

Research Article

Cross-Domain Personalized Learning Resources Recommendation Method

Long Wang,¹ Zhiyong Zeng,² Ruizhi Li,² and Hua Pang³

¹ College of Information, Liaoning University, Shenyang 110036, China

² School of Computer Science and Information Technology, Northeast Normal University, Changchun 130024, China

³ College of Education Technology, Shenyang Normal University, Shenyang 110034, China

Correspondence should be addressed to Hua Pang; 603665868@qq.com

Received 26 June 2013; Accepted 3 September 2013

Academic Editor: William Guo

Copyright © 2013 Long Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

According to cross-domain personalized learning resources recommendation, a new personalized learning resources recommendation method is presented in this paper. Firstly, the cross-domain learning resources recommendation model is given. Then, a method of personalized information extraction from web logs is designed by making use of mixed interest measure which is presented in this paper. Finally, a learning resources recommendation algorithm based on transfer learning technology is presented. A time function and the weight constraint of wrong classified samples can be added to the classic TrAdaBoost algorithm. Through the time function, the importance of samples date can be distinguished. The weight constraint can be used to avoid the samples having too big or too small weight. So the Accuracy and the efficiency of algorithm are improved. Experiments on the real world dataset show that the proposed method could improve the quality and efficiency of learning resources recommendation services effectively.

1. Introduction

With the continuous development of network technology and educational informatization, the network learning system is widely applied in every stage of education. How to improve the intelligence of the network learning system and the users' efficiency has become a common concern for all researchers. It is more and more difficult for students to look for their interested resources from the increasing learning resources in the network learning system. So adding learning resources recommendation service to the network learning system can save the students' time and effort on information searching and can free the students from enormous network information resources. Under these conditions and needs, the personalized learning resources recommendation service technique has been gradually developed.

The personalized learning resources recommendation service refers to predicting whether the new learning resources can meet the students' personalized demand which is extracted from the previous learning information and recommending the interested resources to the students.

In order to obtain good recommendation results, a certain amounts of personalized information should be accumulated. The accumulation time is proportional to the student's learning time in the network learning system. Shortening the accumulation time is an effective method to improve learning resources recommendation service efficiency. The cross-domain learning resources recommendation is an effective method to shorten the process of accumulating personalized information.

In this paper, a cross-domain personalized learning resources recommendation method is presented. First, a cross-domain personalized learning resources recommendation service model is introduced. Next, a personalized information mining method is designed. Then, a cross-domain learning resources recommendation algorithm is presented. Finally, the simulation results and the experiment analysis are given.

2. Service Model

There are large amount of learning resources in network learning system. Generally, the student does not know if

it contains the interested contents when he studies new resource. So, he can only check the content or the resource description. If it does not contain the interested content, the checking is useless. In order to reduce the useless checking, the recommendation service should be added to the system. For improving the service efficiency, a cross-domain personalized recommendation service model is presented in this paper. The service model is shown in Figure 1.

As shown in Figure 1, when students use the network learning system, some web logs data are generated. These data include user name, access date, access times, access time, and requested URL. Through mining the data, the personalized information of students are obtained. Put the personalized information as a basis for personalized learning resources recommendation into the student personalized information database. When a student learns some subject, the service can extract learning information from current subject and the subjects learned before based on his personalized information, and train the recommendation model. Through the recommendation model, the service can find the learning resources which the student is interested in the current subject and recommend them to the student.

3. Personalized Information Mining Based on Mixed Interest Measure

For a student, what learning resources they are interested in is the most easy and effective way to reflect his personalized demands. From the web logs, the service can find out the learning resources that the student has browsed, analyze what they are interested in, what they are uninterested in, and then compose the personalized information of the student.

Now, browsing interest measure [1] is widely used to estimate whether the users are interested in the browsed contents or not. The browsing interest measure P is computed as:

$$P = \frac{C \times T}{B}, \quad (1)$$

where C is the browsing times, T is the each browsing time, and B is the number of bytes of the browsing content.

In network learning system, the file types of learning resources are not only text but also image, animation, audio, and video. In formula (1), the difference from different file types is not considered. In addition, the learning resources not only include the resource content but also include the resource description which is called the brief of resource. According to the characteristics of learning resources, the mixed interest measure I is computed as follow.

