be more reasonable in case of multidomain or multidistributional instance adaptation.

We first apply a clustering algorithm to cluster the training data into several subdomains. Thereafter, we develop a multiclustering in-target-domain logistic approximation model to adapt data from different clusters in the manner of a weighted combination.

In MLA, the k-means clustering algorithm is employed to split the source domain data into k subdomains. The cosine distance is used as the similarity measure. After clustering, the cosine distance between different instances within a domain is small, while the distance between different domains is large.

Based on the k clusters of training data, we propose to conduct multiclustering logistic approximation. Let $p_{(s_d)}(x)$ be the distribution of the dth cluster of the training data, and $q_{(t_d)}(x)$ be the (component) target-domain approximated distribution adapted from $p_{(s_d)}(x)$. We suppose the target-domain approximated distribution $q_t(x)$ in MLA is a weighted sum of $q_{(t_d)}(x)$:

$$q_t(x) = \sum_{d=1}^{k} \eta_d q_{t_d}(x) = \sum_{d=1}^{k} \eta_d a_d / (1+e^{(-\beta_d^T x)}) p_{(s_d)}(x) \quad (1)$$

in which a_d and β_d. are respectively the normalization factor and feature weight of the dth cluster, and η_d is a tradeoff parameter controlling the importance of each cluster. The weighted ensemble of the normalized in-target-domain probabilities

$$w(x) = \sum_{d=1}^{K} \eta_d a_d / (1+e^{(-\beta_d^T x)}) \quad (2)$$

will be used as sampling weights for training a weighted classification model.

Under this assumption, the K-L distance between $p_t(x)$ and $q_t(x)$ can be written approximately as

$$\begin{aligned} KL(p_t \,\|\, q_t) &\approx 1/N_t \sum_{d=1}^{K} \eta_d \sum_{i=1}^{N_t} \log(p_t(x)) / (q_{(t_d)}(x)) \\ &= 1/N_t \sum_{d=1}^{K} \eta_d \sum_{i=1}^{N_t} log(p_t(x)) / ((a_d p_{(s_d)}(x)) / 1+e^{(-\beta_d^T x_i)}) \qquad (3) \\ &= 1/N_t \sum_{d=1}^{K} \eta_d \sum_{i=1}^{N_t} log \frac{p_t(x)}{p_{s_d}(x)} - \frac{1}{N_t} \sum_{d=1}^{K} \eta_d \sum_{i=1}^{N_t} log \frac{a_d}{(1+e^{(-\beta_d^T x_i)})} \end{aligned}$$

The parameters are learned by minimizing $KL(p_t \,\|\, q_t)$ subject to the normalization constraint

$$\begin{aligned} & min(a_d, \beta_d)\, KL(p_t \,\|\, q_t) = min(a_d, \beta_d) - \frac{1}{N_t} \sum_{d=1}^{K} \eta_d \sum_{i=1}^{N_t} log \frac{a_d}{(1+e^{(-\beta_d^T x_i)})} \\ & s.t. \int_{x \epsilon X} q_{(t_d)}(x) dx = 1/N_d \sum_{(j=1)}^{N_d} a_d / (1+e^{(-\beta_d^T x_j)}) = 1, d = 1, \ldots, k. \end{aligned} \quad (4)$$

in which N_d is the number of training instances in the dth cluster.

Similarly, we get the final cost function by solving a_d analytically and plugging it back:

$$\begin{aligned} J_{mila} &= -\frac{1}{N_t} \sum_{d=1}^{K} \eta_d \sum_{i=1}^{N_t} log \frac{N_d}{\dfrac{\sum_{j=1}^{N_d} \frac{1}{1+e^{(-\beta_d^T x_j^T)}}}{1+e^{(-\beta_d^T x_i)}}} \\ &= \frac{1}{N_t} \sum_{d=1}^{K} \eta_d \sum_{i=1}^{N_t} [log(1+e^{(-\beta_d^T x_i)}) + log \sum_{j=1}^{N_d} \frac{1}{1+e^{(-\beta_d^T x_j^T)}}] \end{aligned} \quad (5)$$

For each clustering, once the parameters a_d and β_d are learned, we could calculate the instance weights according to Equation (2). Based on this, we train an instance-weighted classifier for the cross-domain classification task.

