THE UNIVERSITY OF TEXAS AT SAN ANTONIO, COLLEGE OF BUSINESS

Working Paper SERIES

Date May 3, 2013

WP # 0033MSS-061-2013

Behavior-aware User Response Modeling in Social Media: Learning from Diverse Heterogeneous Data

Zhen-Yu Chen
Department of Management Science and Engineering
Northeastern University
Shenyang China

Zhi-Ping Fan
Department of Management Science and Engineering
Northeastern University
Shenyang China

Minghe Sun
Department of Management Science and Statistics
The University of Texas at San Antonio

UTSA COLLEGE OF BUSINESS KNOWLEDGE FOR A NEW WORLD

ONE UTSA CIRCLE
SAN ANTONIO, TEXAS 78249-0631
210 458-4317 | BUSINESS.UTSA.EDU

Behavior-aware User Response Modeling in Social Media: Learning from Diverse Heterogeneous Data

Zhen-Yu Chen[a], Zhi-Ping Fan[a], Minghe Sun[b]

[a]*Department of Management Science and Engineering, School of Business Administration,*

Northeastern University, Shenyang 110819, China

[b]*Department of Management Science and Statistics, College of Business,*

The University of Texas at San Antonio, San Antonio, TX 78249-0632, USA

Abstract:

With the rapid development of Web 2.0 applications, social media have increasingly become a major factor influencing the purchase decisions of customers. Massive user behavioral, *i.e.*, longitudinal individual behavioral and engagement behavioral, data generated on social media sites post challenges to integrate diverse heterogeneous data to improve prediction performance in customer response modeling. In this study, a hierarchical ensemble learning framework is proposed for behavior-aware user response modeling using diverse heterogeneous data. In the framework, a general-purpose data preprocessing and transformation strategy is developed to transform the large-scale and multi-relational datasets into customer-centered external and behavioral datasets and to extract prediction attributes. An improved hierarchical multiple kernel support vector machine (H-MK-SVM) is developed to integrate the external, tag and keyword, individual behavioral and engagement behavioral data for feature selection from multiple correlated attributes and for ensemble learning in user response modeling. Computational experiments using a real-world microblog database were conducted to investigate the benefits of integrating diverse heterogeneous data. Computational results show that the H-MK-SVM using longitudinal individual behavioral data exhibits superior performance over some commonly used methods using aggregated behavioral data and the H-MK-SVM using engagement behavioral data exhibits superior performance over the methods using only the external and individual behavioral data.

Keywords: Data mining; Direct marketing; Response modeling; Social media; Engagement behavior; Support vector machine; Multiple kernel learning

JEL Classification: C32, C38, C51, C61

1. Introduction

Mass marketing and direct marketing are two commonly used approaches for product (service) advertising and promotional activities (Bose and Chen, 2009). For direct marketing, a marketing message is delivered to target customers without an intermediary person or indirect media involved (Bose and Chen, 2009). Customer response modeling aims at identifying the target customers who will respond to a specific marketing campaign from the existing customer base (Cui *et al.*, 2010; Kang *et al.*, 2012). With more and more companies adopting direct marketing, customer response modeling has become one of the most effective direct marketing strategies to increase total revenue and lower the marketing cost (Cui *et al.*, 2006; Kang *et al.*, 2012; Lee *et al.*, 2010).

For customer response modeling, external and behavioral data are usually used to predict the likely respondents and non-respondents (Bose and Chen, 2009; Lee *et al.*, 2010). Therefore, customer response modeling is a binary classification problem and many supervised and semi-supervised machine learning techniques have been used to solve this problem (Lessmann and Voß, 2008). These techniques include artificial neural networks (ANN) (Crone *et al.*, 2006; Kaefer *et al.*, 2005; Kim *et al.*, 2005), decision trees (Crone *et al.*, 2006), Bayesian networks (Baesens *et al*., 2002; Cui *et al.*, 2006), logistic regression (Kang *et al.*, 2012), bagging (Ha *et al*., 2005), support vector machines (SVM) (Crone *et al.*, 2006; Kang *et al.*, 2012; Lessmann and Voß, 2009; Shin and Cho, 2006) and transductive SVMs (Lee *et al.*, 2010). Moreover, some other techniques including clustering (Kang *et al.*, 2012), sampling (Crone *et al.*, 2006; Kang *et al.*, 2012), sequential pattern discovery (Chen *et al.*, 2011), feature selection (Cui *et al.*, 2010) and other preprocessing methods (Crone *et al.*, 2006) have been combined with classification techniques to refine the customer base and improve prediction accuracy.

In the age of Web 2.0, social media sites develop rapidly. Social media refers to a group of online applications which allows the creation and exchange of user-generated contents (Kaplan and Haenlein, 2010). The most popular types of social media include wikis, blogs, microblogs, social networks, video and photo sharing and online communities. They become popular communication tools due in part to the open accessibility of the tools and the fast social interactions among users. Social media have increasingly become a major factor influencing the opinions, attitudes and the purchase behavior of customers (Mangold and Faulds, 2011).

User behavioral data generated and collected on social media sites include two categories, *i.e*., individual behavioral data and engagement behavioral data. Moreover, according to the ways of using the behavioral data in the prediction models, user behavioral data can be classified as the longitudinal behavioral data and aggregated behavioral data. Figure 1 illustrates the different types of behavioral data. For traditional customer response modeling, the longitudinal individual behavioral variables derived from the transactional databases are usually transformed into the aggregated variables such as recency, frequency and monetary (RFM) variables which have been included in most direct marketing datasets and adopted in most response models (Baesens *et al*., 2002; Crone *et al.*, 2006; Cui *et al.*, 2010).

Figure 1 approximately here

In comparison with individual behavior, customer engagement behavior, as an emerging concept, focuses on the customers' behavioral manifestation beyond purchase such as electronic word-of-mouth (EWOM), customer-customer interaction, recommendations, blogging and online reviews (van Doorn *et al*., 2010). In social media, customer engagement behavior has great effect on the individual purchase decisions (Cheung and Thadani, 2012; Dellarocas, 2003; van Doorn *et al*., 2010). For example, Dell gained high income by posting offers to its followers on Twitter (Li and Li, 2013). A survey showed that 91% of respondents said that they consulted online reviews before purchasing, and 46% of respondents believed that the online reviews influenced their purchase decisions (Cheung and Thadani, 2012). Therefore, incorporating engagement behavioral data into customer analytical models is increasingly recognized as a new direction of customer relationship management (CRM) and direct marketing (Bijmolt *et al*., 2010).

The aggregated individual behavioral attributes are usually used as predictors in most response models. Few existing studies of customer response modeling pay attention to the engagement behavioral data and longitudinal individual behavioral data which are widely available in the social media databases. In recent years, the analysis of engagement behavior has been used widely in the areas of recommendation and customer churn prediction. Some researchers extended the factorization model to predict the top-N items the customer was most likely to follow using the aggregated customer-customer interaction data (Chen *et al.*, 2013; Chen, Liu *et al.*, 2012). The information of individual customers and a group of customers which have similar characteristics was used in a novel customer profile model for product recommendation (Park and Chang, 2009). For CRM of the telecommunication industry, the customer-customer interaction data have been recognized as important complements to traditional behavioral data. The aggregated engagement behavioral attributes were combined with traditional attributes to predict customer churn (Zhang *et al.*, 2012).

Some researchers recognized that customer purchase behavior varies over time and the use of the longitudinal individual behavioral data can improve prediction performance (Chen, Fan and Sun, 2012; Liu *et al.*, 2009). Sequential pattern analysis was combined with collaborative filtering for temporal purchase behavioral data to improve recommendation performance (Cho *et al.*, 2005; Choi *et al.*, 2012; Huang and Huang, 2009; Liu *et al.*, 2009; Min and Han, 2005). Prinzie and Van den Poel (2006, 2007, 2011) incorporated customer purchase sequence into dynamic Bayesian networks and Markov models to predict the next product for a customer to buy. Ballings and Van den Poel (2012) studied the problem of how long the customer historical data should be for customer churn prediction. They suggested that selecting a good length of historical data can decrease computational burden.

For social media, the term Item may represent a specific user, organization, product (service) or event such as the appearance of a new term or keyword, the announcement of a new product (service) or activity, or a new price of an existing product (service). The rich behavioral data generated on social media sites can be used for managers to predict user responses to an Item, make marketing policies and allocate marketing resources to influence customer behavior (Daniel and Gloria, 2011). For social media, customer response modeling is also called user response modeling, and the two terms are used interchangeably. In this study, customer response modeling taking into consideration the user behavioral, *e.g.*, longitudinal individual and engagement behavioral, data is called behavior-aware user response modeling. However, the large, diverse and heterogeneous data generated on social media sites bring great challenges on behavior-aware user response modeling (Bijmolt *et al*., 2010; Cao *et al.*, 2012; Chau and Xu, 2012).

How to deal with diverse heterogeneous data is the first challenge. A variety of methods can be used for customer response modeling using external and aggregated individual behavioral data. However, to the best of our knowledge, this study is the first attempt of combining the individual behavioral and the engagement behavioral data, as well as the longitudinal and the external data for user response modeling in social media.

How to deal with large amount of data is another challenge. Social media sites produce large amount of user data. For example, the daily volume of posts mentioning some well-known brands or products such as Google, Microsoft, Sony, iPhone and iPad in Twitter is in the millions (Li and Li, 2013). It is necessary to use marketing intelligence methods to automatically analyze the massive data. The analysis of the massive data requires efficient preprocessing and excellent scalability of the prediction models.

In this study, a hierarchical ensemble learning framework is developed for behavior-aware user response modeling in social media. In the framework, a general-purpose preprocessing and transformation strategy is proposed to transform the large-scale and multi-relational user datasets derived from social media sites into customer-centered datasets and to extract prediction attributes. An improved hierarchical multiple kernel SVM (H-MK-SVM), as an extension of the SVM and the multiple kernel SVM (MK-SVM), is developed to model diverse heterogeneous data including external, tag and keyword, individual behavioral and engagement behavioral data. Because of the multi-relations of the individual behavioral and engagement behavioral data, one advantage of the improved H-MK-SVM is to adaptively select associated attributes. Another advantage of this method is to integrate the diverse heterogeneous social media data into a unified ensemble classifier to improve the prediction performance.