First, the content browsing interest measure P^c is computed as the following:

$$P^c = \eta \times \alpha \times \sum_{i=1}^n \frac{T_i^c}{B^c}, \quad (2)$$

where n is the browsing times, T^c is the each content browsing time, B^c is the number of bytes of the resource content, α is the coefficient of resource content, and η is the coefficient of resource type.

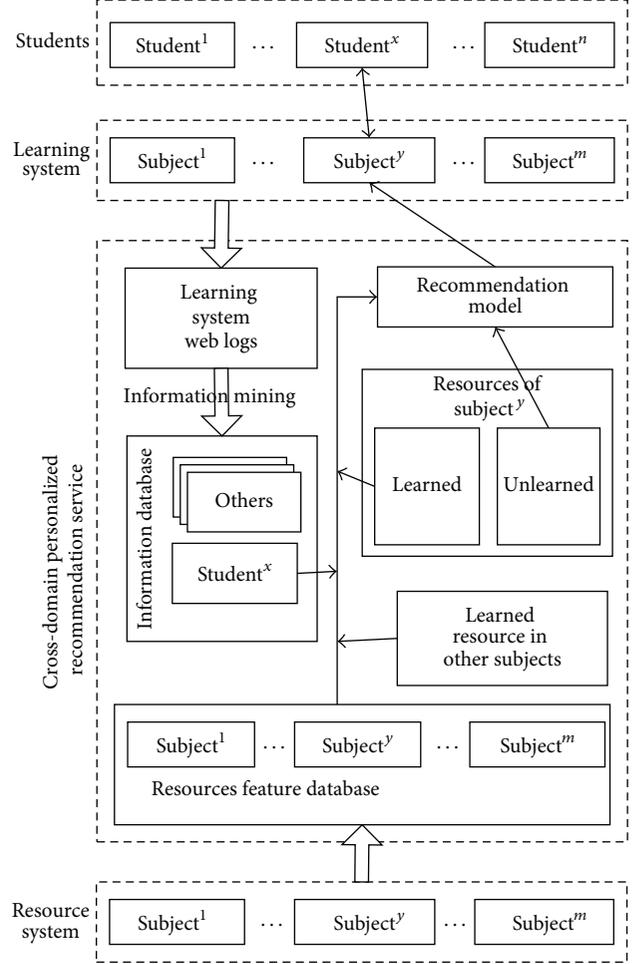


FIGURE 1: Service Model.

Then, the brief browsing interest measure P^b is computed as the following:

$$P^b = \beta \times \sum_{i=1}^m \frac{T_i^b}{B^b}, \quad (3)$$

where m is the browsing times, T^b is the each brief browsing time, B^b is the number of bytes of the resource brief, and β is the coefficient of resource brief.

Next, the browsing interest measure P is computed as the following:

$$P = P^c + P^b. \quad (4)$$

Finally, the mixed interest measure I is computed as the following:

$$I = \frac{P / \sum_{j=1}^N P_j}{1/N}, \quad (5)$$

where N is the total number of the browsed resources.

If $I \geq 1$, the learning resource is interested; otherwise, the learning resource is uninterested.

4. Cross-Domain Learning Resources Recommendation Algorithm

The basic theory of resources recommendation is analyzing new learning resources for students based on their personalized information and finding their interested resources. At present, traditional machine learning technology is used in most of resources recommendation service, and most of traditional machine learning is based on statistic learning, in which training data and testing data should be under the same distribution.

But in the cross-domain recommendation field, traditional machine learning cannot solve this problem, and the transfer learning is an effective method to solve it.

4.1. Transfer Learning. Now, transfer learning includes instance-based transfer learning and feature-based transfer learning [2–7]. Source domain data and target domain data may have different distribution, yet similar content, in which case, instance-based transfer learning has more obvious effect. The instance-based transfer learning refers to estimate instances from source domain data by some method and applies the instances evaluated better to the learning task of the target domain. Researchers have proposed many instance-based transfer learning algorithms [8–12]. Dai and others [13] have expanded traditional Boosting ensemble learning algorithm and proposed TrAdaBoost algorithm that has ability to transfer. The basic idea of TrAdaBoost is that the source domain data and the target domain data are mixed, and then they are trained. And after each iteration training, adapt the weight of the instance which is classified to be wrong. If the instance belongs to the target domain, the weight of it which is considered important should be increased and the effect of the sample should be improved at the next iteration training. If the instance belongs to the source domain, the weight of it which is considered unimportant should be decrease and the effect of sample should be reduced at next iteration training.

