

ИНТЕГРИРАНЕ НА DATA MINING СРЕДСТВАТА В СРЕДИ ЗА ЕЛЕКТРОННО ОБУЧЕНИЕ

Даниела Орозова

Бургаски свободен университет
orozova@bfu.bg

Резюме: Разглеждат се възможностите за прилагане на средства за извличане на знания в среди за електронно обучение, с цел анализ и прогнозиране. Конструиран и описан е обобщено-мрежови модел на процеса на прилагане на Data Mining средства, който дава възможности за изследване на поведението на студентите, анализиране на данните и оценяване на ефективността от работата на обучаващата среда.

Ключови думи: e-learning, datamining tools, generalized net (GN)

1. Въведение

Съвременната образователна реформа е базирана на различни обучаващи среди, които нямат явна граница между материалния и виртуалния свят. Взимането на решения в процеса на обучение изисква изследване на поведението на студентите, анализиране на данни и оценяване на ефективността от педагогическите стратегии.

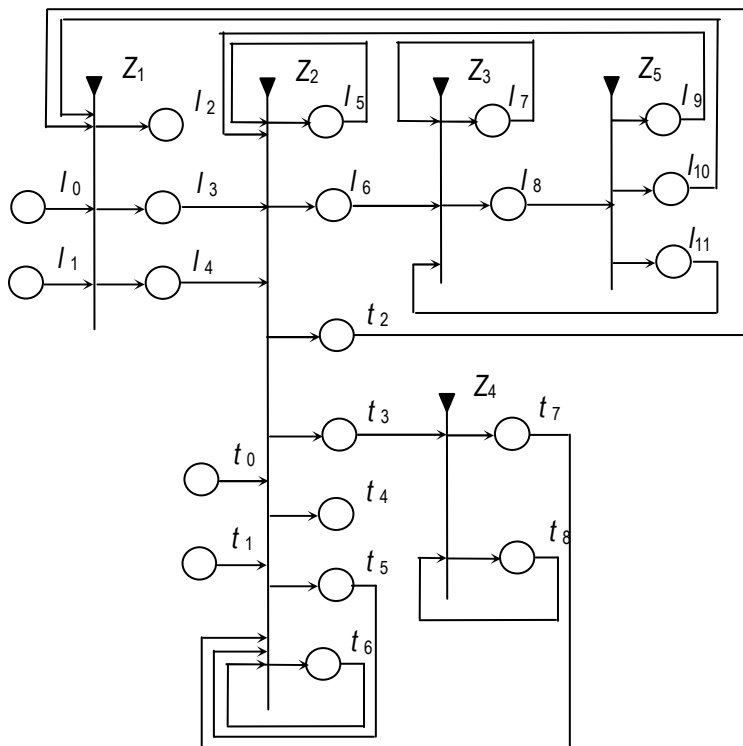
Data mining е процеса на откриване на скрити шаблони и връзки между данните [3,4]. В превод на български език често се използва термина извличане на знания от данните. Моделирането и изучаването на процеса на електронно обучение и прилагането на такива техники са необходими за персонализиране на курсовете за дистанционно обучение.

2. Модел на процеса на прилагане на Data Mining средства чрез апарата на обобщените мрежи

Създаденият модел в [7] чрез средствата на обобщените мрежи [1, 2], представя процеса на прилагане на средства за извличане на знания от данните в медицински бази от данни. Предложеният в този доклад модел е допълнен с възможности за прилагане на критерии за ограничаване на средствата и избор на подходящи техники за извличане на знания от данните в средите за електронно обучение.

Моделът, представен на фигура 1, съдържа 5 прехода и 21 позиции, групирани в две групи и свързани с два типа ядра, които постъпват в съответните типове позиции: α -ядра и I -позиции представят процеса на за

извличане на знания от данните, β -ядра и t -позиции представят критериите за ограничаване на средствата и избор на подходящи техники за извличане на знания от данните. За краткост ще се използва означението α - и β -ядра вместо α_i - и β_j -ядра, където i, j са номерата на съответните ядра.



Фиг.1. OM модел на процеса на прилагане на средства за извличане на знания от данните в среда за електронно обучение.

