SCIENCE CHINA
Information Sciences

• **Supplementary File** •

Pre-course student performance prediction with multi-instance multi-label learning

Yuling MA[1,2], Chaoran CUI[3], Xiushan NIE[3,1], Gongping YANG[1], Kashif Shaheed[1] & Yilong YIN[1*]

[1]*School of Computer Science and Technology, Shandong University, Jinan* 250100*, China;*
[2]*School of Information Engineering, Shandong Yingcai College, Jinan* 250104*, China;*
[3]*School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan* 250014*, China*

Appendix A Notations and algorithm

In this study, we use capital letters (e.g., X), bold lowercase letters (e.g., **x**), and non-bold lowercase letters (e.g., x) to denote sets, vectors, and scalars, respectively. We use $\mathcal{X}$ and $\mathcal{Y}$ to represent the input space (i.e., the set of all possible instances) and the output space (i.e., the class label set), respectively. If not otherwise clarified, all vectors are in column form. Table A1 shows notations used in this study, and Table A2 gives the proposed method.

Table A1 Notations

Variable	Description
p	the target semester for which courses will be predicted
t	the number of semesters
k_i	the number of courses offered in the i^{th} semester
$\mathbf{c}_{ij}$	the description vector of the j^{th} course in the i^{th} semester
S	the student dataset for training
P	an unseen student sample
r	the number of the neighboring examples in S
c	the number of the citers in S
R_P^r	the set of P's r nearest neighbors identified in S
C_P^c	the set of P's c citers identified in S
$\vec{\delta}_{X_i}$	$\lvert\mathcal{Y}\rvert$-dimension vector about the sample X_i, which can be calculated using Eq.(A1)
W	weight matrix to be learned
$Y_i(l)$	the ground truth of X_i on the l^{th} class

$\vec{\delta}_{X_i}$ in Table A1 can be obtained according to Eq.(A1):

$$\vec{\delta}_{X_i}(l) = \sum\nolimits_{Z\in(R^r_{X_i}\cup C^c_{X_i})} [\![\vec{L}_Z(l) == 1]\!], \quad l \in \mathcal{Y}. \tag{A1}$$

Here, for any predictor π, $[\![\pi]\!]$ takes the value of 1 if π holds and 0 otherwise. $\vec{L}_Z(l)$ is the label of the sample Z on the l^{th} class.

* Corresponding author (email: ylyin@sdu.edu.cn)

Table A2 The student performance prediction algorithm based on MIML-kNN

Inputs:
training set $S = \{(X_i, Y_i) \mid 1 \leqslant i \leqslant m\}$, r, c, and an unseen student sample P
Output:
L_P : label set for P
Begin
% training %
1. for each student sample $X_i \in S$
(1) Calculate $R^r_{X_i}$ and $C^c_{X_i}$ for X_i according to the similarity metric defined as Eq.(1) in the letter
(2) Calculate $\vec{\delta}_{X_i}$ according to Eq.(A1)
endfor
2. Obtain weight matrix W through minimizing sum-of-squares error function defined as Eq.(2) in the letter
% testing %
3. Calculate R^r_P and C^c_P for P according to the similarity metric defined as Eq.(1) in the letter
4. Calculate $\vec{\delta}_P$ for P according to Eq.(A1)
5. for each target course (i.e., label) $l \in \mathcal{Y}$ do
(1) Calculate $f(P, l)$ according to Eq.(3) in the letter
(2) Set the final output $L_P(l)$ for the course l to be 1 if $f(P, l) > 0$ holds, otherwise -1.
endfor
6. Output $L_P = \{l \mid L_P(l) = 1,\ l \in \mathcal{Y}\}$
End

Appendix B Datasets and Experimental results

After data preprocessing and integration, we obtain three MIML datasets, namely “BCS08_11”, “BET08_11”, and “GIM08_11”, as shown in Table B1. The datasets “BCS08_11”, “BET08_11”, and “GIM08_11” correspond to three specialized subjects “electronic technology”, “computer science and technology”, and “computer information management”, respectively. Table B2, Table B3, and Table B4 show the performance of each algorithm on the three datasets after balanced using “over sampling” technology. In order to understand the performance of each algorithm in detail, we also give their macro-$F_{1.5}$ values for each semester, which are shown as Figure B1, Figure B2 and Figure B3. Drawing from the experience in [1], we set the value of β to 1.5 in this study, in order to impose its preference over macro-R.

