

Working Paper SERIES

Date April 25, 2014

WP # 0005MSS-061-2014

A Multi-task Multi-kernel Transfer Learning Method for Customer Response Modeling in Social Media

Minghe Sun

*Department of Management Science and Statistics,
College of Business,
The University of Texas at San Antonio*

Zhen-Yu Chen

Zhi-Ping Fan

*Department of Management Science and Engineering,
School of Business Administration,
Northeastern University,
Shenyang 110819, China*

Copyright © 2014, by the author(s). Please do not quote, cite, or reproduce without permission from the author(s).

A Multi-task Multi-kernel Transfer Learning Method for Customer Response Modeling in Social Media

Minghe Sun^a, Zhen-Yu Chen^{b1}, Zhi-Ping Fan^b

^a*Department of Management Science and Statistics, College of Business, The University of Texas at San Antonio, San Antonio, TX 78249-0632, USA*

^b*Department of Management Science and Engineering, School of Business Administration, Northeastern University, Shenyang 110819, China*

Abstract

Customer response modeling is essential for a firm to allocate the marketing resources to active customers who have potential values. With the development of social media, customer response modeling in social media plays important roles in the firms' marketing decisions. For customer response modeling in social media, the inputs involve multiple types of data and the purposes are to identify respondents to multiple items. In this study, a multi-task multi-kernel transfer learning (MT-MKTL) method is proposed to integrate shared, task-specific and transferred features in a framework for customer response modeling in social media. A two-phase algorithm is applied to solving the MT-MKTL problem. A computational experiment is conducted on microblog data. The experimental results show that the MT-MKTL method exhibits good performance.

Keywords: Customer response modeling; Social media; Multi-task learning; Transfer learning; Multi-kernel learning

JEL Classification: C32, C38, C51, C61

1. Introduction

Customer response modeling aims at finding active customers from the customer base who will respond to a firm's marketing activities [1]. Customers are also called users and customer response modeling is also called user response modeling. It is essential for a firm to allocate the marketing resources to active customers who have potential values.

Social media, as popular communication tools, have been widely used by more and more people in the last several years. The most important characteristic of social media is that their contents are generated by

¹ Corresponding author. Tel.: +86-024-83671630:

E-mail address: minghe.sun@utsa.edu (M. Sun); zychen@mail.neu.edu.cn (Z.-Y. Chen); zpfan@mail.neu.edu.cn (Z.-P. Fan).

users themselves. Social media include social networks, blogs, microblogs, Wiki forums, content sharing, among others. Nowadays, social media have become important components of promotional mix for many firms [2]. Thus, user response modeling in social media plays important roles in the firms' marketing decisions.

Compared with traditional customer response modeling, relatively few studies have focused on user response modeling in social media. Chen *et al.* [3] developed a hierarchical ensemble learning framework to combine the longitudinal individual behavioral and customer-customer interaction, called engagement behavioral, data in customer response modeling in social media. The results show that the use of customer-customer interaction data can improve the prediction performance of the response models. Chua [4] developed two generative models to predict the missing links in the user-user social graph and item-user adoption graph using social correlation data. Tang and Liu [5] proposed the concept of social dimension to represent user latent affiliations and used social dimension to construct a classification framework. Fang *et al.* [6] developed a locally weighted expectation-maximization method for Naïve Bayesian learning to predict customer adoption probabilities in the social network.

The purposes of user response models are to identify potential respondents for multiple items. These items may be persons, products, services and events which are automatically recommended to users by the social media platform. User response modeling for multiple items in social media can be viewed as a multi-task learning problem. Multi-task learning deals with multiple tasks associated with each other simultaneously [7-9]. Multi-task learning has been successfully applied to web page classification, text categorization, image annotation, microarray and protein data classification, and so on [7, 8].

Support vector machines (SVMs) can usually have good classification performance due to margin maximization and the usage of kernels. With kernels, SVMs can construct nonlinear classification functions [3, 9, 10] in high dimensional feature spaces without actually mapping the input data from the input space into the high dimensional feature spaces. Multi-kernel learning (MKL) methods are the most popular strategies to learn the weights of a preselected set of some basic kernels [3, 9, 11]. When multiple basic kernels are used in a SVM, the SVM model is a multi-kernel SVM (MK-SVM). MKL methods can also be used for the kernel methods beyond the SVM [11]. Multi-task learning has been formulated as a multi-task MKL problem. The multi-task MKL problem can be modeled with the standard MKL formulations such as quadratically constrained quadratic programs and semi-infinite linear programs and can be solved by standard optimization algorithms [9].

The inputs to user response models involve multiple types of data which have been considered by some researchers [3-5]. User response modeling for multiple items in social media involves external, tag and keyword, individual behavioral and engagement behavioral data [3]. Among these types of data,

external data and tag and keyword data are shared among all tasks, while individual behavioral data and engagement behavioral data are task-specific. As a current topic, transfer learning aims at applying the knowledge of some source tasks to a target task [12] or at generalizing knowledge across tasks [13]. Whether or not the performance of user response modeling for a specific task can be improved using the task-specific features of other tasks, called transferred features in this study, is an interesting problem to be addressed.