This paper is organized as follows. The next section presents the hierarchical ensemble learning framework for behavior-aware user response modeling in social media. Section 3 describes the database used in the study and presents the data preprocessing and transformation strategy. The model formulation of the improved H-MK-SVM is presented in Section 4. The computational results are reported in Section 5. Conclusions and directions for future research are given in Section 6.

2. The Hierarchical Ensemble Learning Framework

In this section, diverse heterogeneous data used for user response modeling in social media are discussed. A hierarchical ensemble learning framework is then proposed for user response modeling in social media using the diverse heterogeneous data.

2.1 Diverse heterogeneous data in social media

User response modeling in social media involves diverse heterogeneous data. In general, two categories of data, *i.e.*, external data and behavioral data (Bose and Chen, 2009), are used for customer response modeling. The external data include the demographic, lifestyle and geographic data of the customers (Bose and Chen, 2009). For social media, tags and keywords make up another type of external data. A tag is a word, sign or image selected by a user as his/her descriptions and a keyword is a word with special meaning extracted from the contents of a media site such as a tweet, a retweet and comments generated by users. Tags and keywords are usually used for the descriptions of users' interests (Chen, Liu *et* *al.*, 2012). In comparison with the external data, the behavioral data are more diverse and informative. As shown in Figure 1, the behavioral data can be grouped into individual behavioral data and engagement behavioral data, and can also be grouped into aggregated behavioral data and longitudinal behavioral data. The aggregated individual behavioral data are usually used in more traditional user response models. The RFM and historical records of user responses are the commonly used behavioral attributes. With the rapid development of social media, firms can easily collect large amount of longitudinal engagement behavioral data. These informative and valuable data have the potential to significantly improve the prediction performance of response models.

For the external data and the behavioral data, each customer is treated as an observation and n is used to represent the number of observations in the dataset. A customer is a respondent if the customer responds to an Item or takes action after an event such as a specific marketing campaign or a non-respondent otherwise. In the binary classification problem of customer response modeling, customer i is assigned the class label $y_i = 1$ if the customer is a respondent or $y_i = -1$ if the customer is a non-respondent.

External data can be described as a matrix $\mathbf{s}$ in which each row represents an observation and each column represents a static variable. In a dataset with m_1 variables, the attributes of a customer i is usually represented by the vector $\mathbf{s}_i = \{s_{ij} \mid j = 1, \cdots, m_1\}$. Different from numerical data, tags and keywords are usually described as textual or symbolic data. They can be represented by $\hat{\mathbf{s}}_i = \{\hat{s}_{i\hat{j}} \mid \hat{j} = 1, \cdots, m_3\}$ where m_3 is the number of tags and keywords.

All standard data mining tasks, including classification, regression and clustering, and the corresponding data mining methods require the input data be organized as a rectangular matrix. However, the longitudinal individual behavioral data are described as customer-centered multivariate time series of fixed length (Chen, Fan and Sun, 2012). A tensor is a multi-dimensional array which can be considered as the generalization of vectors and matrices. A first-order tensor is a vector, a second-order tensor is a matrix, and a tensor with three or higher orders is called a high-order tensor (Kolda and Bader, 2009). Therefore, the longitudinal individual behavioral data can be represented by a third-order tensor $\mathcal{B} = \{\mathbf{B}_i \mid i = 1, \cdots, n\}$. Each input of a customer i is represented by a rectangular matrix $\mathbf{B}_i = \{b_{i\tilde{j}t} \mid \tilde{j} = 1, \cdots, m_2;\ t = 1, \cdots, T\}$ where m_2 represents the number of longitudinal individual behavioral attributes and T represents the number of time points in each longitudinal behavioral variable.

In social media, social links among users carry informative information for user response modeling. For example, because of the social links between two users A and B, incorporating the behavioral data of user B into the response models may improve the prediction accuracy of the response of user A. In this study, the engagement behavioral data are defined as the customer-centered behavioral data of a fixed number of followees of a customer. As shown in Figure 2, the longitudinal engagement behavioral data for a user i can be represented by a fourth-order tensor $\hat{\mathcal{B}} = \{\hat{\mathcal{B}}_i \mid i = 1, \cdots, n\}$. Each input of a customer i is represented by a third-order tensor $\hat{\mathcal{B}}_i = \{\hat{b}_{ij'tf} \mid j' = 1, \cdots, m_4;\ t = 1, \cdots, \hat{T};\ f = 1, \cdots, N\}$ where m_4 represents the number of longitudinal individual behavior attributes of each followee f, $\hat{T}$ represents the number of time points in each longitudinal engagement behavioral variable and N represents the number of followees. Each input of a followee f, as individual behavioral data, can be represented by a third-order tensor $\hat{\mathcal{B}}_f = \{\hat{b}_{ij'tf} \mid i = 1, \cdots, n;\ j' = 1, \cdots, m_4;\ t = 1, \cdots, T\}$.

These four types of data are illustrated in Figure 2. Dealing with the heterogeneous and high-order tensor data is an essential problem in user response modeling and is discussed in the next sub-section.

Figure 2 approximately here

2.2 The hierarchical ensemble learning framework

Targeting potential customers using large, diverse and heterogeneous data generated on social media sites is a difficult task. Three difficult issues need be addressed: (1) identifying the most useful data and generating the customer-centered individual and engagement behavioral datasets; (2) selecting associated attributes from coupled individual and engagement behavioral data; (3) integrating the diverse heterogeneous social media data into a classification model to predict user responses.

A hierarchical ensemble learning framework, as illustrated in Figure 3, is proposed for user response modeling using external, tag and keyword, longitudinal individual and engagement behavioral data. The framework can be organized into three layers. In Layer 1, the original datasets are transformed into customer-centered external, tag and keyword, longitudinal individual and engagement behavioral datasets. In Layer 2, features are selected from the longitudinal individual and engagement behavioral data. In Layer 3, an ensemble classifier is trained using the four types of data. The major tasks of hierarchical ensemble learning framework are described in the following.

Data preprocessing, data transformation and feature extraction. In Layer 1 of the proposed hierarchical ensemble learning framework, the original large-size and multi-relational datasets are preprocessed to generate relatively small-size datasets and are transformed into customer-centered datasets including external data $\mathbf{s}$ and $\hat{\mathbf{s}}$, longitudinal individual behavioral data $\mathcal{B}$ and customer-centered social network data. Longitudinal individual behavioral data $\mathcal{B}$ and social network data are simultaneously used for feature extraction to obtain longitudinal engagement behavioral data $\hat{\mathcal{B}}$. The details of data preprocessing, transformation and feature extraction are discussed using a real dataset in Section 3.

Associated attribute selection. The customer-customer interactions make the individual and engagement behavioral data coupled with each other (Cao *et al.*, 2012). It is difficult to analyze and model the coupled behavioral data partially because of the multi-correlation among the large amount of longitudinal behavioral attributes. Associated attribute selection, as an important task in Layer 2 of the proposed hierarchical ensemble learning framework, is an effective method to reduce the redundant attributes to improve prediction performance (Buckinx *et al.*, 2004; Crone *et al.*, 2006). For this task, a sparse modeling method is adopted to learn the weights of the longitudinal behavioral attributes, and the attributes with non-zero weights are kept as associated attributes.

Ensemble learning. In Layer 3 of the hierarchical ensemble learning framework, different types of kernels are adopted to model the external, tag and keyword, longitudinal individual and engagement behavioral data, respectively. An ensemble classifier is developed to combine these types of data using the weights of the longitudinal behavioral attributes obtained by the associated attribute selection.

Most existing classifiers cannot combine the above mentioned four types of data to predict customer responses. Therefore, a hierarchical ensemble learning method, the improved H-MK-SVM based on the work of Chen, Fan and Sun (2012), is developed for associated attribute selection and ensemble learning. This method is discussed in detail in Section 4.

Figure 3 approximately here

3. The Data

In this section, the database used for the computational experiments is introduced first. The data preprocessing and transformation strategy for large-scale and multi-relational datasets is then described in detail using this database.

3.1 The database

A real-world database provided by Tencent Weibo[1] is used in the computational experiments. Micoblogs, as mainstream social media, become a new marketing platform of EWOM (Li and Li, 2013). Tencent Weibo is one of the largest microblog websites in China. The characteristics of the database are given in Table 1. In the database, four datasets including Training, User Profile, Item and User SNS were used in the following experiments.

The Training dataset contains 73,209,277 historical records about users' responses to different Items over a span of 32 days. Each observation in the Training dataset records the response of a user to an Item at a time. The time period of the dataset is from October 12 to November 12, 2011. The User Profile dataset is the only customer-centered dataset in the database. The dataset records the year of birth, the gender and the number of tweets of each of the 2,320,895 users with numerical values. It also records the tags of the users with strings. There are 6,095 Items in the Item dataset. The category and keywords of each Item are recorded in the Item dataset with strings. The User SNS dataset contains the follow history of each user. There are 50,655,143 records in the User SNS dataset. The relationships of the customer-customer interactions are derived from the follow history.

Table 1 approximately here

3.2 Data preprocessing and transformation

In the computational experiments, Microsoft Access 2010 was used to store the original database and Microsoft Excel 2010 was used to transform the original large-scale and multi-relational datasets into customer-centered datasets. For each single Item, there is a small number of labeled samples. Each Item belongs to a hierarchical category. Therefore, data on the responses to Items belonging to specific categories, rather than to a single Item, are analyzed. Data on responses to Items belonging to Category 1.1.1.1 are analyzed and used in the computational experiments. The historical records in the period from October 12 to November 10 were used to train the improved H-MK-SVM models and those in the period from November 11 to November 12 were used to test the models.

Computational experiments were conducted first without using the engagement behavioral data. Microsoft query in Microsoft Excel 2010 was used in selecting the observations with the Items in Category 1.1.1.1 into the Training dataset. Pivot Table in Microsoft Excel 2010 was used to transform the selected

[1]http://kddcup2012.org/c/kddcup2012-track1/data.

observations into customer-centered longitudinal individual behavioral dataset. In the transformed dataset, each row represents the historical record of a user and each column represents the historical records of all users in a day. The observations (users) having non-empty fields in the period from November 11 to November 12 were kept as samples and others were deleted from the dataset. Pivot Table in Microsoft Excel 2010 was then used to transform the User Profile dataset into the external dataset to obtain the same observations (users) as those in the longitudinal individual behavioral dataset. Observations with missing values in either the longitudinal individual behavioral dataset or the external dataset were deleted. The empty values of the tags in the external dataset were all set to zeros.