4.2. Cross-Domain Resources Recommendation Algorithm Design. In TrAdaBoost, the instances are treated in the same way during the initial weight assignment stage, and the weights of source domain instances and the weights of target domain instances are assigned averagely. For learning resources recommendation, the learning resources that the students browsed recently are more likely to reflect the current personalized information, so a time function can be added to weight assignment stage of the algorithm. In this way, the importance of learning resources can be distinguished.

In addition, the instance's weight which is classified wrong in target domain in TrAdaBoost algorithm will be increased continuously. On the contrary, the instance's weight which is classified wrong in source domain will be decreased continuously. When the number of instances which have too big or too small weight reaches a certain degree, it will decline the classifier ability of the algorithm. In order to solve this

problem, the weight of the wrong classified samples will be constrained.

The detailed description of the cross-domain personalized resource recommendation algorithm is shown below.

Step 1. Let $D = D_s \cup D_o$, where $D_o = \{(d_{m+1}, C(d_{m+1})), (d_{m+2}, C(d_{m+2})), \dots, (d_n, C(d_n))\}$ is the learned resources set in the current subject, $D_s = \{(d_1, C(d_1)), (d_2, C(d_2)), \dots, (d_m, C(d_m))\}$ is the learned resources set in other subjects, and $C(d_i) \in \{0, 1\}$ is the category label of d_i .

Step 2. Set R ($R < m$) and initialize $r = 0$; R is the times of iteration, and r is the count of iteration.

Step 3. Initialize the weight of instances:

$$w_i^0 = \begin{cases} \left(\frac{1}{m}\right) \times f(t_i), & 1 \leq i \leq m, \\ \left(\frac{1}{n} - m\right) \times f(t_i), & m + 1 \leq i \leq n. \end{cases} \quad (6)$$

The time function $f(t_i)$ is as the following:

$$f(t_i) = \begin{cases} 1, & t_i = 0 \\ 1 - \lambda \times \ln\left(1 + \frac{1}{t_i}\right)^{t_i}, & t_i > 0. \end{cases} \quad (7)$$

Step 4. $r = r + 1$.

Step 5. The weight of sample should be normalized, and it should be adjusted as the following:

$$w_i^r = \frac{w_i^{r-1}}{\sum_{i=1}^n w_i^{r-1}},$$

$$w_i^r = \begin{cases} u, & w_i^r > u \\ l, & w_i^r < l \\ w_i^r, & \text{others.} \end{cases} \quad (8)$$

Step 6. Call weak learner; then get back a hypothesis $h^r(d) : d- > C(d)$.

Step 7. Calculate the error of $h^r(d)$:

$$\varepsilon^r = \sum_{i=m+1}^n \frac{w_i^r \times |h^r(d_i) - c(d_i)|}{\sum_{i=m+1}^n w_i^r}. \quad (9)$$

Let

$$\varepsilon^r = \begin{cases} \varepsilon^r & \varepsilon^r < \frac{1}{2} \\ \frac{1}{2} & \varepsilon^r \geq \frac{1}{2}. \end{cases} \quad (10)$$

Step 8. Update the weights of the instances:

$$w_i^r = w_i^r \times \begin{cases} 1, & h^r(d_i) = c(d_i) \quad \text{or } w_i^r \in \{u, l\} \\ \beta_s, & h^r(d_i) \neq c(d_i), \quad 1 \leq i \leq m, w_i^r > l \\ (\beta_o^r)^{-1}, & h^r(d_i) \neq c(d_i), \quad m + 1 \leq i \leq n, w_i^r < u, \end{cases} \quad (11)$$

TABLE 1: Experimental data.