In this study, a multi-task multi-kernel transfer learning (MT-MKTL) method is proposed to integrate the multiple types of input data of multiple learning tasks in a framework for user response modeling in social media. A multi-task multi-kernel transfer SVM (MT-MKT-SVM) model is developed and a two-phase method is applied to training the MT-MKT-SVM. In the first phase, multiple tasks with shared, task-specific and transferred features are modeled as standard SVMs and the SVMs are trained in parallel. In the second phase, the weights of shared, task-specific and transferred features for each task are learned by solving a linear program.

The paper is organized as follows. Section 2 gives a framework of user response modeling in social media using the MT-MKTL method. The proposed MT-MKTL method and the MT-MKT-SVM are described in Section 3. Some computational results are reported in Section 4. Conclusions are given in Section 5.

2. The Framework

The framework for user response modeling in social media using the MT-MKTL method includes the following three main components:

Multi-tasks: User response modeling in social media using the MT-MKTL method simultaneously considers multiple tasks. The tasks may be the identification of potential respondents to multiple items to be recommended to the users.

Multi-features: Each task is learned from multiple features. The features used to predict the users' responses to a task are classified into multiple categories, *i.e.*, shared features, task-specific features and transferred features. Hence, multi-task transfer learning is ensemble learning on multiple feature subspaces. The subspaces include the shared feature, task-specific feature and transferred feature subspaces.

Ensemble learning: The MKL method can be used to ensemble diverse heterogeneous features. Thus, the multi-task transfer learning mentioned above can be formulated as a MT-MKT-SVM model. In the MT-MKT-SVM, the shared features are modeled by shared multi-kernels, the task-specific features are modeled by task-specific multi-kernels, and the transferred features are modeled by transferred multi-

kernels. Thus, the two-phase training algorithm can be used to learn the weights of the shared, task-specific and transferred multi-kernels, respectively.

The MT-MKTL method will be presented in the next section in details. The framework of user response modeling in social media using the MT-MKTL method for two tasks is depicted in Fig. 1.

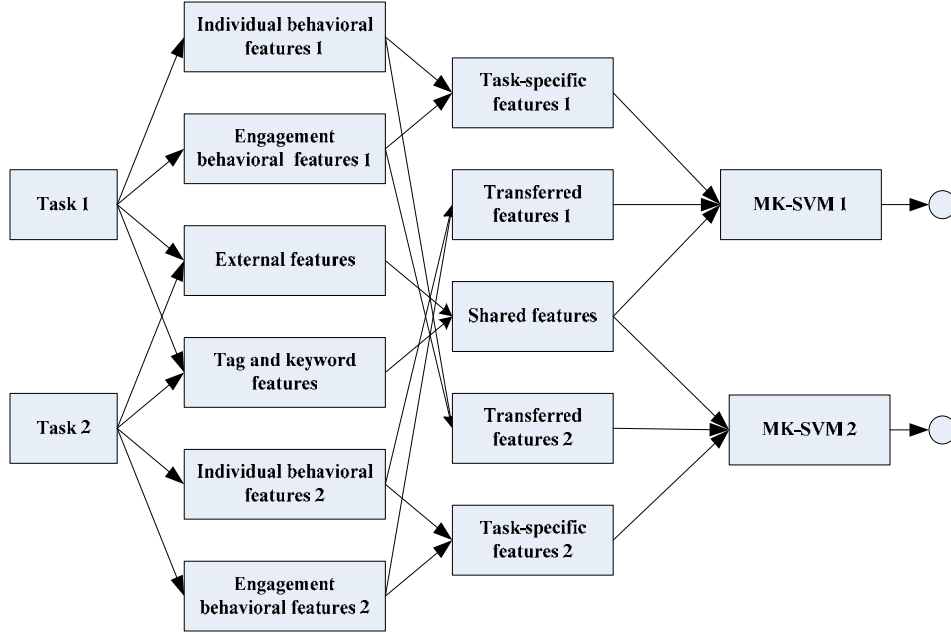


Fig. 1. The framework of user response modeling in social media using the MT-MKTL method for two tasks

3. The Method

In this section, the MKL method is briefly discussed, and the MT-MKTL method is then given in details. The number of tasks is represented by Q and the number of observations in the training dataset is represented by n .