The characteristics of the external data with numerical values, the tag data and the longitudinal individual behavioral data are presented in the first three rows in Table 2. After data preprocessing and transformation, each of these three datasets contains 22,548 observations (users) in which 1,979 observations are positive instances (respondents) and 20,569 observations are negative instances (non-respondents). There are m_1 =2 variables, *i.e.*, gender and the number of tweets, in the external dataset and m_3 =10 tags in the tag dataset. In the longitudinal individual behavioral dataset, there are m_2 =2 variables including the number of responses per day (Quantity) and whether or not the user accepted the recommendation of the Items in the category (Acceptance). A common normalization method was applied to the external and longitudinal individual behavioral data to rescale the values of each variable to the range between 0 and 1.

A holdout validation approach was used to partition the datasets into training sets, validation sets and testing sets with 458, 11045 and 11045 observations, respectively. For customer response modeling, the number of respondents is usually much smaller than the number of non-respondents (Cui *et al.*, 2006; Lee *et al.*, 2010). As shown in Table 2, the response rate is 8.78% with 1,979 respondents and 20,569 non-respondents. The undersampling method, one of the most commonly used techniques for dealing with highly imbalanced data (Burez and Van den Poel, 2009; Chen, Fan and Sun, 2012; Kang *et al.*, 2012; Verbeke *et al.*, 2011), is used to select balanced training sets. The sampling ratio θ is defined to be the number of non-respondents over the number of respondents. In the training sets, the sampling ratio was set to $\theta = 1$, while those in the validation and testing sets were almost equal to that in the whole datasets, *i.e.*, $\theta = 10.39$.

Computational experiments were then conducted by incorporating the engagement behavioral data into user response modeling. Pivot Table of Microsoft Excel 2010 was used to transform the User SNS dataset into the customer-centered social network dataset. Pivot Table of Microsoft Excel 2010 was then used to transform two relational datasets, *i.e.*, the customer-centered social network and the longitudinal individual behavioral datasets, into customer-centered longitudinal engagement behavioral dataset. In the transformed dataset, observations (users) without followees were automatically deleted. As a result, the number of observations in this transformed dataset is smaller than that in other transformed datasets. As shown in the fourth row of Table 2, the dataset contains 9,105 observations in which 875 observations are positive instances (respondents) and 8,230 observations are negative instances (non-respondents). Pivot Table in Microsoft Excel 2010 was used to filter the external, tag and the longitudinal individual behavioral dataset to obtain the same observations as those in the longitudinal engagement behavioral dataset. As shown in Figure 2, the longitudinal engagement behavioral data is a fourth-order tensor. A simple weighted average strategy was used to aggregate the fourth-order tensor into a third-order tensor along the followee dimension $\hat{\mathcal{B}} = \{\hat{b}_{ij't} \mid i = 1, \cdots, n;\ j' = 1, \cdots, m_4;\ t = 1, \cdots, T\}$.

The datasets for the computational experiments with the engagement behavioral data were also partitioned into training sets, validation sets and testing sets with 458, 4323 and 4324 observations, respectively. In the training sets, the sampling ratio θ was also set to $\theta = 1$. The sampling ratio θ in the validation sets was equal to that in the testing sets.

4. The Model

An improved H-MK-SVM, based on the work of Chen, Fan and Sun (2012), is developed in the hierarchical ensemble learning framework. The H-MK-SVM is an extension of the SVM and the MK-SVM. The SVM is one of the most popular and effective machine learning techniques and usually has excellent classification performance in practical applications (Chapelle *et al*., 2002; Vapnik, 1998). The MK-SVM, as an important extension of the SVM, can integrate heterogeneous data and adaptively select the best combinations of multiple basic kernels in the learning process (Bach *et al.*, 2004; Chen, Fan and Sun, 2012; Gönen and Alpaydın, 2011; Lanckrient *et al.*, 2004). The H-MK-SVM was developed to model longitudinal individual behavioral data for the application of customer churn prediction (Chen, Fan and Sun, 2012). A three phase training algorithm for the H-MK-SVM is developed to sequentially learn the Lagrange multipliers, the weight of each longitudinal behavioral attribute and the weight of each single feature basic kernel. Chen, Fan and Sun (2012) provided more details about the MK-SVM and the H-MK-SVM.

The improved H-MK-SVM includes two sequential tasks, *i.e.*, the associated attribute selection and ensemble learning. Each task adopts a two phase training algorithm to sequentially learn the Lagrange multipliers and the weight of each basic kernel. The associated attribute selection is adopted to deal with multi-relations of the individual and engagement behavioral data. Different types of kernels used to model the four types of data are then combined to obtain the final model by ensemble learning. As shown in the hierarchical ensemble learning framework discussed in Section 2.2, the associated attribute selection by the improved H-MK-SVM in Layer 2 is discussed in Section 4.1, and the ensemble learning by the improved H-MK-SVM in Layer 3 is discussed in Section 4.2.

4.1 Associated attribute selection by the improved H-MK-SVM

In Layer 2 of the hierarchical ensemble learning framework, the input data to the improved H-MK-SVM consist of a training dataset $G = \{(\mathbf{B}_{1,} \hat{\mathbf{B}}_{1,} y_1), \cdots, (\mathbf{B}_{n,} \hat{\mathbf{B}}_{n,} y_n)\}$ with the longitudinal individual and engagement behavioral data. The improved H-MK-SVM in Layer 2 of the hierarchical ensemble learning framework constructs an optimal hyperplane in a high dimensional feature space

$$f(\mathbf{B}, \hat{\mathbf{B}}) = \sum_m \mathbf{w}_m^{\mathbf{T}} \cdot \phi_m + b', \tag{1}$$

where ϕ_m is the nonlinear map, $\mathbf{w}_m$ is the vector of weights and b' is the bias.

For the longitudinal individual behavioral data, the multiple kernel (2) in the following is used to map the elements of the input matrices $\mathbf{B}_i$ onto high-dimensional feature spaces via the nonlinear maps $\phi_1(\mathbf{B}_i) \cdots \phi_{m_2 \times T}(\mathbf{B}_i)$

$$K_2(\mathbf{B}_i, \mathbf{B}_{\tilde{i}}) = \sum_{\tilde{j}=1}^{m_2} \sum_{t=1}^{T} \gamma_{\tilde{j},t} k_{\tilde{j},t}(b_{i,\tilde{j},t}, b_{\tilde{i},\tilde{j},t}), \tag{2}$$

where $k_{\tilde{j},t}(b_{i,\tilde{j},t}, b_{\tilde{i},\tilde{j},t}) = \phi_{\tilde{j},t}(\mathbf{B}_i) \cdot \phi_{\tilde{j},t}(\mathbf{B}_{\tilde{i}})$ is the basic kernel and $\gamma_{\tilde{j},t}$ is the weight of $k_{\tilde{j},t}(b_{i,\tilde{j},t}, b_{\tilde{i},\tilde{j},t})$. For the longitudinal engagement behavioral data, a similar multiple kernel (3) in the following is used to map the elements of the input matrices $\hat{\mathbf{B}}_i$ onto feature spaces via the nonlinear maps $\phi_{m_2 \times T+1}(\hat{\mathbf{B}}_i) \cdots \phi_{(m_2 T + m_4 \hat{T})}(\hat{\mathbf{B}}_i)$

$$K_4(\hat{\mathbf{B}}_i, \hat{\mathbf{B}}_{\tilde{i}}) = \sum_{j'=1}^{m_4} \sum_{t'=1}^{\hat{T}} \hat{\gamma}_{j',t'} k_{j',t'}(\hat{b}_{i,j',t'}, \hat{b}_{\tilde{i},j',t'}), \tag{3}$$

where $k_{j',t'}(\hat{b}_{i,\tilde{j},t'},\hat{b}_{\tilde{i},\tilde{j},t'})=\phi_{j',t'}(\hat{\mathbf{B}}_i)\cdot\phi_{j',t'}(\hat{\mathbf{B}}_{\tilde{i}})$ is the basic kernel and $\hat{\gamma}_{j',t'}$ is the weight of $k_{j',t'}(\hat{b}_{i,j',t'},\hat{b}_{\tilde{i},j',t'})$. When convenient, $\boldsymbol{\gamma}$ and $\hat{\boldsymbol{\gamma}}$ are used to denote the vectors of all the weights of the basic kernels in (2) and (3), respectively, and $\boldsymbol{\gamma}'=(\boldsymbol{\gamma},\ \hat{\boldsymbol{\gamma}})$ is used to denote the composite vector consisting of the elements of $\boldsymbol{\gamma}$ and $\hat{\boldsymbol{\gamma}}$. The values of the elements of $\boldsymbol{\gamma}'$ are determined in the attribute selection process.

When the multiple kernels in (2) and (3) are used, the improved H-MK-SVM in Layer 2 of the ensemble learning framework is formulated as the following quadratic program

$$\min_{\boldsymbol{\gamma}}\min_{\mathbf{w}_m,\xi,b'} \quad \frac{1}{2}\sum_{m=1}^{m_2T+m_4\hat{T}}\frac{\|\mathbf{w}_m\|^2}{\gamma'_m}+C\sum_{i=1}^{n}\xi_i \tag{4}$$

$$\text{s.t.} \quad y_i(\sum_m \mathbf{w}_m^T\phi_m+b')\geq 1-\xi_i \qquad i=1,\cdots,n \tag{5}$$

$$\xi_i\geq 0 \qquad i=1,\cdots,n \tag{6}$$

$$\gamma'_m\geq 0 \qquad m=1,\cdots,(m_2T+m_4\hat{T}), \tag{7}$$

where C is the regularization parameter, $\mathbf{w}_m$ is the vector of weights, b' is the bias, ξ_i is the relaxation or the error term for observation i and $\gamma'_m\in\boldsymbol{\gamma}'$.