Първоначално едно β_0 -ядро стои в позиция t_6 с начална характеристика: „налични средства за извличане на знания от данните“.

На следващия преход от функционирането на мрежата β -ядрото се разделя на две. Оригиналното β -ядро ще продължи да стои в позиция t_6 , докато другото β -ядро ще се придвижи към прехода Z_5 , преминавайки през прехода Z_3 .

Ядрата α_0 и α_1 , постъпващи в мрежата през позиции I_0 и I_1 , получават характеристики съответно: „начални хипотези“;
„начални данни“.

Ядрата β_1 и β_2 постъпват в мрежата през позиции t_0 и t_1 . Тези ядра получават съответно начални характеристики:

„нова техника за извличане на знания от данните“;

„критерии за избор на техника за извличане на знания от данните“.

Следва описание на отделните преходи.

	l_2	l_3	l_4	
l_0	<i>false</i>	<i>false</i>	<i>true</i>	$\vee(\wedge(l_0, l_1), l_0, t_2)>$,
l_1	<i>false</i>	<i>false</i>	<i>true</i>	
l_{10}	$W_{10,2}$	$W_{10,3}$	<i>false</i>	
t_2	<i>false</i>	<i>true</i>	<i>false</i>	

където:

$W_{10,2}$ = “техниката за извличане на знания от данните е приложена”,

$W_{10,3}$ = $\neg W_{10,2}$.

Като първо действие на прехода Z_1 α_0 и α_1 -ядрата, които постъпват през позиция l_4 (от позиции l_0 и l_1) се сливат в едно ново α -ядро с характеристика:

“Начални хипотези, начални данни”.

При следващото активиране на прехода Z_1 β -ядрото, което постъпва в позиция l_3 (от позиция t_2) получава характеристиката:

“Цел, техника за извличане на знания от данните”.

При следващото активиране на същия преход β -ядрата могат да постъпят в позиция l_2 или l_3 като не получават нова характеристика.

$$Z_2 = \langle \{l_3, l_4, l_5, l_9, t_0, t_1, t_5, t_6, t_7\}, \{l_5, l_6, t_2, t_3, t_4, t_5, t_6\},$$

	l_5	l_6	t_2	t_3	t_4	t_5	t_6
l_3	<i>true</i>	<i>false</i>	<i>false</i>	<i>false</i>	<i>false</i>	<i>false</i>	<i>false</i>
l_4	<i>false</i>	<i>false</i>	<i>false</i>	<i>false</i>	<i>false</i>	<i>false</i>	<i>true</i>
l_5	$W_{5,5}$	$W_{5,6}$	<i>false</i>	<i>false</i>	<i>false</i>	<i>false</i>	<i>false</i>
l_9	<i>true</i>	<i>false</i>	<i>false</i>	<i>false</i>	<i>false</i>	<i>false</i>	<i>false</i>
t_0	<i>false</i>	<i>false</i>	<i>false</i>	<i>false</i>	<i>false</i>	<i>false</i>	<i>true</i>
t_5	<i>false</i>	<i>false</i>	<i>false</i>	<i>true</i>	<i>false</i>	<i>false</i>	<i>false</i>
t_6	<i>false</i>	<i>false</i>	<i>false</i>	<i>false</i>	$W_{6,4}$	$W_{6,5}$	<i>true</i>
t_7	<i>false</i>	<i>false</i>	<i>true</i>	<i>false</i>	<i>false</i>	<i>false</i>	<i>false</i>
t_1	<i>false</i>	<i>false</i>	<i>false</i>	<i>true</i>	<i>false</i>	<i>false</i>	<i>false</i>

$$\vee(l_3, l_4, l_5, l_9, t_0, \wedge(t_1, t_5), t_6, t_7)>$$

където:

$W_{5,5}$ = “има техника за извличане на знания от данните, която още не е приложена”,

$W_{5,6}$ = “техниката за извличане на знания от данните е вече е приложена”,

$W_{6,4}$ = “техниката за извличане на знания от данните е отхвърлена”,

$W_{6,5}$ = “техника за извличане на знания от данните е избрана”.