3.1. The MKL method

For a binary classification problem, *i.e.*, when $Q = 2$, the training dataset is represented by $G = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$, where $\mathbf{x}_i \in \mathcal{R}^m$, $y_i \in \{0, 1\}$ and m is the number of features in the input data. A multi-kernel $K(\mathbf{x}_i, \mathbf{x}_j)$ is a linear combination of P basic kernels $k_p(\mathbf{x}_i, \mathbf{x}_j)$ for $p = 1, \dots, P$. Let $\phi_p(\mathbf{x}) : \mathcal{R}^m \mapsto \mathcal{R}^{\tilde{m}_p}$ with $m \ll \tilde{m}_p$ be the nonlinear map for the basic kernel $k_p(\mathbf{x}_i, \mathbf{x}_j)$. The nonlinear map

$\phi_p(\mathbf{x}_i)$ maps the input $\mathbf{x}_i \in \mathfrak{R}^m$ onto $\phi_p(\mathbf{x}_i) \in \mathfrak{R}^{\tilde{m}_p}$ in a high-dimensional feature space and \tilde{m}_p is the dimension of the p th feature space. When kernels are used, the mappings are not actually carried out and the functional forms of the mappings are not necessarily known. In this study the Gaussian kernel [10], also called the radial basis function (RBF) kernel, is used as the basic kernels.

The purpose of the MKL method is to construct a classification function of the form

$$f(\mathbf{x}_{i_0}) = \sum_{p=1}^P \mathbf{w}_p^T \phi_p(\mathbf{x}_{i_0}) + b \quad (1)$$

for any observation i_0 with an input $\mathbf{x}_{i_0} \in \mathfrak{R}^m$ by training a MK-SVM, where \mathbf{w}_p is the vector of weights and b is the bias. Typically, the following formulation of the MK-SVM is considered:

$$\min_{\boldsymbol{\beta}} \min_{\mathbf{w}_p, \boldsymbol{\xi}, b} \frac{1}{2} \sum_{p=1}^P \beta_p \|\mathbf{w}_p\|^2 + C \sum_{i=1}^n \xi_i \quad (2)$$

$$\text{s.t.} \quad y_i \left(\sum_{p=1}^P \mathbf{w}_p^T \phi_p(\mathbf{x}_i) + b \right) \geq 1 - \xi_i, \quad i = 1, \dots, n \quad (3)$$

$$\xi_i \geq 0, \quad i = 1, \dots, n \quad (4)$$

$$\beta_p \geq 0, \quad p = 1, \dots, P, \quad (5)$$

where C is the regularization parameter, ξ_i is the error term for observation i , $\boldsymbol{\xi}$ is the vector of ξ_i for $i = 1, \dots, n$, β_p is the weight of the basic kernel $k_p(\mathbf{x}_i, \mathbf{x}_j)$ and $\boldsymbol{\beta}$ is the vector of β_p for $p = 1, \dots, P$.

In the last decade, many methods have been proposed to solve the MKL problem in (2)-(5). For the MT-MKTL method, the two-phase method proposed in [12] is used to solve the sub-problems corresponding to the specific tasks.

3.2. The MT-MKTL method

For user response modeling in social media, there are four types of data, *i.e.*, external, tag and keyword, individual behavioral and engagement behavioral data. The number of observations in the training dataset n is the number of users or customers. The label of observation i for task q in the training dataset is represented by $y_i^q \in \{0, 1\}$ and the vector of the labels of all n customers for task q is represented by \mathbf{y}^q for $q = 1, \dots, Q$. The external data are represented by $\mathbf{S} = \{s_{ij} \mid i = 1, \dots, n; j = 1, \dots, m_1\}$, where m_1 is the number of features of external data. The external features of observation i are represented by $\mathbf{s}_i = \{s_{ij} \mid j = 1, \dots, m_1\}$. The tag and keyword data are represented by $\hat{\mathbf{S}} = \{\hat{s}_{ij} \mid i = 1, \dots, n; \hat{j} = 1, \dots, m_2\}$, where m_2 is the number of features of tag and keyword data. The tag and keyword features of observation

i are represented by $\hat{\mathbf{s}}_i = \{\hat{s}_{ij} | j=1, \dots, m_2\}$. The individual behavioral data are represented by $\mathbf{B} = \{b_{ij}^q | i=1, \dots, n; j=1, \dots, m_3; t=1, \dots, T; q=1, \dots, Q\}$, where m_3 is the number of features of individual behavioral data and T is the length of each feature of the individual behavioral data. The individual behavioral features for task q of observation i are represented by $\mathbf{B}_i^q = \{b_{ij}^q | j=1, \dots, m_3; t=1, \dots, T\}$ and the \tilde{j} th individual behavioral feature for task q of observation i is represented by $\mathbf{b}_{ij}^q = \{b_{ij}^q | t=1, \dots, T\}$. The engagement behavioral data are the average individual behavioral data of the followers of the customers. The engagement behavioral data are represented by $\hat{\mathbf{B}} = \{\hat{b}_{ij'}^q | i=1, \dots, n; j'=1, \dots, m_4; t'=1, \dots, T'; q=1, \dots, Q\}$, where m_4 is the number of features of the engagement behavioral data and T' is the length of each feature of the engagement behavioral data. The features of engagement behavioral data for task q of observation i are represented by $\hat{\mathbf{B}}_i^q = \{\hat{b}_{ij'}^q | j'=1, \dots, m_4; t'=1, \dots, T'\}$ and the j' th feature of the engagement behavioral data for task q of observation i is represented by $\hat{\mathbf{b}}_{ij'}^q = \{\hat{b}_{ij'}^q | t'=1, \dots, T'\}$. Individual and engagement behavioral features for task q are the transferred features when they are used for predicting the users' responses for a different task q' .