A two-phase iterative procedure (Chen *et al.*, 2007) is used to decompose the problem in (4)-(7) into two sub-problems and to solve them iteratively. In phase 1, the values of the elements of $\boldsymbol{\gamma}'$ are fixed and the dual of (4)-(7) is solved. The dual is rewritten as the following quadratic program

$$\max_{\boldsymbol{\alpha}} \quad \sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,\tilde{i}=1}^{n}\alpha_i\alpha_{\tilde{i}}y_iy_{\tilde{i}}\left(\sum_{\tilde{j}=1}^{m_2}\sum_{t=1}^{T}\gamma_{\tilde{j},t}k(b_{i,\tilde{j},t},b_{\tilde{i},\tilde{j},t})+\sum_{j'=1}^{m_4}\sum_{t'=1}^{\hat{T}}\hat{\gamma}_{j',t'}k(\hat{b}_{i,j',t'},\hat{b}_{\tilde{i},j',t'})\right) \tag{8}$$

$$\text{s.t.} \quad \sum_{i=1}^{n}y_i\alpha_i=0 \tag{9}$$

$$0\leq\alpha_i\leq C,\quad i=1,\cdots,n, \tag{10}$$

where α_i is the Lagrange multiplier of observation i. The values of the dual variables are determined after the dual in (8)-(10) is solved. Using the values of the dual variables obtained in phase 1, the primal variables $\mathbf{w}_m$ can be written as

$$\mathbf{w}_m=\gamma'_m\sum_{i=1}^{n}y_i\alpha_i\phi_m,\ m=1,\cdots,(m_2T+m_4\hat{T}). \tag{11}$$

In phase 2, the values of b' and the components of $\mathbf{w}_m$ are fixed. Using the primal variables $\mathbf{w}_m$ in (11), the original problem in (4)-(7) can be rewritten as

$$\min_{\gamma,\hat{\gamma}} \quad \frac{1}{2}\sum_{i,\tilde{i}=1}^{n} \alpha_i\, \alpha_{\tilde{i}}\, y_i y_{\tilde{i}} \left(\sum_{\tilde{j}=1}^{m_2}\sum_{t=1}^{T} \gamma_{\tilde{j},t} k(b_{i,\tilde{j},t}, b_{\tilde{i},\tilde{j},t}) + \sum_{j'=1}^{m_4}\sum_{t'=1}^{\hat{T}} \hat{\gamma}_{j',t'} k(\hat{b}_{i,j',t'}, \hat{b}_{\tilde{i},j',t'}) \right) + \lambda \sum_{i=1}^{n} \xi_i \tag{12}$$

$$\text{s.t.} \quad y_i \left\{ \sum_{\tilde{i}=1}^{n} \alpha_{\tilde{i}}\, y_{\tilde{i}} \left(\sum_{\tilde{j}=1}^{m_2}\sum_{t=1}^{T} \gamma_{\tilde{j},t} k(b_{i,\tilde{j},t}, b_{\tilde{i},\tilde{j},t}) + \sum_{j'=1}^{m_4}\sum_{t'=1}^{\hat{T}} \hat{\gamma}_{j',t'} k(\hat{b}_{i,j',t'}, \hat{b}_{\tilde{i},j',t'}) \right) + \tilde{b} \right\} \geq 1 - \xi_i \tag{13}$$

$$\xi_i \geq 0 \qquad i = 1,\cdots,n \tag{14}$$

$$\gamma_{\tilde{j},t} \geq 0 \qquad \tilde{j} = 1,\cdots,m_2,\ t = 1,\cdots,T \tag{15}$$

$$\hat{\gamma}_{j',t'} \geq 0 \qquad j' = 1,\cdots,m_4,\ t' = 1,\cdots,\hat{T}\ , \tag{16}$$

where λ is the regularization parameter and α_i is the Lagrange multiplier of observation i obtained in phase 1. Because minimizing the L_1-norm based regularization function leads to sparse solutions for the elements of $\boldsymbol{\gamma}$ and $\hat{\boldsymbol{\gamma}}$, the training process of the problem (12)-(16) is also a feature selection process. The solutions of the problem can be obtained by solving its dual. The dual is stated in (17)-(21) in the following

$$\max \quad \sum_{i=1}^{n} u_i \tag{17}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} u_i y_i \sum_{\tilde{i}=1}^{n} \alpha_{\tilde{i}}\, y_{\tilde{i}} \sum_{t=1}^{T} k(b_{i,\tilde{j},t}, b_{\tilde{i},\tilde{j},t}) \leq \frac{1}{2}\sum_{i=1}^{n}\sum_{\tilde{i}=1}^{n} \alpha_i\, \alpha_{\tilde{i}}\, y_i y_{\tilde{i}} k(b_{i,\tilde{j},t}, b_{\tilde{i},\tilde{j},t}) \qquad \tilde{j} = 1,\cdots,m_2 \times T \tag{18}$$

$$\sum_{i=1}^{n} u_i y_i \sum_{\tilde{i}=1}^{n} \alpha_{\tilde{i}}\, y_{\tilde{i}} \sum_{t'=1}^{\hat{T}} k(\hat{b}_{i,j',t'}, \hat{b}_{\tilde{i},j',t'}) \leq \frac{1}{2}\sum_{i=1}^{n}\sum_{\tilde{i}=1}^{n} \alpha_i\, \alpha_{\tilde{i}}\, y_i y_{\tilde{i}} k(\hat{b}_{i,j',t'}, \hat{b}_{\tilde{i},j',t'}) \qquad j' = 1,\cdots,m_4 \times \hat{T} \tag{19}$$

$$\sum_{i=1}^{n} u_i y_i = 0 \tag{20}$$

$$0 \leq u_i \leq \lambda \qquad i = 1,\cdots,n\ , \tag{21}$$

where u_i is the dual variable associated with observation i. Sometimes, the weights of the kernels, *i.e.*, the elements of $\boldsymbol{\gamma}$ and $\hat{\boldsymbol{\gamma}}$, in the objective function (12) can be set to constants to make the linear program in (12)-(16) easier to solve. For example, (22) and (23) in the following can be used in (12)

$$\frac{1}{2}\sum_{i=1}^{n}\sum_{\tilde{i}=1}^{n} \alpha_i\, \alpha_{\tilde{i}}\, y_i y_{\tilde{i}} k(b_{i,\tilde{j},t}, b_{\tilde{i},\tilde{j},t}) = 1 \tag{22}$$

$$\frac{1}{2}\sum_{i=1}^{n}\sum_{\tilde{i}=1}^{n} \alpha_i\, \alpha_{\tilde{i}}\, y_i y_{\tilde{i}} k(\hat{b}_{i,j',t'}, \hat{b}_{\tilde{i},j',t'}) = 1. \tag{23}$$

When the variables $\mathbf{w}_m$ and the kernels in (2) and (3) are used in (1), the classification function constructed by the improved H-MK-SVM in Layer 2 of the hierarchical ensemble learning framework is

$$Y(\mathbf{B}_{\tilde{i}},\hat{\mathbf{B}}_{\tilde{i}})=\operatorname{sgn}\left\{\sum_{i=1}^{n}\alpha_i\ y_i\left(\sum_{\tilde{j}=1}^{m_2}\sum_{t=1}^{T}\gamma_{\tilde{j},t}k(b_{i,\tilde{j},t},b_{\tilde{i},\tilde{j},t})+\sum_{j'=1}^{m_4}\sum_{t'=1}^{\hat{T}}\hat{\gamma}_{j',t'}k(\hat{b}_{i,j',t'},\hat{b}_{\tilde{i},j',t'})\right)+b'\right\}. \tag{24}$$

The bias b' in (24) is computed using (25) in the following

$$b'=y_{\tilde{i}}-\sum_{i=1}^{n}\alpha_i\ y_i\left(\sum_{\tilde{j}=1}^{m_2}\sum_{t=1}^{T}\gamma_{\tilde{j},t}k(b_{i,\tilde{j},t},b_{\tilde{i},\tilde{j},t})+\sum_{j'=1}^{m_4}\sum_{t'=1}^{\hat{T}}\hat{\gamma}_{j',t'}k(\hat{b}_{i,j',t'},\hat{b}_{\tilde{i},j',t'})\right),\ \text{for}\ \alpha_{\tilde{i}}\in(0,C)\,. \tag{25}$$

4.2 Ensemble learning by the improved H-MK-SVM

In Layer 3 of the hierarchical ensemble learning framework, the input data of the improved H-MK-SVM consist of a training dataset $\tilde{G}=\{(\mathbf{s}_1,\hat{\mathbf{s}}_1,\mathbf{B}_1,\hat{\mathbf{B}}_1,y_1),\cdots,(\mathbf{s}_n,\hat{\mathbf{s}}_n,\mathbf{B}_n,\hat{\mathbf{B}}_n,y_n)\}$ with external and behavioral data. The improved H-MK-SVM in Layer 3 of the hierarchical ensemble learning framework constructs an optimal hyperplane in a high dimensional feature space

$$\tilde{f}(\mathbf{s},\hat{\mathbf{s}},\mathbf{B},\hat{\mathbf{B}})=\sum_{\hat{m}=1}^{M}\tilde{\mathbf{w}}_{\hat{m}}^{\mathbf{T}}\cdot\phi_{\hat{m}}+\tilde{b}\,, \tag{26}$$

where $\phi_{\hat{m}}$ is the nonlinear map, $\tilde{\mathbf{w}}_{\hat{m}}$ is the vector of weights, $M=4$ is the number of types of data and $\tilde{b}$ is the bias.

