A MOVIE RECOMMENDER SYSTEM BASED ON ENSEMBLE OF TRANSDUCTIVE SVM CLASSIFIERS

Aristomenis S. Lampropoulos, Paraskevi S. Lampropoulou and George A. Tsihrintzis
Department of Informatics, University of Piraeus, 80 Karaoli and Dimitriou St, 18534 Piraeus, Greece

Keywords: Transductive SVM, Recommender system, Ensemble of classifiers.

Abstract: In this paper, we address the recommendation process as a classification problem based on content features and a bank of Transductive SVM classifiers that capture user preferences. Specifically, we develop an ensemble of Transductive SVM (TSVM) classifiers, each of which utilizes a different feature vector extracted from different semantic meta-data such as actors, directors, writers, editors and genres. The ensemble classifier allows our system to utilize feature vectors of meta-data from a database and to make personalized recommendations to users. This is achieved through the property of TSVM classifiers to utilize a large amount of available unlabeled data together with a small amount of labeled data that constitute the rated movies of a user. The proposed method is compared to a TSVM classifier which utilizes a feature vector extracted from only ratings of users. The experimental results based on the MovieLens data set indicated that our classifier based on an ensemble of TSVM with content meta-data yield higher accuracy recommendations when compared to the TSVM classifier that utilized only user ratings.

1 INTRODUCTION

The huge quantities of information that are available online to broad classes of computer users often result in the users facing difficulties or lacking the knowledge to make efficient use of the information. In turn, this has led to the need for systems that have the ability to identify user needs automatically and help users to choose appropriate sets of files from those available to them. Such systems are known as *recommender systems* and, somehow, represent a process similar to the social process of recommendation. Recommender systems help relieve some of the pressure of information overload by taking into account personal needs and interests of users and by providing information in the most appropriate and valuable way.

During the recent years, recommender systems have received a lot of attention from several research groups worldwide and have distinguished themselves from simple search engines and retrieval systems. The main difference between a recommender system and a search engine or a retrieval system is that a recommender system not only returns results, but also selects objects (items) that satisfy the specific querying user's needs. Thus, recommender systems must be equipped with an individualization/personalization process of the results they return to their user. Ultimately, recommender systems attempt to predict items that a user might be interested in.

In work of ours, we present a movie recommender system which is trained with a small number of examples of user-preferred movies. The system computes features that are automatically extracted from semantic content about actors, directors, genres etc. which are provided by the IMDB database (IMDB, 2010). Therefore, our system makes recommendations on a personalized basis, i.e., without having to match the user's interests to some other user's interests. In this way, our system overcomes well-known problems associated with Collaborative Filtering, such as non-association or user bias[1].

More specifically, the paper is organized as follows: in Section 2, we present an overview of related work on movie recommendation systems. In Section 3, we present briefly the Transductive Inference Paradigm. In Section 4, we formulate the recommendation problem as an ensemble of Transduc-

[1]The problem of non-association arises when two similar items have never been wanted by the same user, their relationship is not known explicitly or, in item based Collaboration Filtering, those two items cannot be classified into the same group. On the other hand, the problem of user bias may be present in past ratings.

S. Lampropoulos A., S. Lampropoulou P. and A. Tsihrintzis G..
A MOVIE RECOMMENDER SYSTEM BASED ON ENSEMBLE OF TRANSDUCTIVE SVM CLASSIFIERS.
DOI: 10.5220/0003682802420247
In *Proceedings of the International Conference on Neural Computation Theory and Applications* (NCTA-2011), pages 242-247
ISBN: 978-989-8425-84-3

tive SVM classifiers. We evaluate recommendation methods and present experimental results. Finally, in Section 5, we draw conclusions and point to future related work.

2 RELATED WORK

In general, the recommendation problem refers to methods for selecting and suggesting items to a user. These methods attempt to enhance ratings given by users to items and information that describes or characterizes users or items. There are three main approaches to recommender systems:

- Content-based,
- Collaborative filtering, and
- Hybrid.