For learning Q binary classification tasks represented by the labels \mathbf{y}^q for $q=1, \dots, Q$, the input data in the training dataset $(\mathbf{s}_i, \hat{\mathbf{s}}_i, \mathbf{B}_i^q, \hat{\mathbf{B}}_i^q, \mathbf{y}_i^q)$ for $i=1, \dots, n$ are available. Among the input features, the external features \mathbf{s}_i and tag and keyword features $\hat{\mathbf{s}}_i$ are shared features, and the individual behavioral features \mathbf{B}_i^q and engagement behavioral features $\hat{\mathbf{B}}_i^q$ are task-specific features for task q . However, the individual behavioral features \mathbf{B}_i^q and engagement behavioral features $\hat{\mathbf{B}}_i^q$ are transferred features for a different task q' .

The purpose of the MT-MKTL method for task q is to construct a classification function of the form

$$f(\mathbf{s}_{i_0}, \hat{\mathbf{s}}_{i_0}, \mathbf{B}_{i_0}^q, \hat{\mathbf{B}}_{i_0}^q) = \hat{\mathbf{w}}_1^T \phi_1(\mathbf{s}_{i_0}) + \hat{\mathbf{w}}_2^T \phi_2(\hat{\mathbf{s}}_{i_0}) + \sum_{q'=1}^Q \sum_{j=1}^{m_3} \tilde{\mathbf{w}}_j^{(T)q'} \tilde{\phi}_j(\mathbf{B}_{i_0}^{q'}) + \sum_{q'=1}^Q \sum_{j'=1}^{m_4} \hat{\mathbf{w}}_{j'}^{(T)q'} \hat{\phi}_{j'}(\hat{\mathbf{B}}_{i_0}^{q'}) + \hat{b} \quad (6)$$

for any observation i_0 with input $(\mathbf{s}_{i_0}, \hat{\mathbf{s}}_{i_0}, \mathbf{B}_{i_0}^q, \hat{\mathbf{B}}_{i_0}^q)$ by training a MT-MKT-SVM. In (6), $\hat{\mathbf{w}}_1$ is the vector of weights of the external data, $\hat{\mathbf{w}}_2$ is the vector of weights of the tag and keyword data, $\tilde{\mathbf{w}}_j^{q'}$ is the vector of weights of the \tilde{j} th feature of the individual behavioral data with task q' , $\hat{\mathbf{w}}_{j'}^{q'}$ is the vector of weights of the j' th feature of the engagement behavioral data with task q' , and \hat{b} is the bias. Also in (6), $\phi_1(\mathbf{s})$,

$\phi_2(\hat{\mathbf{s}})$, $\tilde{\phi}_j(\mathbf{B}_i^{q'})$ and $\hat{\phi}_j(\hat{\mathbf{B}}_i^{q'})$ are the nonlinear mappings. Using the shared, task-specific and transferred features, the MT-MKT-SVM for task q is formulated as

$$\min_{\hat{\boldsymbol{\beta}}} \min_{\hat{\mathbf{w}}_1, \hat{\mathbf{w}}_2, \tilde{\mathbf{w}}_j^{q'}, \hat{\mathbf{w}}_j^{q'}, \hat{\boldsymbol{\xi}}, \hat{b}} \frac{1}{2} \hat{\beta}_1 \|\hat{\mathbf{w}}_1\|^2 + \frac{1}{2} \hat{\beta}_2 \|\hat{\mathbf{w}}_2\|^2 + \frac{1}{2} \sum_{q'=1}^Q \sum_{j=1}^{m_3} \tilde{\gamma}_j^{q'} \|\tilde{\mathbf{w}}_j^{q'}\|^2 + \frac{1}{2} \sum_{q'=1}^Q \sum_{j=1}^{m_4} \hat{\gamma}_j^{q'} \|\hat{\mathbf{w}}_j^{q'}\|^2 + \tilde{C} \sum_{i=1}^n \hat{\xi}_i \quad (7)$$

$$\text{s.t.} \quad y_i^q \left(\hat{\mathbf{w}}_1^T \phi_1(\mathbf{s}_i) + \hat{\mathbf{w}}_2^T \phi_2(\hat{\mathbf{s}}_i) + \sum_{q'=1}^Q \sum_{j=1}^{m_3} \tilde{\mathbf{w}}_j^{(T)q'} \tilde{\phi}_j(\mathbf{B}_i^{q'}) + \sum_{q'=1}^Q \sum_{j=1}^{m_4} \hat{\mathbf{w}}_j^{(T)q'} \hat{\phi}_j(\hat{\mathbf{B}}_i^{q'}) + \hat{b} \right) \geq 1 - \hat{\xi}_i, \quad i=1, \dots, n \quad (8)$$