Student (data set)	Literature		History	
	Interested	Uninterested	Interested	Uninterested
$S_0 (U_0)$	812	614	789	543
$S_1 (U_1)$	785	586	856	654
$S_2 (U_2)$	914	689	758	542
$S_3 (U_3)$	1021	754	699	488
$S_4 (U_4)$	654	562	951	623
\vdots	\vdots	\vdots	\vdots	\vdots
$S_{59} (U_{59})$	948	643	751	523
$S_{60} (U_{60})$	987	763	928	616
$S_{61} (U_{61})$	1012	670	789	563
$S_{62} (U_{62})$	965	543	851	652

where

$$\beta_s = \frac{1}{1 + \sqrt{2 \times \ln(m/R)}}, \quad (12)$$

$$\beta_o^r = \frac{\varepsilon^r}{1 - \varepsilon^r}.$$

Step 9. If $r < R$, go to Step 4; otherwise, get back the final hypothesis:

$$h^R(d) = \begin{cases} 1, & \sum_{r=R/2}^R \ln\left(\frac{1}{\beta_o^r}\right) \times h^r(d) \geq \frac{1}{2} \times \sum_{r=R/2}^R \ln\left(\frac{1}{\beta_o^r}\right). \\ 0, & \text{others.} \end{cases} \quad (13)$$

In this algorithm, the formula (7) has the following properties: the function $f(t_i)$ on $[0, +\infty)$ is monotonically decreasing; $f(0) = 1$; $\lim_{t_i \rightarrow +\infty} f(t_i) = 0$; the decrease of function $f(t_i)$ is from fast to slow. Its properties satisfy the time characteristics of the learning resources. When $1 \leq i \leq m$, t_i is the time interval between the browsed time of the current instance and the recent browsed time of the instances in D_s . when $m + 1 \leq i \leq n$, t_i is the time interval between the browsed time of the current instance and the recent browsed time of the instances in D_o . $\lambda \in [0, 1]$ reflects the descending rate of the instance's importance. When $\lambda = 0$, the time characteristics of instances are not considered. When $0 < \lambda \leq 1$, the bigger the λ , the faster the descending rate of the instance's importance, and the value of $f(t_i)/f(0)$ is bigger than $1 - \lambda$.

In formula (11), u is the upper threshold of the weight, l is the limit threshold of the weight, $u = \max(w_i^1) \times \theta$, and $l = \min(w_i^1)/\theta$. The $\max(w_i^1)$ is the maximum value of w_i^1 , $\min(w_i^1)$ is the minimum value of w_i^1 , $\theta \geq 1$ is a real and can be adjusted. When θ is large enough, it means that the weights of the wrong classified instances are not constrained.

5. Experimental Evaluation

This paper chooses the Literature subject and the History subject to verify the effectiveness of the recommendation method and select the learning logs of a class with 62 students as the experimental data from a network learning system without recommendation service. The learning resources which are placed in "favorite" by students are used as the interested resources, and the learning resources which are placed in "recovery" by students are used as the uninterested resources. The experimental data sets are composed of the resources feature expression and their browsed time. Part of the information is shown in Table 1.

Firstly, test the effectiveness of the personalized information mining based on mixed interest measure; the specific experimental process of the personalized information mining for each student is shown as follows.

- (1) Extract the records of student S_i from web logs.
- (2) Extract the records which are corresponding to the resources in U_i .
- (3) According to the proposed method, calculate the student's interest measure of each learning resources in U_i .
- (4) Classify the learning resources according to whether the interest measure is bigger than 1, compare the results with the actual categorical attribute, and calculate the accuracy f of the method.

The experimental results are shown in Figure 2.

In 62 groups experimental data, 8 groups are between 84 and 85%, 25 groups are between 85 and 86%, 27 groups are between 86 and 87%, 2 groups are between 87 and 88%, and the average accuracy rate can reach 86.48%, and these meet the demands of personalized learning resources recommendation.

Then, apply the data set that was achieved from personalized information mining to test cross-domain resources recommendation algorithm, and the specific experimental process is as follows.

- (1) For each data set, randomly select 10% instances from History subject and all instances from Literature subject, use this data set as training set, and use the rest instances of History subject as test set.
- (2) The values of θ are 5 and 10^6 , the values of λ are 0.5 and 0.
- (3) Train and test, and the iteration times are 1 to 50. We should compute the accuracy, precision, and recall.
- (4) Repeat 10 times, calculate the average of accuracy, precision and recall.
- (5) Calculate the average of accuracy, precision, and recall of all data sets.