Modern information systems embed the ability to monitor and analyze users' actions to determine the best way to interact with them. Ideally, each user's actions are logged separately and analyzed to generate an individual user profile. All the information about a user, extracted either by monitoring user actions or by examining the objects the user has evaluated (Burke, 2002), is stored or utilized to customize services offered. This user modeling approach is known as content-based learning. The main assumption behind it is that a user's behavior remains unchanged through time; therefore, the content of past user actions may be used to predict the desired content of their future actions (Mooney and Roy, 2000). Therefore, in content-based recommendation methods, the rating $R(u,i)$ of the item i by the user u is typically estimated based on ratings assigned by user u to the items in a subset I_n of the full set of items I that are "similar" to item i in terms of their content as defined by their associated features.

To be able to search through a collection of items and make observations about the similarity between objects that are not directly comparable, we must first transform raw data at a certain level of information granularity. Information granules refer to a collection of data that contain only information essential to the recommendation process. Such granulation allows more efficient processing for extracting features and computing numerical representations that characterize an item. As a result, the large amount of detailed information of one item is reduced to a limited set of features. Each feature captures some aspects of the item that are essential and sufficient to determine item similarity.

Collaborative filtering is based on collecting ratings for items, comparing commonalities between users (or items) on the basis of their ratings, and finally producing recommended items according to inter-user (or inter-item) comparisons. The problem space of collaborative filtering can be formulated as a matrix of users versus items, with each cell representing a user's rating on specific item (Schafer et al., 2007; Herlocker et al., 2000). In (Sarwar et al., 2001), an automated collaborative filtering algorithm is presented to generate movie recommendations. An comparative evaluation of collaborative filtering methods are presented in (Herlocker et al., 2004).

Hybrid methods combine two or more recommendation techniques to achieve better performance and to address drawbacks of each non-hybrid techniques. Usually, collaborative filtering methods are combined with content-based methods. There are several different ways of combining these two separate systems. Hybrid recommender systems for movies are presented in (Christakou and Stafylopatis, 2005), (Mukherjee et al., 2003). In (Christakou and Stafylopatis, 2005), a hybrid approach, based on Multilayer Perceptron neural networks combined with collaborative information, is used to construct a recommender system for movies.

More generally, there are four groups of hybrid methods according to the combination of content-based and collaborative methods.

In the Weighted Hybridization Method, the outputs (ratings) acquired by individual recommender systems are combined together to produce a single final recommendation using either linear combination (Claypool et al., 1999) or a voting scheme (Pazzani, 1999). In Switched Hybridization, the system switches between recommendation techniques selecting the method that gives better recommendations for the current situation depending on some recommendation "quality" metric (Billsus and Pazzani, 2000). Finally, the Cascade Hybridization recommendation technique can be analyzed into two sequential stages. The first stage (content-based method/collaborative) selects intermediate recommendations. Then, the second stage (collaborative/content-based method) selects appropriate items from the recommendations of the first stage. This method is more efficient than the weighted hybridization method which applies all of its techniques on all items. The computational burden of this hybrid approach is relatively small because recommendation candidates in the second level are partially eliminated during the first level. Moreover, this method is more tolerant to noise in the operation of low-priority recommendations, since ratings of the high level recommender can only be refined, but never over-turned (Burke, 2007; Lampropoulos et al., 2011).

3 TRANSDUCTIVE INFERENCE PARADIGM

Vladimir Vapnik proposed the *Transductive Inference Paradigm* (Vapnik, 1982) as the next step beyond the previously-proposed *Model Prediction Paradigm*. The key ideas behind the transductive inference paradigm arose from the need to create efficient methods of inference from small sample sizes. Specifically, in transductive inference an effort is made to estimate the values of an unknown predictive function at a given restricted subset of its domain in which we are interested and not in the entire domain of its definition. This led Vapnik to formulate the *Main Principle* (Vapnik, 1982), (Vapnik, 1998), (Cherkassky and Mulier, 2007):

"*If you possess a restricted amount of information for solving some problem, try to solve the problem directly and never solve a more general problem as an intermediate step. It is possible that the available information is sufficient for a direct solution, but may be insufficient to solve a more general intermediate problem.*"

The main principle constitutes the essential difference between newer approaches and the classical paradigm of statistical inference based on the estimation of a number of free parameters. While the classical paradigm is useful in simple problems that can be analyzed with few variables, real world problems are much more complex and require large numbers of variables. Thus, the goal when dealing with real life problems that are by nature of high dimensionality is to define a less demanding problem which admits well-posed solutions. This fact involves finding values of the unknown function reasonably well only at given points of interest, while outside of the given points of interest that function may not be well-estimated.