$$\hat{\xi}_i \geq 0, \quad i=1, \dots, n \quad (9)$$

$$\hat{\beta}_1 \geq 0, \quad \hat{\beta}_2 \geq 0 \quad (10)$$

$$\tilde{\gamma}_j^{q'} \geq 0, \quad \tilde{j}=1, \dots, m_3, \quad q'=1, \dots, Q \quad (11)$$

$$\hat{\gamma}_j^{q'} \geq 0, \quad j'=1, \dots, m_4, \quad q'=1, \dots, Q, \quad (12)$$

where \tilde{C} is the regularization parameter, $\hat{\beta}_1$ is the weight of $\hat{\mathbf{w}}_1$, $\hat{\beta}_2$ is the weight of $\hat{\mathbf{w}}_2$, $\tilde{\gamma}_j^{q'}$ is the weight of $\tilde{\mathbf{w}}_j^{q'}$, $\hat{\gamma}_j^{q'}$ is the weight of $\hat{\mathbf{w}}_j^{q'}$, $\tilde{\boldsymbol{\gamma}}$ is the vector of $\tilde{\gamma}_j^{q'}$ for $\tilde{j}=1, \dots, m_3$, $\hat{\boldsymbol{\gamma}}$ is the vector of $\hat{\gamma}_j^{q'}$ for $j'=1, \dots, m_4$, $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, \tilde{\boldsymbol{\gamma}}, \hat{\boldsymbol{\gamma}})$ is a composite vector, $\hat{\xi}_i$ is the error term for observation i and $\hat{\boldsymbol{\xi}}$ is the vector of $\hat{\xi}_i$ for $i=1, \dots, n$.

A two-phase algorithm is proposed to solve the problem in (7)-(12). In the first phase of the two-phase algorithm, the estimated parameters $\hat{\boldsymbol{\beta}}$ are fixed and the dual of the q th MKL problem is solved for all $q=1, \dots, Q$. The dual of the q th MKL problem is stated as follows

$$\max_{\boldsymbol{\alpha}^q} \sum_{i=1}^n \alpha_i^q - \frac{1}{2} \sum_{i=1}^n \sum_{i'=1}^n \alpha_i^q \alpha_{i'}^q y_i^q y_{i'}^q \left(\beta_1 k_1(\mathbf{s}_i, \mathbf{s}_{i'}) + \beta_2 k_2(\hat{\mathbf{s}}_i, \hat{\mathbf{s}}_{i'}) + \sum_{q'=1}^Q \sum_{j=1}^{m_3} \tilde{\gamma}_j^{q'} k_3^{q'}(\mathbf{b}_{ij}^{q'}, \mathbf{b}_{i'j}^{q'}) + \sum_{q'=1}^Q \sum_{j=1}^{m_4} \hat{\gamma}_j^{q'} k_4^{q'}(\hat{\mathbf{b}}_{ij}^{q'}, \hat{\mathbf{b}}_{i'j}^{q'}) \right) \quad (13)$$

$$\text{s.t.} \quad \sum_{i=1}^n y_i^q \alpha_i^q = 0 \quad (14)$$

$$0 \leq \alpha_i^q \leq \tilde{C}, \quad i=1, \dots, n, \quad (15)$$

where α_i^q is the Lagrange multiplier of observation i for task q , $\boldsymbol{\alpha}^q$ is the vector of α_i^q for $i=1, \dots, n$, $k_1(\mathbf{s}_i, \mathbf{s}_{i'})$ is the basic kernel for the external data, $k_2(\hat{\mathbf{s}}_i, \hat{\mathbf{s}}_{i'})$ is the basic kernel for the tag and keyword data, $k_3^{q'}(\mathbf{b}_{ij}^{q'}, \mathbf{b}_{i'j}^{q'})$ is the basic kernel of the \tilde{j} th feature of the individual behavioral data for task q' , and $k_4^{q'}(\hat{\mathbf{b}}_{ij}^{q'}, \hat{\mathbf{b}}_{i'j}^{q'})$ is the basic kernel of the j' th feature of the engagement behavioral data for task q' .