The final comparison results of the average of accuracy, precision, and recall are shown in Table 2.

Using the same data set, the experimental results of the traditional methods SVM and TrSVM [12] are as in Table 3.

TABLE 2: Final results comparison.

Coefficient		Accuracy	Precision	Recall	Comment
θ	λ				
10^6	0	80.98%	81.63%	81.17%	TrAdaBoost
10^6	0.5	82.11%	82.15%	82.06%	Add time function only
5	0	85.59%	86.10%	86.06%	Add weight constraints only
5	0.5	86.54%	86.63%	86.42%	In this paper

TABLE 3: Experimental results.

SVM			TrSVM			In this paper		
Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall
55.13%	55.89%	56.03%	82.3%	82.5%	82.4%	86.54%	86.63%	86.42%

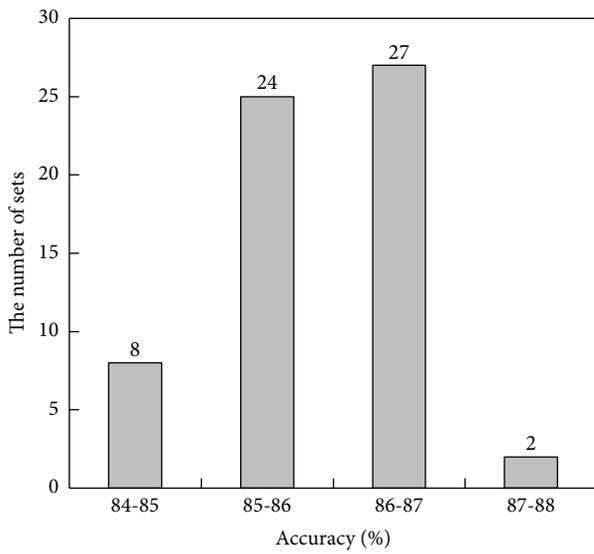


FIGURE 2: Personalized information mining experimental results.

Using the transfer learning recommendation method, the accuracy, precision and recall are all much higher than the traditional SVM method and higher than the TrSVM method.

It is shown in Table 2 that the accuracy, precision, and recall of TrAdaBoost algorithm are 80.98%, 81.63%, and 80.98%. After adding the time function to the weight assignments stage, the accuracy, precision and recall are 82.11%, 82.15% and 82.06%, just a slight improvement, this is because the algorithm uses Boosting integration learning framework, and so the initial weight assignment has a small effect on the result of final recommendation. When adding weight constraints only, the accuracy, precision, and recall are 85.59%, 86.10%, and 85.59%, better than TrAdaBoost algorithm; this is because there are error data in the data set obtained from personalized information mining. Without weight constraint, the error data will greatly affect the quality of recommendation service, and under the help of the weight constraint, the quality of recommend service is greatly improved. The accuracy, precision, and recall in this paper

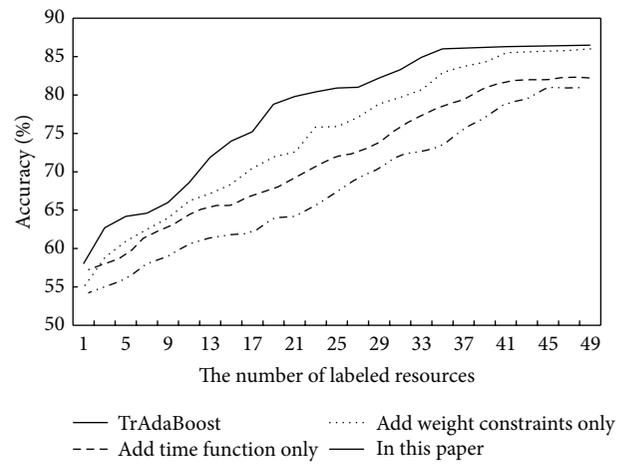


FIGURE 3: Convergence comparison.

are 86.54%, 86.63%, and 86.42%, apparently higher than the TrAdaBoost algorithm; the recommendation algorithm in this paper is more effective.

The convergence comparison is shown in Figure 3.