Different kernels are used for different types of data. For the external data with numerical values, the standard single Gaussian kernel is used. For the external data with strings, the string kernel in (27) in the following is used

$$K_3(\hat{\mathbf{s}}_i,\hat{\mathbf{s}}_{\tilde{i}})=\sum_{m=1}^{m_3}\mathrm{I}(\hat{s}_{i,\hat{j}},\hat{s}_{\tilde{i},\hat{j}})\,, \tag{27}$$

where m_3 is the number of external attributes with strings. In (27), $\mathrm{I}(\hat{s}_{i,\hat{j}},\hat{s}_{\tilde{i},\hat{j}})$ is given in (28) in the following

$$\mathrm{I}(\hat{s}_{i,\hat{j}},\hat{s}_{\tilde{i},\hat{j}})=\begin{cases}1,\ \text{if}\ (\hat{s}_{i,\hat{j}}=\hat{s}_{\tilde{i},\hat{j}})\\ 0,\ \text{if}\ (\hat{s}_{i,\hat{j}}=\hat{s}_{\tilde{i},\hat{j}})\end{cases}. \tag{28}$$

The following multiple kernel (29) similar to (2) is used for the longitudinal individual behavioral data

$$\tilde{K}_2(\mathbf{B}_i,\mathbf{B}_{\tilde{i}})=\sum_{\tilde{j}=1}^{m_2}\sum_{t=1}^{T}\gamma_{\tilde{j},t}k_{\tilde{j},t}(b_{i,\tilde{j},t},b_{\tilde{i},\tilde{j},t})\,, \tag{29}$$

where $k_{\tilde{j},t}(b_{i,\tilde{j},t}, b_{\tilde{i},\tilde{j},t})$ is the basic kernel and $\gamma_{\tilde{j},t}$ is the known weight obtained in Layer 2 of the hierarchical ensemble learning framework. The following multiple kernel (30) similar to (3) is used for the longitudinal engagement behavioral data

$$\tilde{K}_4(\hat{\mathbf{B}}_i, \hat{\mathbf{B}}_{\tilde{i}}) = \sum_{j'=1}^{m_4}\sum_{t'=1}^{\hat{T}} \hat{\gamma}_{j',t'} k_{j',t'}(\hat{b}_{i,j',t'}, \hat{b}_{\tilde{i},j',t'}) , \tag{30}$$

where $k_{j',t'}(\hat{b}_{i,\tilde{j},t'}, \hat{b}_{\tilde{i},\tilde{j},t'})$ is the basic kernel and $\hat{\gamma}_{j',t'}$ is the known weight obtained in Layer 2 of the hierarchical ensemble learning framework.

When the kernels in (29) and (30) above are used, the model of the improved H-MK-SVM in Layer 3 of the hierarchical ensemble learning framework is formulated as

$$\min_{\boldsymbol{\beta}} \min_{\mathbf{w}_{\hat{m}}, \tilde{\xi}, \tilde{b}} \quad \frac{1}{2}\sum_{\hat{m}=1}^{M} \frac{\left\|\mathbf{w}_{\hat{m}}\right\|^2}{\beta_{\hat{m}}} + \tilde{C}\sum_{i=1}^{n}\tilde{\xi}_i \tag{31}$$

$$\text{s.t.} \quad y_i(\sum_{\hat{m}} \mathbf{w}_{\hat{m}}^T \phi_{\hat{m}} + \tilde{b}) \geq 1 - \tilde{\xi}_i \qquad i = 1, \cdots, n \tag{32}$$

$$\tilde{\xi}_i \geq 0 \qquad i = 1, \cdots, n \tag{33}$$

$$\beta_{\hat{m}} \geq 0 \qquad \hat{m} = 1, \cdots, M , \tag{34}$$

where $\hat{C}$ is the regularization parameter, $\tilde{\xi}_i$ is the relaxation or error term for observation i and $\beta_{\hat{m}}$ is the weight of each type of data.

The two-phase procedure is used to solve the problem in (31)-(34). In phase 1, the values of the elements of $\boldsymbol{\beta}$ are fixed and the dual of (31)-(34) is solved. The dual is written as

$$\max_{\boldsymbol{\alpha}} \quad \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i,\tilde{i}=1}^{n} \alpha_i \alpha_{\tilde{i}} y_i y_{\tilde{i}} \mathbf{K} \tag{35}$$

$$\text{s.t.} \sum_{i=1}^{n} y_i \alpha_i = 0 \tag{36}$$

$$0 \leq \alpha_i \leq \tilde{C}, \quad i = 1, \cdots, n \tag{37}$$

$$\beta_{\hat{m}} \geq 0, \quad \hat{m} = 1, \cdots, M . \tag{38}$$

In (35), $\mathbf{K}$ is shown in (39) in the following

$$\mathbf{K} = \beta_1 \sum_{j=1}^{m_1} k(s_{i,j}, s_{\tilde{i},j}) + \beta_2 \sum_{\tilde{j}=1}^{m_2}\sum_{t=1}^{T} \gamma_{\tilde{j},t} k(b_{i,\tilde{j},t}, b_{\tilde{i},\tilde{j},t}) + \beta_3 \sum_{\hat{j}=1}^{m_3} k(\hat{s}_{i,\hat{j}}, \hat{s}_{\tilde{i},\hat{j}}) + \beta_4 \sum_{j'=1}^{m_4}\sum_{t'=1}^{\hat{T}} \hat{\gamma}_{j',t'} k(\hat{b}_{i,j',t'}, \hat{b}_{\tilde{i},j',t'}) . \tag{39}$$

The values of the dual variables are determined after the dual in (35)-(38) is solved. The primal variables $\mathbf{w}_{\hat{m}}$ are then represented by the dual variables as

$$\mathbf{w}_{\hat{m}} = \beta_{\hat{m}} \sum_{i=1}^{n} y_i \alpha_i \phi_{\hat{m}}, \ \hat{m} = 1, \cdots, M \ . \tag{40}$$

In phase 2, the values of $\tilde{b}$ and the elements of $\mathbf{w}_{\hat{m}}$ are fixed. Using the primal variables $\mathbf{w}_{\hat{m}}$ in (40), the original problem in (31)-(34) can be written as the linear program in (41)-(44) in the following

$$\min_{\boldsymbol{\beta}} \quad \frac{1}{2} \sum_{i,\tilde{i}=1}^{n} \alpha_i \, \alpha_{\tilde{i}} \, y_i y_{\tilde{i}} \mathbf{K} + \tilde{\lambda} \sum_{i=1}^{n} \tilde{\xi}_i \tag{41}$$

$$\text{s.t.} \quad y_i \left\{ \sum_{\tilde{i}=1}^{n} \alpha_{\tilde{i}} \, y_{\tilde{i}} \mathbf{K} + \tilde{b} \right\} \geq 1 - \tilde{\xi}_i \tag{42}$$

$$\tilde{\xi}_i \geq 0 \qquad\qquad i = 1, \cdots, n \tag{43}$$

$$\beta_{\hat{m}} \geq 0 \qquad\qquad \hat{m} = 1, \cdots, M \ . \tag{44}$$

where $\tilde{\lambda}$ is the regularization parameter.

When the variables $\mathbf{w}_{\hat{m}}$ and the kernels for the four types of data are used in (26), the classification function constructed by the improved H-MK-SVM in Layer 3 of the hierarchical ensemble learning framework is

$$\tilde{Y}(\mathbf{s}_{\tilde{i}}, \hat{\mathbf{s}}_{\tilde{i}}, \mathbf{B}_{\tilde{i}}, \hat{\mathbf{B}}_{\tilde{i}}) = \operatorname{sgn} \left\{ \sum_{i=1}^{n} \alpha_i \, y_i \mathbf{K} + \tilde{b} \right\} . \tag{45}$$

The bias $\tilde{b}$ in (45) is computed using (46) in the following

$$\tilde{b} = y_{\tilde{i}} - \sum_{i=1}^{n} \alpha_i \, y_i \mathbf{K} \text{ , for } \alpha_{\tilde{i}} \in (0, \tilde{C}) \ . \tag{46}$$

It should be noted that the improved H-MK-SVM in Layer 2 and Layer 3 of the hierarchical ensemble learning framework play different roles. The one in Layer 2 using two types of behavioral data is for associated attribute selection and the one in Layer 3 obtains the final classification function (45) for user response modeling. After associated attribute selection, the H-MK-SVM in Layer 2 can also construct a classification function for user response modeling using just the two types of behavioral data. The improved H-MK-SVM based hierarchical ensemble learning combines the advantages of sparse modeling with reduced feature sets and the ensemble learning with diverse heterogeneous data and improves the performance of user response modeling.

4.3 Parameter tuning

In the improved H-MK-SVM, the Gaussian kernel, also known as the radial basis function (RBF) kernel, is used for the external data with numerical values and as the basic kernels for the longitudinal behavioral data in (2), (3), (29) and (30). For the external data with numerical values, the Gaussian kernel is stated in (47) in the following

$$K_1(\mathbf{s}_i, \mathbf{s}_{\tilde{i}}) = \exp\left(-\frac{1}{\sigma_1^2}\left\|\mathbf{s}_i - \mathbf{s}_{\tilde{i}}\right\|^2\right), \tag{47}$$

where σ_1^2 is the kernel parameter. For the longitudinal individual behavioral attribute Quantity, the following Gaussian kernel in (48) is used as the basic kernel of the multiple kernel (2)

$$k_{\tilde{j},t}(b_{i,\tilde{j},t}, b_{\tilde{i},\tilde{j},t}) = \exp\left(-\frac{1}{\sigma_2^2}\left\|b_{i,\tilde{j},t} - b_{\tilde{i},\tilde{j},t}\right\|^2\right), \tag{48}$$

where σ_2^2 is the kernel parameter. In the improved H-MK-SVM in Layer 3 of the hierarchical ensemble learning framework, the same Gaussian kernel is used as the basic kernel of the multiple kernel (29) with a different kernel parameter $\tilde{\sigma}_2^2$. For the longitudinal individual behavioral attribute Acceptance, Gaussian kernels with parameters σ_4^2 and $\tilde{\sigma}_4^2$ are used as the basic kernels of the multiple kernels in the improved H-MK-SVM in Layers 2 and 3, respectively, of the hierarchical ensemble learning framework.

For the longitudinal engagement behavioral attribute Quantity, the Gaussian kernel (49) in the following is used as the basic kernel of the multiple kernel (3)

$$k_{\tilde{j},t'}(\hat{b}_{i,\tilde{j},t'}, \hat{b}_{\tilde{i},\tilde{j},t'}) = \exp\left(-\frac{1}{\sigma_5^2}\left\|\hat{b}_{i,\tilde{j},t'} - \hat{b}_{\tilde{i},\tilde{j},t'}\right\|^2\right), \tag{49}$$

where σ_5^2 is the kernel parameter. In the improved H-MK-SVM in Layer 3 of the hierarchical ensemble learning framework, the Gaussian kernel with parameter $\tilde{\sigma}_5^2$ is used as the basic kernel in the multiple kernel (30). For the longitudinal engagement behavioral attribute Acceptance, the Gaussian kernel with parameters σ_6^2 and $\tilde{\sigma}_6^2$ are used in the improved H-MK-SVM in Layers 2 and 3, respectively, of the hierarchical ensemble learning framework.