Table B1 Datasets for three different computer specialties

Collective Name	Size	Courses Number	Semesters Number	Positive/Negative samples number
BCS08_11	68	24	8	10 / 58
BET08_11	97	24	8	15 / 82
GIM08_11	228	15	5	34 / 194

Table B2 Results of each algorithm on BCS08_11

Methods	ave-acc ± std.	macro-P ± std.	macro-R ± std.	macro-$F_{1.5}$
NB	0.7138 ± 0.1012	0.5008 ± 0.1402	0.6611± 0.1251	0.6018
ANN	0.7169±0.0658	0.4453±0.1269	0.5654±0.1575	0.5221
LR	0.6845±0.0761	0.4424±0.1192	0.6185±0.1294	0.5510
SVM	0.7010±0.1251	0.5145±0.1283	0.7614±0.1023	0.6634
DT	0.7475±0.0657	0.4543±0.1906	0.3490±0.1501	0.3758
KNN	0.6833±0.1185	0.4744±0.1325	0.7463±0.1320	0.6344
Citation-KNN	0.7014±0.0920	**0.7144±0.0841**	0.7347±0.1137	0.7283
Our method	**0.7819±0.1151**	0.6470±0.1524	**0.9045±0.0886**	**0.8058**

References

1 Hu Y H, Lo C L, Shih S P. Developing early warning systems to predict students’ online learning performance. Computers in Human Behavior, 2014, 36:469-478

Table B3 Results of each algorithm on BET08_11

Methods	ave-acc ± std.	macro-P ± std.	macro-R ± std.	macro-$F_{1.5}$
NB	0.7409±0.0450	0.4582±0.0836	0.7210±0.1142	0.6129
ANN	0.7311±0.0668	0.4346±0.1473	0.5545±0.1121	0.5112
LR	0.7165±0.0764	0.4171±0.1346	0.6365±0.1589	0.5478
SVM	0.7113±0.0570	0.4483±0.0872	0.7815±0.1145	0.6361
DT	0.7398±0.0554	0.3890±0.1467	0.3941±0.1565	0.3926
KNN	0.7017±0.0822	0.4329±0.0964	0.7359±0.1029	0.6055
Citation-KNN	0.6890±0.1003	**0.6871±0.0848**	0.7118±0.1412	0.7040
Our method	**0.7978±0.0798**	0.6134±0.0948	**0.9667**±0.0550	**0.8212**

Table B4 Results of each algorithm on GIM08_11

Methods	ave-acc ± std.	macro-P ± std.	macro-R ± std.	macro-$F_{1.5}$
NB	0.7274±0.0436	0.3667±0.0568	0.6816±0.0634	0.5392
ANN	0.6989±0.0578	0.3104±0.0703	0.5495±0.0889	0.4443
LR	0.7157±0.0501	0.3494±0.0582	0.6482±0.0843	0.5131
SVM	0.7043±0.0702	0.3540±0.0650	0.7115±0.1060	0.5428
DT	0.7665±0.0317	0.3548±0.0747	0.3720±0.0823	0.3665
KNN	0.6579±0.0868	0.3084±0.0571	0.6898±0.0655	0.4997
Citation-KNN	0.6485±0.0570	0.6571±0.0607	0.6268±0.0825	0.6358
Our method	**0.8612±0.0464**	**0.6964±0.0690**	**0.9776±0.0358**	**0.8695**

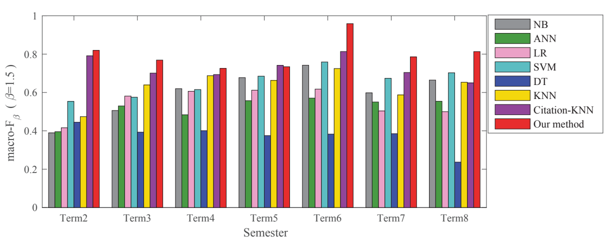

Figure B1 (color online) macro-F_β ($\beta = 1.5$) on the class-balanced data set BCS08_11

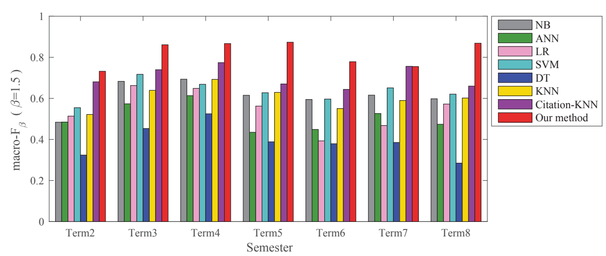

Figure B2 (color online) macro-F_β ($\beta = 1.5$) on the class-balanced data set BET08_11

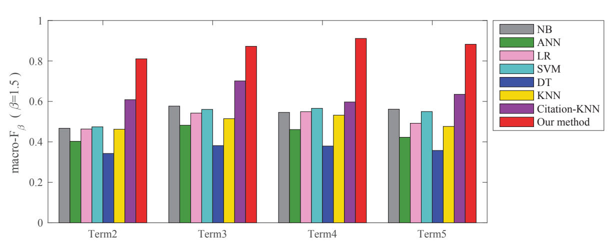

Figure B3 (color online) macro-F_β ($\beta = 1.5$) on the class-balanced data set GIM08_11