It is shown in Figure 3 that the TrAdaBoost algorithm needs about 45 times of iteration to achieve convergence. While adding time functions or weight constraints, the algorithm needs about 40 times to achieve convergence, and the proposed algorithm in this paper only needs about 35 times of iteration. The proposed algorithm is more effective.

The experimental results show that the recommendation service based on the proposed algorithm in this paper is superior to the recommendation service based on TrAdaBoost algorithm both in quality and efficiency and meets the demands of personalized learning resources recommendation service.

6. Conclusion

In the personalized learning resources recommendation service, the learning resources recommendation method requires a certain amount of personalized information to

achieve a satisfactory effect. How to achieve the information accurately and conveniently and make full use of the information to recommend is an urgent puzzle to be solved. We extract students' personalized information from web log based on the mix interest measure, put forward a kind of cross-domain learning resources recommendation method, and this method not only can take advantage of the students' personalized information in current subjects but also can use the personalized information in other subjects they have learned. The experimental results show that the proposed learning resources recommendation method can effectively improve the quality and efficiency of learning resources recommendation.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grants nos. 60973136 and 61073164) and the Youth Research Foundation of Liaoning University (Grant no. 2011LDQN14).

References

- [1] S. D. Pierrako and G. Paliouras, *Web Usage Mining as a Tool for Personalization: A Survey*, Kluwer Academic Publishers, 2003.
- [2] S. J. Pan, V. W. Zheng, Q. Yang, and D. H. Hu, "Transfer learning for wifi-based indoor localization," in *Proceedings of the Workshop on Transfer Learning for Complex Task of the 23rd AAAI Conference on Artificial Intelligence*, Chicago, Ill, USA, July 2008.
- [3] J. Ramon, K. Driessens, and T. Croonenborghs, "Transfer learning in reinforcement learning problems through partial policy recycling," in *Proceedings of the 18th European Conference on Machine Learning*, pp. 699–707, Springer, 2007.
- [4] M. E. Taylor and P. Stone, "Cross-domain transfer for reinforcement learning," in *Proceedings of the 24th International Conference on Machine Learning (ICML '07)*, pp. 879–886, ACM, June 2007.
- [5] R. Raina, A. Battle, H. Lee, B. Packer, and A. Y. Ng, "Self-taught learning: transfer learning from unlabeled data," in *Proceedings of the 24th International Conference on Machine Learning (ICML '07)*, pp. 759–766, Corvallis, Ore, USA, June 2007.
- [6] H. Daumé III and D. Marcu, "Domain adaptation for statistical classifiers," *Journal of Artificial Intelligence Research*, vol. 26, pp. 101–126, 2006.
- [7] Z. Wang, Y. Song, and C. Zhang, "Transferred dimensionality reduction," in *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases (ECML/PKDD '08)*, pp. 550–565, Springer, September 2008.
- [8] J. Huang, A. Smola, A. Gretton et al., "Correcting sample selection bias by unlabeled data," in *Proceedings of the 19th Annual Conference on Neural Information Processing Systems*, pp. 601–608, MIT Press, 2007.
- [9] M. Sugiyama, S. Nakajima, H. Kashima, P. Von Büna, and M. Kawanabe, "Direct importance estimation with model selection and its application to covariate shift adaptation," in *Proceedings of the 21st Annual Conference on Neural Information Processing Systems (NIPS '07)*, pp. 1433–1440, MIT Press, December 2007.
- [10] S. Bickel, C. Sawade, and T. Scheffer, "Transfer learning by distribution matching for targeted advertising," in *Proceedings of the 21th Annual Conference on Neural Information Processing Systems*, pp. 145–152, MIT Press, 2009.
- [11] A. Storkey and M. Sugiyama, "Mixture regression for covariate shift," in *Proceedings of the 19th Annual Conference on Neural Information Processing Systems*, pp. 1337–1344, MIT Press, 2007.
- [12] J. Hong, J. Yin, Y. Huang, Y. Liu, and J. Wang, "TrSVM: a transfer learning algorithm using domain similarity," *Journal of Computer Research and Development*, vol. 48, no. 10, pp. 1823–1830, 2011.
- [13] W. Dai, Q. Yang, G. Xue et al., "Boosting for transfer learning," in *Proceedings of the 24th International Conference on Machine Learning*, pp. 193–200, ACM, 2007.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