The grid search method (Hsu *et al.*, 2003) was used to tune the free parameters in the improved H-MK-SVM. For the improved H-MK-SVM in Layer 2 of the hierarchical ensemble learning framework, exponentially growing values for λ, C, $1/\sigma_2^2$, $1/\sigma_4^2$, $1/\sigma_5^2$ and $1/\sigma_6^2$ from 10^{-2} to 10^2 were tried in turn. For the improved H-MK-SVM in layer 3 of the hierarchical ensemble learning framework, a nested grid search strategy was used. Exponentially growing values for $\tilde{\lambda}$, $\tilde{C}$, $1/\tilde{\sigma}_1^2$, $1/\tilde{\sigma}_2^2$, $1/\tilde{\sigma}_4^2$, $1/\tilde{\sigma}_5^2$ and $1/\tilde{\sigma}_6^2$

from 10^{-2} to 10^{2} were tried first in turn. These parameters were then finely tuned. Additively growing values in the best intervals obtained earlier for $\tilde{C}$, $1/\tilde{\sigma}_1^2$, $1/\tilde{\sigma}_2^2$, $1/\tilde{\sigma}_4^2$, $1/\sigma_5^2$ and $1/\sigma_6^2$, *i.e.*, $\tilde{C}=0.01$, $0.02,\cdots,\ 0.1$; ($1/\tilde{\sigma}_1^2$)=($1/\sigma_6^2$)= 0.1, $0.2,\cdots,\ 1$; ($1/\tilde{\sigma}_2^2$)=($1/\tilde{\sigma}_4^2$)=($1/\sigma_5^2$)=$1,\ 2,\cdots,\ 10$, were tried in turn. The values of these parameters with the best performance on the validation sets were used to test the performance of the model. Criteria measuring performance in this study are introduced in Section 5.

5. Computational Experiments

Computational results are reported in this section. These results include the comparisons of user response modeling performances with and without hierarchical ensemble learning; the comparisons of user response modeling performances using longitudinal and aggregated behavioral data with other more traditional method; the effects of varying time lengths and aggregation scales of the longitudinal data on prediction performance; and the effects of using the engagement behavioral data on prediction performance. Matlab 7.4 was used to conduct the computational experiments. The laptop computer used for the computation has an Intel Core i7 processor with a 2.80 GHz clock speed and has 4GB of RAM.

Five criteria are used to measure the performances of the improved H-MK-SVM and some other competitive methods. These criteria include the overall hit rate (PCC), the true positive rate (Sensitivity), the true negative rate (Specificity), AUC (the area under the receiver operating characteristic curve) and Lift (Burez and Van den Poel, 2009; Lessmann and Voß, 2008; Verbeke *et al.*, 2011). The LSSVMlab v1.8 toolbox[2] was used for the computation of the AUC. The computational results reported in the following are obtained on the testing sets.

5.1 Results with and without hierarchical ensemble learning

To measure the benefits of using the hierarchical ensemble learning framework, the following five experiments are conducted: (S1) user response modeling based on the hierarchical ensemble learning framework using the external, tag and keyword, and longitudinal individual behavioral data; (S2) user response modeling by the improved H-MK-SVM in Layer 2 using only the longitudinal individual behavioral variable Quantity; (S3) user response modeling by the improved H-MK-SVM in Layer 2 using only the variable Acceptance; (S4) user response modeling by the improved H-MK-SVM in Layer 3 using

---

[2] http://www.esat.kuleuven.be/sita/lssvmlab/

the same data as those under S1; and (S5) user response modeling based on the hierarchical ensemble learning framework without learning the weights of these three types of data, *i.e.*, $\beta_{\hat{m}} = 1$.

Results of these five experiments are reported in Table 3. As shown in Table 3, the H-MK-SVM under S1 obtained the highest PCC (66.50%), Sensitivity (88.50%) and top 10% Lift (9.87), while the H-MK-SVM under S5 obtained the best AUC (72.45%). In comparison with the H-MK-SVM without hierarchical ensemble learning (S2, S3 and S4), the H-MK-SVM based on hierarchical ensemble learning (S1 and S5) obtained higher PCC, Sensitivity, AUC and top 10% Lift. For example, the H-MK-SVM under S1 demonstrated more than 5%, 15%, 10% and 0.5 improvements of the PCC, Sensitivity, AUC and Lift over the H-MK-SVM under S2, S3 and S4, respectively.

Table 3 approximately here

The size n of the training sets is varied to examine its effects on the performance of the improved H-MK-SVM. The performance criteria and the computational time in seconds are reported in Table 4. Results in Table 4 show that the improved H-MK-SVM used the least computational time with n=200, obtained the highest PCC, Sensitivity and Lift with n=450, and obtained the highest AUC and Lift with n=700. The performance of the improved H-MK-SVM with n=450 is almost as good as that with n=700. However, much less computational time is used with n=450 than with n=700. Therefore, the moderate size, *i.e.*, n=450, of the training sets is used in the rest of the computational experiments.

Table 4 approximately here

5.2 Comparisons of performance of the improved H-MK-SVM with other methods

Four competitive methods including the SVM, feed-forward ANN (FFANN), radial basis function neural network (RBFNN) and decision tree (DT) were used in the experiments to compare their performances with that of the improved H-MK-SVM. The Neural Network and the Statistics toolboxes in Matlab 7.4 were used to implement the FFANN, RBFNN and DT. Because these four methods cannot be directly used to model heterogeneous and tensor data, the longitudinal individual behavioral attributes represented by a third-order tensor were aggregated as a matrix $\mathbf{ts} = \{ts_{i\tilde{j}} \mid i = 1, \cdots, n; \tilde{j} = 1, \cdots, m_2\}$. Both the aggregated behavioral attributes and the external attributes, *i.e.*, the composite vector $\mathbf{x}_i = [\mathbf{s}_i, \mathbf{ts}_i]$, were used as inputs in these four methods.

The results of the improved H-MK-SVM and of these four methods are presented in Table 5. Except for the Specificity, the improved H-MK-SVM obtained the highest PCC, Sensitivity, AUC and top 10% Lift.

The AUC is a robust estimator of prediction performance (Lee *et al.*, 2010). As shown in Table 3, the improved H-MK-SVM under S1 demonstrated 4.55%, 4.50%, 4.38% and 15.48% improvements in AUC over the SVM, FFANN, RBFNN and DT, respectively. The Lift is one of the most commonly used measures in direct marketing applications (Cui *et al.*, 2006). The improved H-MK-SVM demonstrated 1.14, 3.51, 0.57 and 2.88 improvements in the top 10% Lift over the SVM, FFANN, RBFNN and DT, respectively. These results show that the improved H-MK-SVM using longitudinal behavioral data outperforms the more traditional methods using aggregated behavioral data. Therefore, using the longitudinal individual behavioral data in the hierarchical ensemble learning framework improves the performance of user response modeling.

Table 5 approximately here

5.3 Effects of varying time lengths and aggregation scales

In this section, the effects of varying time lengths and aggregation scales of the longitudinal individual behavioral data on the prediction performance are examined. The results of the improved H-MK-SVM using the longitudinal individual behavioral data with different time lengths are shown in Table 6. Results in Table 6 show that the improved H-MK-SVM using the longest behavioral data ($T=30$) obtained the highest PCC, Sensitivity, AUC and Lift. Specially, the AUC and Lift of the improved H-MK-SVM using the longest behavioral data are much higher than those using shorter behavioral data. These results show that long enough behavioral data need to be stored in the data warehouse and used in user behavioral analysis.

The longitudinal individual behavioral data are aggregated at different scales to examine the effects of these scales on performance. Specifically, the behavioral data are aggregated per day ($Scale=1$), *i.e.*, no aggregation, per two days ($Scale=2$), per six days ($Scale=6$) and per month ($Scale=30$), respectively. The results of the improved H-MK-SVM using longitudinal individual behavioral data with different aggregation scales are shown in Table 7. As shown in Table 7, the improved H-MK-SVM with $Scale=2$ obtained an AUC 0.95% higher and a Lift 0.1 higher than the improved H-MK-SVM with $Scale=1$. Therefore, it is helpful to select a suitable aggregation scale of the longitudinal data.

Tables 6-7 approximately here

5.4 Effects of incorporating the engagement behavioral data

The performances of the improved H-MK-SVM, as well as the other four competitive methods, incorporating the engagement behavioral data using the AUC, the top 10%, 20%, 30% and 40% Lift as criteria are reported in Table 8. For the four competitive methods, both the longitudinal individual and the engagement behavioral attributes were aggregated as $\mathbf{ts}'=\{ts'_{ij} \mid i=1,\cdots,n;\ j=1,\cdots,m_2+m_4\}$, and the aggregated behavioral attributes and the external attributes, *i.e*., the composite vector $\mathbf{x}_i = (\mathbf{s}_i, \mathbf{ts}'_i)$, were used as inputs. For the results reported in Table 8, only users with followees are selected into the datasets. For comparison purpose, the computational experiments were re-conducted for the improved H-MK-SVM and the other four methods using this dataset but using only the external and individual behavioral data. Results both with and without engagement behavioral data are reported in Table 8 for comparison purpose.

As shown in Table 8, the improved H-MK-SVM using the longitudinal engagement behavioral data obtained the highest AUC and the highest Lift. The improved H-MK-SVM using the engagement behavioral data demonstrated 0.81%, 0.6, 0.05, 0.13 and 0.13 improvements in the AUC, the top 10%, 20%, 30% and 40% Lift, respectively, over the improved H-MK-SVM using only the external and individual behavioral data. These results show that the use of the engagement behavioral data in the improved H-MK-SVM can improve the user response modeling performance. Table 8 also shows that the improved H-MK-SVM using the engagement behavioral data demonstrated 1.52% (1.41), 7.39% (1.11), 5.73% (2.12) and 19.15% (3.09) improvements in the AUC (the top 10% Lift) over the SVM, FFANN, RBFNN and DT, respectively, using the aggregated engagement behavioral data. Therefore, the performance of the improved H-MK-SVM is obviously superior to the other four methods using the aggregated behavioral data.