EXPERIMENTAL STUDY

Tasks, Datasets, and Experimental Settings

To fully evaluate the performance of MLA, we conducted experiments on two NLP tasks: cross-domain text categorization and cross-domain sentiment classification.

For cross-domain text categorization, we use the 20 Newsgroups dataset for experiments. It contains seven top categories, under which there are 20 subcategories. We use four top categories as the class labels, and generate source and target domains based on subcategories. Taking "sci vs talk" as an example, the top categories "sci" and "talk" are the class labels. The subcategories "med," "elec" (under category "sci"), "guns," and "mideast" (under category "talk") are used as the multidistributional training data. Subcategories "crypt" (under category "sci") and "misc" (under category "talk") are used as the target-domain test data.

For cross-domain sentiment classification, we randomly chose two domains from four mutidomain sentiment datasets as the source domain, and chose one from the remaining two domains as the target domain. For example, "book + kitchen → dvd" represents that "book" and "kitchen" are the two subdomains in the source domain, and "dvd" is the target domain. It is worth pointing out that here that we only present the result of MLA when the number of subdomains is two. Similar behaviors can be observed when the number of subdomains increases.

In both tasks, unigrams and bigrams with term frequency no less than four are used as features for classification. We randomly repeat the experiments for 10 times and report the average results. The paired *t*-test[18] is employed for significance testing, with a default significant level of 0.05.

The Comparison of System Performance

The following systems are implemented for comparison:

- No-adaptation: the standard approach using all training data without domain adaptation
- KLIEP-Gaussian: the KLIEP model with a Gaussian kernel[12]
- PUIW: the instance-weighting model via PU learning[2]
- ILA: the standard in-target-domain logistic approximation algorithm[5]

We compare the system performance on two tasks respectively.

Text categorization: In Table 1, we can observe that in comparison with the No-adaptation system, the improvements of KLIEP-Gaussian and PUIW are very slight. ILA is a bit more effective. But the effect is quite limited compared with its performance in single-domain instance adaptation.[5] In contrast, MLA outperforms No-adaptation, KLIEP-Gaussian, PUIW, and ILA significantly.

Sentiment classification: In comparison with No-adaptation, three instance adaptation methods (KLIEP-Gaussian, PUIW, and ILA) exhibit effective performance. ILA does not show significant superiority when compared with KLIEP-Gaussian and PUIW. By contrast, MLA outperforms all the other methods significantly.

In general, ILA does not yield significant improvements in comparison with KLIEP-Gaussian and PUIW, in multidistributional cases. It suggests that the effect of ILA is limited for multidistributional instance adaptation. By contrast, the MLA algorithm is rather effective in this setting.

Table 1. Multidomain instance adaptation performance of different systems on text categorization

Dataset	No-adaptation	KLIEP-Gaussian	PUIW	ILA	MLA

talk vs rec	0.809	0.821	0.826	0.851	**0.857**
talk vs com	0.952	**0.961**	0.954	0.958	**0.961**
sci vs talk	0.720	0.728	0.725	0.727	**0.735**
sci vs com	0.727	0.718	0.728	0.725	**0.738**
rec vs com	0.880	0.883	0.885	**0.890**	0.887
rec vs sci	0.770	0.770	0.784	0.801	**0.807**
Average	0.809	0.814	0.817	0.825	**0.831**

Table 2. Multidomain instance adaptation performance of different systems on sentiment classification