The paradigm of transductive inference forms a solution that derives results directly from particular (training samples) to particular (testing samples).

A simple form transductive inference method could be considered the k-nearest neighbor method (k-NN), where a new data vector is classified into one of the existing classes in the data samples based on the majority of classes among k nearest to the new vector samples. The distance is measured with the use of a similarity measure e.g. Euclidean distance.

In many problems, we do not care about finding a specific function with good generalization ability, but rather are interested in classifying a given set of examples (i.e. a test set of data) with minimum possible error. For this this reason, the inductive formulation of the learning problem is unnecessarily complex.

Transductive inference embeds the unlabeled (test) data in the decision making process that will be responsible for their final classification. Transductive inference "works because the test set can give you a non-trivial factorization of the (discrimination) function class" (Chapelle et al., 2006). Additionally, the unlabeled examples provide information on the prior information of the labeled examples and "guide the linear boundary away from the dense region of labeled examples" (Zhu, 2008).

For a given set of labeled data points $Labeled-Set = \{(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)\}$, with $y_i \in \{-1,1\}$ and a set of test data points $Unlabaled-Set = \{x_{n+1},x_{n+2},\ldots,x_{n+k}\}$, where $x_i \in R^d$, transduction seeks among the feasible corresponding labels the one $y^*_{n+1},y^*_{n+2},\ldots,y^*_{n+k}$ that has the minimum number of errors.

Also, transduction would be useful among other ways of inference in which there are either a small amount of labeled data points available or the cost for annotating data points is prohibitive. Hence, the use of the Empirical Risk Minimization(ERM) principle helps in selections of the "best function from the set of indicator functions defined in R^d, while transductive inference targets only the functions defined on the working set $Working-Set = Labeled-Set \bigcup Unlabeled-Set$," which is a discrete space.

The goal of inductive learning is to generalize for any future test set, while the goal of transductive inference is to make predictions for a specific working set. In inductive inference, the error probability is not meaningful when the prediction rule is updated very abruptly and the data point may be not independently and identically distributed, as, for example, in data streaming. On the contrary, Vapnik (Vapnik, 1998) illustrated that the results from transductive inference are accurate even when the data points of interest and the training data are not independently and identically distributed. Therefore, the predictive power of transductive inference can be estimated at any time instance in a data stream for both future and previously observed data points that are not independently and identically distributed. In particular, *empirical findings suggest that transductive inference is more suitable than inductive inference for problems with small training sets and large test sets* (Zhu, 2008).

In this paper we follow the approach based on Transductive learning and SVM presented in (Joachims, 1999), (Joachims, 2008) where it is utilized a TSVM approach for text classification.

More specifically, the TSVM can be viewed as a standard SVM with an extra regularization term defined over the set of unlabeled data. The goal of TSVM is to construct a function f that maximizes the separation between $Labeled - Set$ and $Unlabeled - Set$. The decision function has the following form:

$$f(x) = w \cdot \Phi(x) + b \quad (1)$$

where w, b are the parameters of the model and $\Phi(\cdot)$ is the mapping function from the input space to some other higher dimension space where the data are linearly separable.

The TSVM solves the following optimization problem:

$$\min \frac{1}{2}\|w\|^2 + C \cdot \sum_{i=1}^{n} L(y_i, f(x_i)) + C^* \cdot \sum_{i=n+1}^{n+k} (L\,|\,f(x_i)|) \quad (2)$$

where $L(\cdot)$ is the loss function for labeled data and C, C^* are adjustable parameters.

4 PROPOSED RECOMMENDATION METHOD

In our proposed approach, we improve the performance of a movie recommendation process by utilizing an ensemble of TSVM classifiers. Each of these classifiers utilizes a feature vector from a specific kind of meta-information such as genres, actors, writers and directors. The architecture of our approach is presented in Fig. 1.

We treat our recommendation process as a classification problem where each movie that was rated with a positive rating degree of 4-5 stars belongs to the class of $Labeled - Set$ while movies that rated in the range of 0-3 stars is considered as data from the class of $Unlabeled - Set$. As it is well known, (Kuncheva, 2004) an ensemble of classifiers have proved to improve classification performance in many applications. More specifically, the combination of classifiers achieve better performance than the best single classifier when these classifiers are diverse. As is presented in (Kuncheva, 2004), diversity can be achieved for example by combining classifiers that utilized different feature spaces. Consequently, in this paper we follow this important remark and we use a simple majority voting rule (Kuncheva, 2004) to combine a bank of TSVM classifiers each of them works on different feature vectors extracted from different semantic meta-data.