In the second phase of the two-phase algorithm, the Lagrange multipliers $\boldsymbol{\alpha}^q$ for all $q=1, \dots, Q$ are fixed, the problem in (13)-(15) is transformed into a linear programming problem with all components of $\hat{\boldsymbol{\beta}}$ as the variables

$$\min_{\hat{\beta}, \tilde{\xi}} \frac{1}{2} \sum_{i=1}^n \sum_{i=1}^n \alpha_i^q \alpha_i^q y_i^q y_i^q \left(\beta_1 k_1(\mathbf{s}_i, \mathbf{s}_i) + \beta_2 k_2(\hat{\mathbf{s}}_i, \hat{\mathbf{s}}_i) + \sum_{q'=1}^Q \sum_{j=1}^{m_3} \tilde{\gamma}_j^{q'} k_3^{q'}(\mathbf{b}_{ij}^{q'}, \mathbf{b}_{ij}^{q'}) + \sum_{q'=1}^Q \sum_{j'=1}^{m_4} \hat{\gamma}_{j'}^{q'} k_4^{q'}(\hat{\mathbf{b}}_{ij'}^{q'}, \hat{\mathbf{b}}_{ij'}^{q'}) \right) + \hat{C} \sum_{i=1}^n \tilde{\xi}_i \quad (16)$$

$$\text{s.t. } y_i^q \left\{ \sum_{i=1}^n \alpha_i^q y_i^q \left(\beta_1 k_1(\mathbf{s}_i, \mathbf{s}_i) + \beta_2 k_2(\hat{\mathbf{s}}_i, \hat{\mathbf{s}}_i) + \sum_{q'=1}^Q \sum_{j=1}^{m_3} \tilde{\gamma}_j^{q'} k_3^{q'}(\mathbf{b}_{ij}^{q'}, \mathbf{b}_{ij}^{q'}) + \sum_{q'=1}^Q \sum_{j'=1}^{m_4} \hat{\gamma}_{j'}^{q'} k_4^{q'}(\hat{\mathbf{b}}_{ij'}^{q'}, \hat{\mathbf{b}}_{ij'}^{q'}) \right) + \hat{b} \right\} \geq 1 - \tilde{\xi}_i \quad (17)$$

$$\tilde{\xi}_i \geq 0, \quad i = 1, \dots, n \quad (18)$$

$$\hat{\beta}_1 \geq 0, \quad \hat{\beta}_2 \geq 0 \quad (19)$$

$$\tilde{\gamma}_j^{q'} \geq 0, \quad j = 1, \dots, m_3, \quad q' = 1, \dots, Q \quad (20)$$

$$\hat{\gamma}_{j'}^{q'} \geq 0, \quad j' = 1, \dots, m_4, \quad q' = 1, \dots, Q, \quad (21)$$

where \hat{C} is the regularization parameter, $\tilde{\xi}_i$ is the error term for observation i and $\tilde{\xi}$ is the vector of $\tilde{\xi}_i$ for $i = 1, \dots, n$. The dual of the model in (16)-(21) is easier to solve than the primal. The dual is stated as

$$\max_{\mathbf{u}^q} \sum_{i=1}^n u_i^q \quad (22)$$

$$\text{s.t. } \sum_{i=1}^n y_i^q u_i^q \sum_{i=1}^n \alpha_i^q y_i^q k_1(\mathbf{s}_i, \mathbf{s}_i) \leq \frac{1}{2} \sum_{i=1}^n \sum_{i=1}^n \alpha_i^q \alpha_i^q y_i^q y_i^q k_1(\mathbf{s}_i, \mathbf{s}_i) \quad (23)$$

$$\sum_{i=1}^n y_i^q u_i^q \sum_{i=1}^n \alpha_i^q y_i^q k_2(\hat{\mathbf{s}}_i, \hat{\mathbf{s}}_i) \leq \frac{1}{2} \sum_{i=1}^n \sum_{i=1}^n \alpha_i^q \alpha_i^q y_i^q y_i^q k_2(\hat{\mathbf{s}}_i, \hat{\mathbf{s}}_i) \quad (24)$$

$$\sum_{i=1}^n y_i^q u_i^q \sum_{i=1}^n \alpha_i^q y_i^q k_3^{q'}(\mathbf{b}_{ij}^{q'}, \mathbf{b}_{ij}^{q'}) \leq \frac{1}{2} \sum_{i=1}^n \sum_{i=1}^n \alpha_i^q \alpha_i^q y_i^q y_i^q k_3^{q'}(\mathbf{b}_{ij}^{q'}, \mathbf{b}_{ij}^{q'}), \quad q' = 1, \dots, Q, \quad j = 1, \dots, m_3 \quad (25)$$

$$\sum_{i=1}^n y_i^q u_i^q \sum_{i=1}^n \alpha_i^q y_i^q k_4^{q'}(\hat{\mathbf{b}}_{ij'}^{q'}, \hat{\mathbf{b}}_{ij'}^{q'}) \leq \frac{1}{2} \sum_{i=1}^n \sum_{i=1}^n \alpha_i^q \alpha_i^q y_i^q y_i^q k_4^{q'}(\hat{\mathbf{b}}_{ij'}^{q'}, \hat{\mathbf{b}}_{ij'}^{q'}), \quad q' = 1, \dots, Q, \quad j' = 1, \dots, m_4 \quad (26)$$

$$\sum_{i=1}^n y_i^q u_i^q = 0 \quad (27)$$

$$0 \leq u_i^q \leq \hat{C}, \quad i = 1, \dots, n, \quad (28)$$

where u_i^q is the dual variable for observation i and \mathbf{u}^q is the vector of u_i^q for $i = 1, \dots, n$.