Dataset	**No-adaptation**	**KLIEP-Gaussian**	**PUIW**	**ILA**	**MLA**
book + dvd → kitchen	0.797	0.809	0.802	0.810	**0.818**
book + dvd → elec	0.733	0.782	0.772	0.778	**0.791**
book + elec → kitchen	0.84	0.846	0.847	0.846	**0.854**
book + elec → dvd	0.801	0.807	0.811	0.811	**0.818**
book + kitchen → dvd	0.81	0.811	0.815	0.816	**0.817**
book + kitchen → elec	0.822	0.828	0.833	0.827	**0.843**
dvd + elec → kitchen	0.85	0.851	0.852	0.853	**0.854**
dvd + elec → book	0.789	0.79	0.798	0.798	**0.800**
dvd + kitchen → book	0.798	0.802	0.799	0.780	**0.820**
dvd + kitchen → elec	0.814	0.820	0.819	0.814	**0.842**
elec + kitchen → dvd	0.781	0.783	0.782	0.783	**0.787**
elec + kitchen → book	0.753	0.756	0.758	0.752	**0.763**
Average	0.799	0.807	0.807	0.806	**0.818**

Parameter Stability of the Clustering Weight

In this part, we discuss the sensitivity of the domain-based tradeoff parameter in MLA. In Figures 3 and 4, we presented the results of eight datasets in two tasks. It can be seen that all the accuracy curves are bimodal and stable. When η_d is relatively small, i.e., $\eta_d < 0.5$, the best accuracy is obtained around 0.1 to 0.4. When η_d becomes larger, the best accuracy is obtained when η_d is located at 0.7 to 0.9. It suggests that our approach MLA inclines to choose more target-domain-likely samples from one domain, but less from the others.

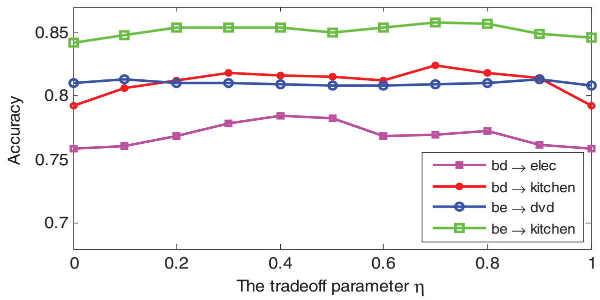

Figure 3. Sensitivity of the tradeoff parameter in MILA in sentiment classification.

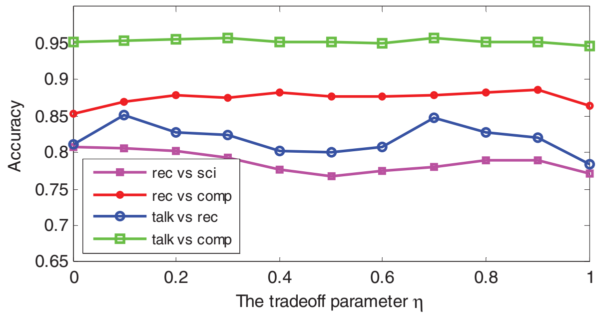

Figure 4. Sensitivity of the tradeoff parameter in MILA in text categorization.

The Relation Between K-L Distance and Instance Adaptation Performance

We further investigate the relation between the K-L distances and domain adaptation performance. In Table 3, we report three kind of K-L distance between the data of source and target domain respectively:

1. KLD-1 represents the K-L distance between source domain #1 and the target domain,
2. KLD-2 represents the K-L distance between source domain #2 and the target domain, and
3. KLD-3 represents the K-L distance between the source domains.