We compare our ensemble of TSVM classifiers with a TSVM classifier which works on feature vectors constructed by ratings of users on a movie. More specifically, for each item (movie) we used a corresponding feature vector coming from the ratings assigned to this by other users. In other words, each

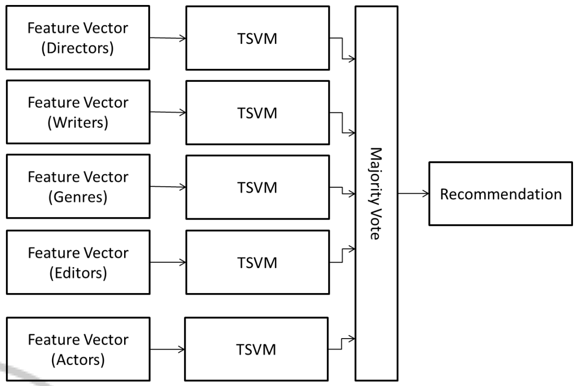

Figure 1: Proposed Recommendation Method.

movie was represented by a feature vector of 942 features equal in number to the number of the remaining 942 users of our dataset, in which the ratings of the active user are not taken into account.

Finally, we examined an ensemble of TSVM classifiers, where we aggregated both classifiers based on content-based features and classifier based on ratings of users.

System Evaluation. In order to illustrate the performance of our recommender method, we utilized the publicly available MovieLens dataset provided by GroupLens project. The MovieLens dataset (MovieLens, 2010), consists of 100,000 ratings which were assigned by 943 users on 1682 movies. Ratings are values from the set $\{1,2,3,4,5\}$, with each user having provided ratings for at least 20 movies.

Content features were derived from IMDB database, an off-line version of which is available at (IMDB, 2010). We reconstructed the relational model of IMDB database into Mysql RDBMS with the use of the JMDB tool (JMDB, 2009). We synchronized the MovieLens dataset with the IMDB database and got a set of 1040 movies. For each of these 1040 movies, we extracted four different feature vectors.

Specifically:

- feature vector of 1526 actors (size: 1 x 1526).
- feature vector of 529 writers (size: 1 x 529).
- feature vector of 205 directors (size: 1 x 205).
- feature vector of 19 genres (size: 1 x 19).

For each of the 943 users, we trained an ensemble of TSVM classifiers using semantic-content feature vectors, a TSVM classifier based on rating feature vectors and an ensemble based on a combination of the previous classifiers. For the evaluation of the various classifiers, we followed a 10-fold cross validation on the labels of each user, where the available labels have been randomly split into a training set (90%) and a test set (10%). The results in terms of classification accuracy are presented in Table 1.

Table 1: Classification Accuracy.

Ensemble of TSVM CB features	67.7%
TSVM Rating features	62.5%
Ensemble of TSVM CB and Rating features	68.3%

As presented in Table 1, the ensemble of TSVM classifiers based on content-based features yielded to higher performance than the performance of classifiers based on feature vectors constructed by the ratings of other users. In other words, the content-based semantic information can describe more efficiently the preferences of users than the opinion of other users for a specific item. Finally, the ensemble of TSVM classifiers based on the aggregation of all available features, improves slightly the accuracy of the ensemble with only content-based feature vectors.

5 CONCLUSIONS AND FUTURE WORK

In this paper, we addressed the movie recommendation process as a classification problem. Specifically, we followed an approach based on an ensemble of classifiers, each of which was fed with different feature vectors extracted from different sematic information about movie. Each classifier was based on Transductive Support Vector Machines which enhances their ability to embed unlabeled data in the decision making process and results in better performance when the available datasets are highly unbalanced. Our recommendation method has been evaluated on the MovieLens dataset. We found that the content-based semantic information can describe more efficiently the preferences of users rather than the opinion of other users, represented as ratings of items.

Currently, we are in the process of conducting further experiments and improvements to our system by extending the proposed method into a hybrid cascade recommender system (Lampropoulos et al., 2011) and by applying different types of classifiers (Lampropoulos et al., 2010). This and other related research work is currently in progress and will be reported elsewhere in the near future.