The classification function (6) with the dual variables has the following form

$$f(\mathbf{s}_{i_0}, \hat{\mathbf{s}}_{i_0}, \mathbf{B}_{i_0}^q, \hat{\mathbf{B}}_{i_0}^q) = \sum_{i=1}^n \alpha_i^q y_i^q \left(\beta_1 k_1(\mathbf{s}_{i_0}, \mathbf{s}_i) + \beta_2 k_2(\hat{\mathbf{s}}_{i_0}, \hat{\mathbf{s}}_i) + \sum_{q'=1}^Q \sum_{j=1}^{m_3} \tilde{\gamma}_j^{q'} k_3^{q'}(\mathbf{b}_{i_0 j}^{q'}, \mathbf{b}_{ij}^{q'}) + \sum_{q'=1}^Q \sum_{j'=1}^{m_4} \hat{\gamma}_{j'}^{q'} k_4^{q'}(\hat{\mathbf{b}}_{i_0 j'}^{q'}, \hat{\mathbf{b}}_{ij'}^{q'}) \right) + \hat{b} \quad (29)$$

for any observation i_0 with input $(\mathbf{s}_{i_0}, \hat{\mathbf{s}}_{i_0}, \mathbf{B}_{i_0}^q, \hat{\mathbf{B}}_{i_0}^q)$.

The input $(\mathbf{s}_i, \hat{\mathbf{s}}_i, \mathbf{B}_i^q, \hat{\mathbf{B}}_i^q)$ of an observation i such that $0 < \alpha_i^q < \tilde{C}$ is a support vector. The bias \hat{b} in (6) and (29) can be determined by (30) in the following using any support vector i ,

$$\hat{b} = y_i^q - \sum_{i=1}^n \alpha_i^q y_i^q \left(\beta_1 k_1(\mathbf{s}_i, \mathbf{s}_i) + \beta_2 k_2(\hat{\mathbf{s}}_i, \hat{\mathbf{s}}_i) + \sum_{q'=1}^Q \sum_{j=1}^{m_3} \tilde{\gamma}_j^{q'} k_3^{q'}(\mathbf{b}_{ij}^{q'}, \mathbf{b}_{ij}^{q'}) + \sum_{q'=1}^Q \sum_{j'=1}^{m_4} \hat{\gamma}_{j'}^{q'} k_4^{q'}(\hat{\mathbf{b}}_{ij'}^{q'}, \hat{\mathbf{b}}_{ij'}^{q'}) \right). \quad (30)$$

4. Computational Experiments

In this section, the dataset used in the experiment is briefly discussed and the experimental results are then reported.

4.1. The data

The microblog data² provided by Tencent Weibo and used in KDD Cup 2012 are used to examine the performance of the proposed MT-MKTL method. The original data are preprocessed by the proposed method in [3]. The transformed data used in the experiment include four datasets, *i.e.*, external, tag and keyword, individual behavioral and engagement behavioral datasets. There are $m_1=2$, $m_2=10$, $m_3=2$ and $m_4=2$ independent features, respectively, in these datasets which are the same as those used in [3]. Two tasks, *i.e.*, Item 1.1.2.2 and Item 1.1.2.5, are used to evaluate the performance of the proposed MT-MKTL method. Each dataset is divided into three subsets, *i.e.*, training set, validation set and testing set. The validation set is used to guide the selection of the kernel parameters while the training set is used to train the MT-MKT-SVM. The computational results on the testing set are reported in the following.

4.2. Experimental results

In the experiment, three scenarios are considered: (s1) the external, tag, individual behavioral and engagement behavioral features of Task A and all two individual behavioral features of Task B are used to predict the users' responses to Task A; (s2) the external, tag, individual behavioral and engagement behavioral features of Task A and one individual behavioral feature, *i.e.*, Acceptance [3], of Task B are used to predict the users' responses to Task A; (s3) only the external, tag, individual behavioral and engagement behavioral features of Task A are used to predict the users' responses to Task A. The scenarios s1 and s2 use transfer learning, *i.e.*, use the transferred features of Task B to predict the users' responses to Task A, while the scenario s3 does not use transfer learning, *i.e.*, predicts the users' responses to Task A without using transferred features of Task B.

Six measures including the overall hit rate (PCC), the hit rate of the positive class (Sensitivity), the hit rate of the negative class (Specificity), the area under the receiver operating characteristic curve (AUC), the top 10% lift (Lift) and the maximum profit (MP) are used to evaluate the performance of the MT-MKTL method on the scenarios mentioned above [3]. Results of the MT-MKTL method on Items 1.1.2.2 and 1.1.2.5 with and without transfer learning are reported in Tables 1 and 2, respectively.