Table 8 approximately here

6. Conclusions

In this study, a hierarchical ensemble learning framework is developed for behavior-aware user response modeling using diverse heterogeneous data. In the framework, a general-purpose data preprocessing and transformation strategy is proposed to transform the large-scale and multi-relational user data into customer-centered data and to extract prediction attributes. An improved H-MK-SVM is developed to combine the external, tag and keyword, individual behavioral and engagement behavioral data to improve the prediction performance.

Computational experiments are conducted using a real-world microblog database. The experimental results show that (1) the improved H-MK-SVM with hierarchical ensemble learning exhibits superior performance over that without hierarchical ensemble learning; (2) the improved H-MK-SVM using the longitudinal individual behavioral data demonstrates noticeable improvements over the SVM, FFANN, RBFNN and DT; (3) the improved H-MK-SVM using the longitudinal engagement behavioral data demonstrates noticeable improvements over the improved H-MK-SVM using only the external and individual behavioral data; and (4) the improved H-MK-SVM using the longitudinal engagement behavioral data demonstrates considerable improvements over the SVM, FFANN, RBFNN and DT using the aggregated engagement behavioral data. Furthermore, this study investigates the usefulness of selecting a suitable training sample size and selecting a suitable time length and aggregation scale of the longitudinal behavioral data for user response modeling.

The hierarchical ensemble learning framework provides valuable implications of how to integrate diverse heterogeneous user data available in the databases of electronic commerce and social media marketing. Integrating the multi-channel, multi-network, multi-media (text, video and audio) data into CRM and direct marketing models to improve the prediction performance and effectively allocate the marketing resources will be a direction for further research.

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. N110706001).

References:

Bach, F. R., Lanckrient, G. R. G., Jordan, M. I., 2004. 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, pp. 41-48.

Baesens, B., Viaene, S., Van den Poel, D., Vanthienen, J., Dedene, G. (2002). Bayesian neural network learning for repeat purchase modelling in direct marketing. *European Journal of Operational Research*, **138**(1), 191–211.

Ballings, M., Van den Poel, D., 2012. Customer event history for churn prediction: How long is long enough? *Expert Systems with Applications*, **39**, 13517–13522.

Bijmolt, T. H. A., Leeflang, P. S. H., Block, F., Eisenbeiss, M., Hardie, B. G. S., Lemmens A., Saffert, P., 2010. Analytics for customer engagement. *Journal of Service Research*, **13**, 341-356

Bose, I., Chen, X., 2009. Quantitative models for direct marketing: A review from systems perspective. *European Journal of Operational Research*, **195**, 1-16.

Buckinx, W., Moons, E., Van den Poel, D., Wets, G., 2004. Customer-adapted coupon targeting using feature selection. *Expert Systems with Applications*, **26**, 509–518.

Burez, J., Van den Poel, D., 2009. Handling class imbalance in customer churn prediction. *Expert Systems with Applications*, **36**, 4626–4636.

Cao, L., Ou, Y., Yu, P. S., 2012. Coupled behavior analysis with applications. *IEEE Transactions on Knowledge and Data Engineering*, **24**, 1378–1392.

Chapelle, O., Vapnik, V., Bousquet, O., Mukherjee, S., 2002. Choosing multiple parameters for support vector machines. *Machine Learning*, **46**, 131–159.

Chau, M., Xu, J. J., 2012. Business intelligence in Blogs: understanding consumer interactions and communities. *MIS Quarterly*, **36**, 1189-1216.

Chen, T., Li, H., Yang, Q., Yu., Y., 2013. General functional matrix factorization using gradient boosting. In: Proceedings of the 30th International Conference on Machine Learning, Atlanta, Georgia, USA.

Chen, W-C., Hsu, C.-C., Hsu, J.-N., 2011. Optimal selection of potential customer range through the union sequential pattern by using a response model. *Expert Systems with Applications*, **38**, 7451–7461.

Chen, Y., Liu, Z., Ji, D., Xin, Y., Wang, W., Yao, L., Zou, Y., 2012. Context-aware ensemble of multifaceted factorization models for recommendation prediction in social networks. In KDD-Cup Workshop, Beijing, China.

Chen, Z. Y., Li, J. P., Wei, L. W., 2007. A multiple kernel support vector machine scheme for feature selection and rule extraction from gene expression data of cancer tissue. *Artificial Intelligence in Medicine*, **41**, 161–175.

Chen, Z. Y., Fan, Z. P., Sun, M., 2012. A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data. *European Journal of Operational Research*, **223**, 461–472.

Cheung, C. M. K., Thadani, D. R., 2012. The impact of electronic word-of-mouth communication: A literature analysis and integrative model. *Decision Support Systems*, **54**(1), 461–470.

Cho, Y. B., Cho, Y. H., Kim, S. H., 2005. Mining changes in customer buying behavior for collaborative recommendations. *Expert Systems with Applications*, **28**, 359–369.

Choi, K., Yoo, D., Kim, G., Suh, Y., 2012. A hybrid online-product recommendation system: Combining implicit rating-based collaborative filtering and sequential pattern analysis. *Electronic Commerce Research and Applications*, **11**, 309–317.

Crone, S. F., Lessmann, S., Stahlbock, R., 2006. The impact of preprocessing on data mining: An evaluation of classifier sensitivity in direct marketing. *European Journal of Operational Research*, **173**, 781–800.

Cui, G., Wong, M. L., Lui, H.-K., 2006. Machine learning for direct marketing response models: Bayesian networks with evolutionary programming. *Management Science*, **52**, 597–612.

Cui, G., Wong M. L., Zhang, G., 2010. Bayesian variable selection for binary response models and direct marketing forecasting. *Expert Systems with Applications*, **37**, 7656–7662.

Dellarocas, C., 2003. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*, **49**, 1407–1424.

Gönen, M., Alpaydın, E., 2011. Multiple kernel learning algorithms. *Journal of Machine Learning Research*, **12**, 2211–2268.

Ha, K., Cho, S., MacLachlan, D., 2005. Response models based on bagging neural networks. *Journal of Interactive Marketing*, **19**, 17–30.

Hsu, C. W., Chang, C. C., Lin, C. J., 2003. A practical guide to support vector classification. http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf.

Huang, C. -L., Huang, W. -L., 2009. Handling sequential pattern decay: Developing a two-stage collaborative recommender system. *Electronic Commerce Research and Applications*, **8**, 117–129.

Kang, P., Cho, S., MacLachlan, D. L., 2012. Improved response modeling based on clustering, under-sampling, and ensemble. *Expert Systems with Applications*, **39**, 6738–6753.

Kaefer, F., Heilman, C. M., Ramenofsky, S. D., 2005. A neural network application to customer classification to improve the timing of direct marketing activities. *Computers & Operations Research*, **32**, 2595–2615.

Kaplan, A. M., Haenlein, M., 2010. Users of the world, unite! the challenges and opportunities of social media. *Business Horizons*, **53**, 59-68.

Kim, Y., Street, W. N., Russell, G. J., Menczer, F., 2005. Customer targeting: A neural network approach guided by genetic algorithms. *Management Science*, **51**, 264–276.

Kolda, T., Bader, B., 2009. Tensor decompositions and applications. *SIAM review*, **51**, 455–500.

Lanckrient, G. R. G., Cristianini, N., Bartlett, P., El Ghaoui, L., Jordan, M. I., 2004. Learning the kernel matrix with semidefinite programming. *Journal of Machine Learning Research*, **5**, 27–72.

Lee, H., Shin, H., Hwang, S., Cho, S., MacLachlan, D., 2010. Semi-supervised response modeling. *Journal of Interactive Marketing*, **24**, 42–54.

Lessmann, S., Voß, S., 2008. Supervised classification for decision support in customer relationship management. In: Bortfeldt, A., Homberger, J., Kopfer, H., Pankratz, G., Stangmeier, R. (Eds.), *Intelligent Decision Support*, Gabler, Wiesbaden, pp. 231–253.

Lessmann, S., Voß, S., 2009. A reference model for customer-centric data mining with support vector machines. *European Journal of Operational Research*, **199**, 520–530.

Li, Y. -M., Li, T. -Y., 2013. Deriving market intelligence from Microblogs. *Decision Support Systems*, doi: 10.1016/j.dss.2013.01.023.

Liu, D. -R., Lai, C. -H., Lee W. -J., 2009. A hybrid of sequential rules and collaborative filtering for product recommendation. *Information Sciences*, **179**, 3505–3519.

Mangold, W. G., Faulds, D. J., 2009. Social media: The new hybrid element of the promotion mix. *Business Horizons*, **52**, 357–365.

Min, S. H., Han, I., 2005. Detection of the customer time-variant pattern for improving recommender systems. *Expert Systems with Applications*, **28**, 189–199.

Park, Y. -J., Chang, K. -N., 2009. Individual and group behavior-based customer profile model for personalized product recommendation. *Expert Systems with Applications*, **36**, 1932–1939.

Power, D. J., Phillips-Wren, G., 2011. Impact of social media and Web 2.0 on decision-making. *Journal of*

Decision Systems, **20**, 249-261.

Prinzie, A., Van den Poel, D., 2006. Investigating purchasing sequence patterns for financial services using Markov, MTD and MTDg models. *European Journal of Operational Research*, **170**, 710–734.

Prinzie, A., Van den Poel, D., 2007. Predicting home-appliance acquisition sequences: Markov/Markov for discrimination and survival analysis for modelling sequential information in NPTB models. *Decision Support Systems*, **44**, 28–45.

Prinzie, A., Van den Poel, D., 2011. Modeling complex longitudinal consumer behavior with dynamic Bayesian networks: an acquisition pattern analysis application. *Journal of Intelligent Information System*, **36**(3), 283–304.

Shin, H., Cho, S., 2006. Response modeling with support vector machines. *Expert Systems with Applications*, **30**, 746–760.

Vapnik, V. N., 1998. *Statistic learning theory*. Wiley, New York.

van Doorn, J., Lemon, K., Mittal, V., Nass, S., Pick, D., Pimer, P., Verhoef, P., 2010. Customer engagement behavior: Theoretical foundations and research directions. *Journal of Service Research*, **13**, 253-266.