Table 3. Domain adaptation performance of different systems on sentiment classification

Dataset	**KLD-1**	**KLD-2**	**KLD-3**	**No adaptation**	**ILA**	**MLA**
book + dvd → kitchen	31.48	31.98	19.15	0.797	0.810	**0.818**
book + dvd → elec	31.44	31.59	19.66	0.733	0.778	**0.791**
book + elec → kitchen	35.38	30.33	74.32	0.840	0.846	**0.854**
book + elec → dvd	18.07	60.88	73.32	0.801	0.811	**0.818**
book + kitchen → dvd	17.35	16.26	80.28	0.810	0.816	**0.817**

book + kitchen → elec	33.24	34.28	81.76	0.822	0.827	**0.843**
dvd + elec → kitchen	36.83	30.24	73.65	0.850	0.853	**0.854**
dvd + elec → book	56.54	60.52	71.15	0.789	0.798	**0.800**
dvd + kitchen → book	19.99	101.24	51.14	0.798	0.780	**0.820**
dvd + kitchen → elec	32.87	56.42	54.73	0.814	0.814	**0.842**
elec + kitchen → dvd	46.44	52.96	49.10	0.781	0.783	**0.787**
elec + kitchen → book	49.28	54.44	51.40	0.753	0.752	**0.763**

Taking "book + kitchen → dvd" for example, KLD-1 is the K-L distance between "book" and "dvd," KLD-2 is the K-L distance between "kitchen" and "dvd," and KLD-3 is the K-L distance between "book" and "kitchen."

First, it can be seen that when the K-L distance between the two source domains is small (KLD-3 < 10), the improvement of MLA is also limited. It shows that when the K-L distance between source domains is small, ILA and MLA exhibit competitive performance. The reason is that when KLD-3 is relatively small, the two source sub-domains have similar distributional difference.

Second, when the K-L distance between the two source domains is relatively large (KLD-3 > 40), we can observe MLA perform much better than ILA. Compared to the ILA method, the average improvements on the last ten subtasks are 1.3 percent. But it is also worth noting that when KLD-1 and KLD-2 are close to each other, the improvement is less than 1 percent, such as "book + kitchen → dvd" and "dvd + electronics → kitchen." By contrast, if the difference between KLD-1 and KLD-2 is large, the improvement is more significant, such as "dvd + kitchen → book" and "dvd + kitchen → elec." The reason is that if the distance between the distribution of the subdomains and target domain is large, the ILA model will assign large weights to few samples, while the weights of other samples in the source domain are close to zero. This might lead to the overfitting of domain adaptation. However, the MLA model could avoid the overfitting problem by tuning the parameter to assign large weights to more samples from the source domain whose distribution is much more similar to the target domain.

This result confirms our motivation of MLA very well. When KLD-3 is small, i.e., the distributions of the two source domains are similar, the effects of ILA (viewing the two distributions as a whole distribution, and choosing samples from it) and MLA (choosing samples from the two distributions respectively) are similar. When KLD-3 is large, it means the difference between the two distributions is large. For MLA, it could tune the parameter to pay more attention to the source domain whose distribution is much more similar to the target domain and assign large weights to more samples from this source domain.

CONCLUSIONS

The traditional instance adaptation methods including in-target-domain logistic approximation (ILA) normally conducted domain adaptation for a single source domain. In this work, a multi-clustering logistic approximation (MLA) model is proposed, to conduct instance adaptation for multidistributional-labelled training data, where the training data might come from many subdomains. MLA extends the ILA algorithm and is more suitable and more efficient in the multi-distributional case. We conduct systematic experiments on the tasks of cross-domain sentiment classification and text categorization. The results indicate that MLA has significant advantages over traditional instance adaptation methods, especially when the gap between each subdomain in the training data is large.

ACKNOWLEDGMENTS

This work was supported by the Natural Science Foundation of China (no. 61672288), and the Natural Science Foundation of Jiangsu Province for Excellent Young Scholars (no. BK20160085).