² <http://kddcup2012.org/c/kddcup2012-track1/data>.

As shown in Tables 1 and 2, the MT-MKTL method under S1 obtained the highest PCC, Specificity, AUC, Lift and MP on Item 1.1.2.2 and the highest Sensitivity, AUC, Lift and MP on Item 1.1.2.5. These results show that the use of the transferred features can improve the classification performance.

Table 1. Results of the MT-MKTL method on Item 1.1.2.2 with and without transfer learning

Methods	PCC	Sensitivity	Specificity	AUC	Lift	MP
MT-MKTL(S1)	86.24	36.07	88.90	73.36	3.12	2.14
MT-MKTL(S2)	77.25	47.54	78.83	67.50	2.87	1.89
MT-MKTL(S3)	7.94	95.49	3.30	67.31	2.87	1.89

Table 2. Results of the MT-MKTL method on Item 1.1.2.5 with and without transfer learning

Methods	PCC	Sensitivity	Specificity	AUC	Lift	MP
MT-MKTL(S1)	87.15	28.16	90.29	59.54	2.58	1.60
MT-MKTL(S2)	87.09	25.71	90.36	54.40	2.43	1.48
MT-MKTL(S3)	92.08	17.14	96.07	53.61	1.72	0.74

5. Conclusions

In this study, a MT-MKTL method is proposed to integrate multiple types of input data of multiple learning tasks in a framework for user response modeling in social media. A two-phase method is applied to solving the MT-MKTL problem. The advantage of the proposed MT-MKTL method is that the integration of transferred features can improve the classification performance.

A computational experiment is conducted on the Tencent Weibo data. The experimental results show that the MT-MKTL method using the transferred features obtained the highest AUC, Lift and MP on two tasks used in the experiment.

How to identify relative from a large number of tasks and select the relevant transferred features to improve the classification performance will be studied as future works. Unlike the engagement behavioral data used in this study, the engagement behavioral data of selected neighbors of an item may also be used as task-specific features and transferred features to learn the responses of the users to the item and to its related items.

Acknowledgements

This work was partially supported by the National Natural Science Foundation of China (Project No. 71101023, 71021061 and 71271051) and the Fundamental Research Funds for the Central Universities, NEU, China (Project No. N120406001 and N110706001).

References

- [1] Schweidel DA, Knox G. Incorporating directmarketing activity into latent attrition models. *Marketing Science* 2013; 32(3):471–487.
- [2] Mangold WG, Faulds DJ. Social media: The new hybrid element of the promotion mix. *Business Horizons* 2009; 52:357–365.
- [3] Chen ZY, Fan ZP, Sun M. Behavior-aware user response modeling in social media: Learning from diverse heterogeneous data. Working paper 2013; The University of Texas at San Antonio.
- [4] Chua FCT, Lauw HW, Lim EP. Generative models for item adoptions using social correlation. *IEEE Transactions on Knowledge and Data Engineering* 2013; 25(9):2036–2048.
- [5] Tang L, Liu H. Leveraging social media networks for classification. *Data Mining and Knowledge Discovery* 2011; 23:447–478.
- [6] Fang X, Hu PJH, Li Z, Tsai W. Predicting adoption probabilities in social networks. *Information Systems Research* 2013; 24(1):128–145.
- [7] Tang L, Chen J, Ye J. On multiple kernel learning with multiple labels. Proceedings of the Twenty-one International Joint Conference on Artificial Intelligence, 2009.
- [8] Zhang W, Xue X, Fan J, Huang X, Wu B, Liu M. Multi-kernel multi-label learning with max-margin concept network. Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, 2012; 1615-1620.
- [9] Li C, Georgiopoulos M, Anagnostopoulos GC. A unifying framework for typical multi-task multiple kernel learning problems. *IEEE Transactions on Neural Network and Learning Systems* 2014; http://www.eecs.ucf.edu/~congli/papers/TNNLS2013_Cong_Li.pdf.
- [10] Vapnik, VN. *Statistic learning theory*. 1998; Wiley, New York.
- [11] Bach FR, Lanckrient, GRG, Jordan, MI. Multiple kernel learning, conic duality and the SMO algorithm. In: G. Russell, S. Dale (Eds.), Proceedings of the Twenty First International Conference on Machine Learning, 2004; 41–48.
- [12] Pan SJ, Yang Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 2010; 22(10):1345–1359.

- [13]Tylor ME, Stone P. Transfer Learning for Reinforcement Learning Domains: A Survey. *Journal of Machine Learning Research*, 2009; 10:1633–1685.
- [14] Chen ZY, Li JP, Wei LW. A multiple kernel support vector machine scheme for feature selection and rule extraction from gene expression data of cancer tissue. *Artificial Intelligence in Medicine 2007*; 41(2):161–175.