Verbeke, W., Dejaeger, K., Martens, D., Hur, J., Baesens, B., 2011. New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *European Journal of Operational Research*, **218**, 211–229.

Zhang, X., Zhu J., Xu, S., Wan, Y., 2012. Predicting customer churn through interpersonal influence. *Knowledge-Based Systems*, **28**, 97–104.

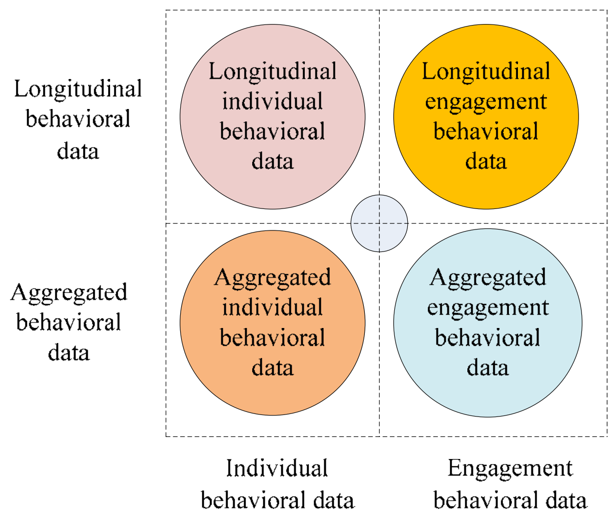

Figure 1. Different types of behavioral data

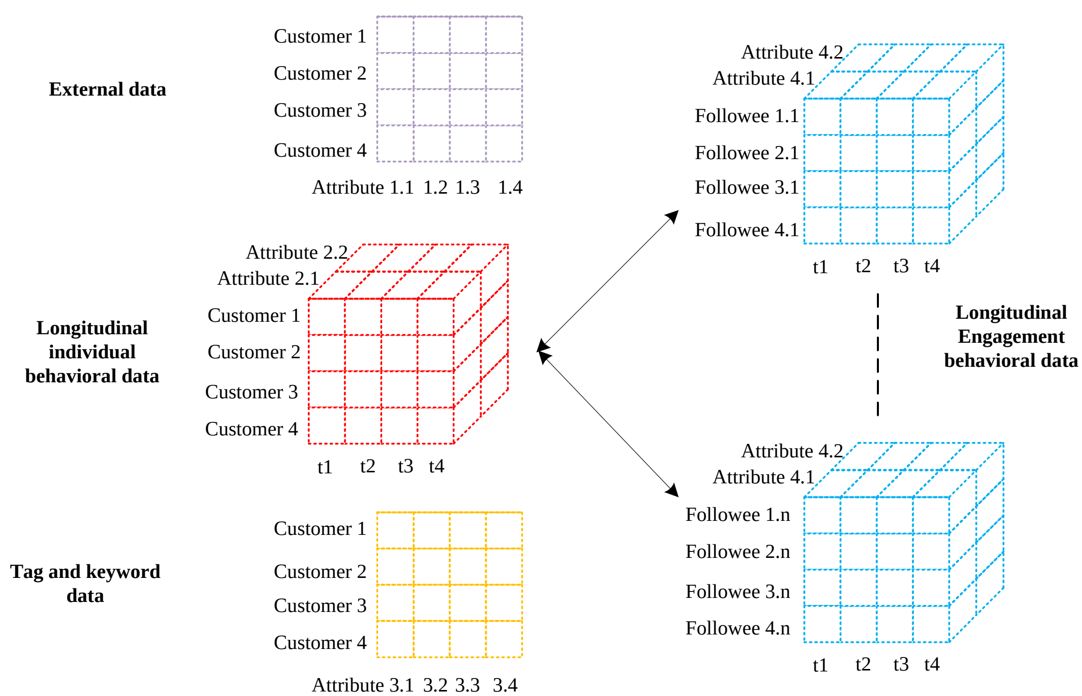

Figure 2. Illustration of diverse heterogeneous user data

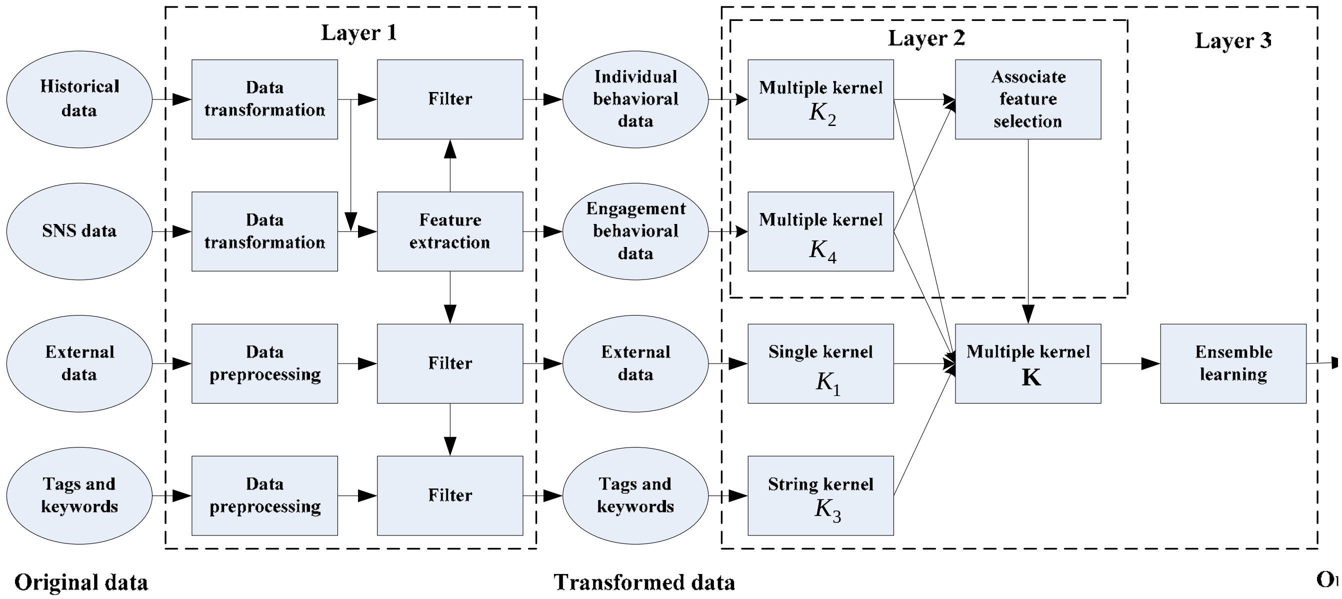

Figure 3. The proposed hierarchical ensemble learning framework for behavior-aware user response modeling in social media

Table 1. Characteristics of the original database

Datasets	Records	Customer-centered	Dimensions	Numerical	String	Timestamp
Training	73,209,277	No	4	Yes	No	Yes
User Profile	2,320,895	Yes	5	Yes	Yes	No
Item	6,095	No	3	No	Yes	No
User SNS	50,655,143	No	2	No	No	No

Table 2. Characteristics of the transformed customer-centered datasets

Datasets	n	$n_{positive}$	$n_{negtive}$	m_1	m_2	m_3	m_4	T
1	22,548	1,979	20,569	2	0	0	0	0
2	22,548	1,979	20,569	0	0	10	0	0
3	22,548	1,979	20,569	0	2	0	0	30
4	9,105	875	8,230	0	0	0	2	30

Table 3. Results of the improved H-MK-SVM with and without hierarchical ensemble learning

Models	PCC	Sensitivity	Specificity	AUC	10% Lift
H-MK-SVM (S1)	66.50	81.50	51.50	72.06	9.87
H-MK-SVM (S2)	59.75	30.67	88.83	59.73	9.21
H-MK-SVM (S3)	60.17	30.50	89.83	59.84	9.30
H-MK-SVM (S4)	60.83	66.17	55.50	61.94	8.45
H-MK-SVM (S5)	65.83	75.67	56.00	72.45	9.78

Table 4. Results of the improved H-MK-SVM with varying sizes of the training set

n	PCC	Sensitivity	Specificity	AUC	Lift	CPU time (s)
200	65.33	75.50	55.17	65.84	7.79	7.77
450	66.50	81.50	51.50	72.06	9.87	126.89
700	66.08	78.00	54.17	72.58	9.87	922.66

Table 5. Results of the improved H-MK-SVM and other competitive methods

Models	PCC	Sensitivity	Specificity	AUC	10% Lift
H-MK-SVM	66.50	81.50	51.50	72.06	9.87
SVM	65.92	77.50	54.33	67.51	8.73
FFANN	53.83	11.17	96.50	67.56	6.36
RBFNN	57.58	19.50	95.67	67.68	9.30
DT	57.25	46.83	67.67	56.58	6.99

Table 6. Results of the improved H-MK-SVM using longitudinal individual behavioral data with varying lengths

T	PCC	Sensitivity	Specificity	AUC	Lift
30	66.50	81.50	51.50	72.06	9.87
15	64.08	74.33	53.83	65.83	8.17
5	65.75	77.50	54.00	66.89	8.45
1	62.50	61.17	63.83	64.18	7.60

Table 7. Results of the improved H-MK-SVM using longitudinal individual behavioral data with varying aggregation scales

Scale	*T*	PCC	Sensitivity	Specificity	AUC	Lift
1	30	66.50	81.50	51.50	72.06	9.87
2	15	66.33	80.33	35.67	73.01	9.97
6	5	65.92	77.50	54.33	67.30	8.36
30	1	62.25	69.67	54.83	62.46	7.69

Table 8. Comparisons of the results of different classifiers with and without engagement behavioral data

Methods	With engagement data					Without engagement data				
	AUC	10%Lift	20%Lift	30%Lift	40%Lift	AUC	10%Lift	20%Lift	30%Lift	40%Lift
H-MK-SVM	72.74	7.17	6.43	5.82	5.92	71.93	6.57	6.38	5.69	5.79
SVM	71.22	5.76	6.03	5.42	5.04	69.59	6.26	5.48	4.95	5.04
FANN	65.35	6.06	5.33	3.55	2.66	70.49	6.36	6.38	5.12	3.84
RBFNN	67.01	5.05	2.51	1.67	1.25	68.38	4.14	2.06	1.37	1.03
DT	53.59	4.08	4.08	4.15	3.51	58.33	4.60	4.62	4.28	3.61