REFERENCES

Billsus, D. and Pazzani, M. J. (2000). User modeling for adaptive news access. *User Modeling and User-Adapted Interaction*, 10(2-3):147–180.

Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4):331–370.

Burke, R. (2007). Hybrid web recommender systems. In Brusilovsky, P., Kobsa, A., and Nejdl, W., editors, *The adaptive web*, pages 377–408. Springer-Verlag, Berlin, Heidelberg.

Chapelle, O., Schölkopf, B., and Zien, A. (2006). *Semi-Supervised Learning*. The MIT Press, Cambridge, Massachusetts, London, England.

Cherkassky, V. and Mulier, F. M. (2007). *Learning from Data: Concepts, Theory, and Methods*. Wiley-IEEE Press.

Christakou, C. and Stafylopatis, A. (2005). A hybrid movie recommender system based on neural networks. In *Proceedings of the 5th International Conference on Intelligent Systems Design and Applications*, ISDA '05, pages 500–505, Washington, DC, USA. IEEE Computer Society.

Claypool, M., Gokhale, A., Miranda, T., Murnikov, P., Netes, D., and Sartin, M. (1999). Combining content-based and collaborative filters in an online newspaper. In *Proc. ACM SIGIR Workshop on Recommender Systems*.

Herlocker, J. L., Konstan, J. A., and Riedl, J. (2000). Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, CSCW '00, pages 241–250, New York, NY, USA. ACM.

Herlocker, J. L., Konstan, J. A., Terveen, L. G., and Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.*, 22:5–53.

IMDB (2010). The internet movie database. Database available at http://www.imdb.com/interfaces#plain.

JMDB (2009). Java movie database. Software available at http://www.jmdb.de/.

Joachims, T. (1999). Transductive inference for text classification using support vector machines. In *Proceedings of the Sixteenth International Conference on Machine Learning*, ICML '99, pages 200–209, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Joachims, T. (2008). Svmlight: An implementation of support vector machines. Software available at http://svmlight.joachims.org/.

Kuncheva, L. I. (2004). *Combining Pattern Classifiers Methods and Algorithms*. Wiley, New York, NY, USA.

Lampropoulos, A. S., Lampropoulou, P. S., and Tsihrintzis, G. A. (2011). A cascade-hybrid music recommender system based on musical genre classification and personality diagnosis for mobile services. *Multimedia Tools and Applications*, pages 1–18. 10.1007/s11042-011-0742-0.

Lampropoulos, A. S., Sotiropoulos, D. N., and Tsihrintzis, G. A. (2010). A music recommender based on artificial immune systems. In Howlett, R. J., Jain, L. C., Tsihrintzis, G. A., Damiani, E., Virvou, M., Howlett, R. J., and Jain, L. C., editors, *Intelligent Interactive Multimedia Systems and Services*, volume 6 of *Smart Innovation, Systems and Technologies*, pages 167–179. Springer Berlin Heidelberg.

Mooney, R. J. and Roy, L. (2000). Content-based book recommending using learning for text categorization. In *Proceedings of the fifth ACM conference on Digital libraries*, DL '00, pages 195–204, New York, NY, USA. ACM.

MovieLens (2010). Movielens data sets. Dataset available at http://www.grouplens.org/node/73.

Mukherjee, R., Sajja, N., and Sen, S. (2003). A movie recommendation system an application of voting theory in user modeling. *User Modeling and User-Adapted Interaction*, 13:5–33.

Pazzani, M. J. (1999). A framework for collaborative, content-based and demographic filtering. *Artificial Intelligence Review*, 13(5-6):393–408.

Sarwar, B., Karypis, G., Konstan, J., and Reidl, J. (2001). Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, WWW '01, pages 285–295, New York, NY, USA. ACM.

Schafer, J. B., Frankowski, D., Herlocker, J., and Sen, S. (2007). Collaborative filtering recommender systems. In Brusilovsky, P., Kobsa, A., and Nejdl, W., editors, *The adaptive web*, pages 291–324. Springer-Verlag, Berlin, Heidelberg.

Vapnik, V. N. (1982). *Estimation of Dependences Based on Empirical Data: Springer Series in Statistics*. Springer, Secaucus, NJ, USA.

Vapnik, V. N. (1998). *Statistical Learning Theory*. Wiley, New York, NY, USA.

Zhu, X. (2008). Semi-supervised learning literature survey. Technical report, University of Winsconsin, Department of Computer Science.