REFERENCES

1. J. Jiang, *A Literature Survey on Domain Adaptation of Statistical Classifiers*, 2008; http://sifaka.cs.uiuc.edu/jiang4/domain_adaptation/survey/da_survey.pdf.
2. R. Xia et al., “Feature Ensemble plus Sample Selection: Domain Adaptation for Sentiment Classification,” *IEEE Intelligent Systems*, vol. 28, no. 3, 2013, pp. 10–18.
3. E. Cambria et al., “Semantic Multidimensional Scaling for Open-Domain Sentiment Analysis,” *IEEE Intelligent Systems*, vol. 29, no. 2, 2014, pp. 44–51.
4. E. Cambria, “Affective Computing and Sentiment Analysis,” *IEEE Intelligent Systems*, vol. 31, no. 2, 2016, pp. 102–107.
5. R. Xia et al., “Instance-Based Domain Adaptation in NLP via In-Target-Domain Logistic Approximation,” *Proc. AAAI Conf. Artificial Intelligence* (AAAI), 2014, pp. 1600–1606.
6. J. Pan and Q. Yang, “A Survey on Transfer Learning,” *IEEE Trans. Knowledge and Data Eng.*, vol. 22, no. 10, 2010, pp. 1345–1359.
7. J. Heckman, “Sample Selection Bias as a Specification Error,” *Econometrica*, vol. 47, no. 1, 1979, pp. 153–161.
8. B. Zadrozny, “Learning and Evaluating Classifiers under Sample Selection Bias,” *Proc. Int'l Conf. Machine Learning* (ICML), 2004, p. 114.
9. H. Shimodaira, “Improving Predictive Inference under Covariate Shift by Weighting the Log-Likelihood Function,” *J. Statistical Planning and Inference*, vol. 90, 2000, pp. 227–244.
10. M. Dudik, R. Schapire, and S. Phillips, “Correcting Sample Selection Bias in Maximum Entropy Density Estimation,” *Proc. Advances in Neural Information Processing Systems* (NIPS), 2005, pp. 3232–330.
11. J. Huang et al., “Correcting Sample Selection Bias by Unlabeled Data,” *Proc. Advances in Neural Information Processing Systems* (NIPS), 2006, pp. 601–608.
12. M. Sugiyama et al., “Direct importance estimation with model selection and its application to covariate shift adaptation,” *Advances in Neural Information Processing Systems* (NIPS), 2007, pp. 1433–1440.
13. Y. Tsuboi et al., “Direct Density Ratio Estimation for Large-scale Covariate Shift Adaptation,” *Proceedings of the SIAM International Conference on Data Mining* (SDM), 2009.
14. T. Kanamori, S. Hido, and M. Sugiyama, “A least-squares approach to direct importance estimation,” *Journal of Machine Learning Research*, vol. 10, 2009, pp. 1391–1445.
15. S. Bickel, M. Brückner, and T. Scheffer, “Discriminative learning for differing training and test distributions.,” *Proceedings of the International Conference on Machine Learning* (ICML), 2007.
16. S. Bikckle, M. Brückner, and T. Scheffer, “Discriminative learning under covariate shift,” *Journal of Machine Learning Research*, vol. 10, 2009, pp. 2137–2155.
17. R. Xia et al., “Instance Selection and Instance Weighting for Cross-domain Sentiment Classification via PU Learning,” *Proceedings of the International Joint Conference on Artificial Intelligence* (IJCAI), 2013, pp. 2276–2182.
18. Y. Yang and X. Liu, “A re-examination of text categorization methods,” *Proceedings of the Annual International SIGIR conference* (SIGIR), 1999.

ABOUT THE AUTHORS

Feng Xu received a master's degree from Southeast University, China, in 2007. She is currently a PhD candidate at Nanjing University of Science and Technology's School of Economics and Management. Her research interests include social computing, data mining, and financial management. Contact her at breezewing@126.com.

Jianfei Yu received a master's degree from Nanjing University of Science and Technology in 2015. He is currently a PhD candidate at Singapore Management University's School of Information Systems. His research interests include natural language processing and machine learning. Contact him at yujianfei1990@gmail.com.

Rui Xia is a professor at Nanjing University of Science and Technology's School of Computer Science & Engineering. He received a PhD from the Chinese Academy of Sciences' Institute of Automation in 2011. His research interests include machine learning, natural language processing, text mining, and sentiment analysis. Rui Xia was the corresponding author for this piece. Contact him at rxia@njust.edu.cn.