α -ядрата получават характеристики съответно:

“Избрани техники за извличане на знания от данните,
критерии за избор на data mining техники” в позиция l_3 ,

“Отхвърлена техника за извличане на знания от данните”, в позиция l_4 .

β_1 -ядрото, което постъпва в прехода Z_2 (от позиция t_0) ще се слее с оригиналното β_0 -ядро, което стои в позиция t_6 .

β_3 -ядрото, което постъпва в позиция t_5 получава характеристиката:

“Избрани техники за извличане на знания от данните”.

β -ядрата, които постъпват в позиции t_2 (от позиция t_7) и t_3 (от позиция t_5) не получават нова характеристика.

$$Z_3 = \langle \{l_6, l_7, l_{11}\}, \{l_7, l_8\}, \begin{array}{c|cc} & l_7 & l_8 \\ \hline l_6 & true & false \\ l_7 & W_{7,7} & W_{7,8} \\ l_{11} & true & false \end{array}, \sqrt{\langle l_6, l_7, l_{11} \rangle}, \rangle$$

където:

$W_{7,7}$ = “Има следващи стъпки от работата на текущата техника за извличане на знания от данните”,

$W_{7,8}$ = “Избрана е следващата стъпка от работата на текущата техника за извличане на знания от данните”

α -ядрото, което постъпва в позиция l_7 не получава нова характеристика, докато α -ядрото, което постъпва в позиция l_8 получава характеристиката:

“Текуща стъпка от прилагането на избраната техника за извличане на знания от данните”

$$Z_4 = \langle \{t_3, t_8\}, \{t_7, t_8\}, \begin{array}{c|cc} & t_7 & t_8 \\ \hline t_3 & false & true \\ t_8 & W_{8,7} & W_{8,8} \end{array}, \rangle,$$

където:

$W_{8,7}$ = “Избрана е техника за поставената цел”,

$W_{8,8}$ = $\neg W_{8,7}$.

β -ядрата, които постъпват в позиция t_8 не получават нова характеристика, докато β -ядрата, които постъпват в позиция t_7 получават характеристиката:

“Цел, избрани техники за извличане на знания от данните”.

$$Z_5 = \langle \{ l_8 \}, \{ l_9, l_{10}, l_{11} \}, \frac{l_9 \quad l_{10} \quad l_{11}}{l_8 \quad W_{8,9} \quad W_{8,10} \quad W_{8,11}} \rangle,$$

където:

$W_{8,9}$ = “Има следваща техника за извличане на знания от данните, която може да бъде приложена”,

$W_{8,10}$ = “Последната възможна техника за извличане на знания от данните е приложена”,

$W_{8,11}$ = “Има следваща стъпка от текущата техника за извличане на знания от данните, която ще се изпълнява”.

α -ядрата, които постъпват в позиции l_9 и l_{11} не получават нови характеристики, докато α -ядрата, постъпващи в позиция l_{10} получават характеристиката:

“Цел, средство за извличане на знания от данните, оценка от работата”.

3. Интегриране на средства за извличане на знания от данните със среди за електронно обучение

През последните години сме свидетели на широко използване на средствата за извличане на знания от данните, с цел анализ и прогнозиране в различни сфери на съвременния живот [5,6]. Интегрирайки тези техники в обучаващите среди могат да се търсят тенденции относно развитието на процесите на електронното обучение и неговото обслужване. Проследявайки процеса на избор и използване на различни средства могат да се изградят оптимални учебни среди с възможности за персонализирано овладяване на ключови знания, умения и компетенции.

Базирайки се на натрупани данни от работата на система за електронно обучение с различни потребители, прилагайки средства от областта на извличане на знания от данните, могат да се взимат различни решения. На базата на извършени изследвания [8,9] тук е направено систематизиране на основните задачи в тази област:

- Чрез прогнозиране на академичните дейности на студентите могат да се планират и подобряват учебните програми, както и процеса на обучение, изпитване, дипломиране и изразстване на обучаемите в обучаващи.

- Провеждане на анализи относно степента на придобиване и забравяне на знанията на обучаемите в различни интервали от време и сравняване на показателите през годините.

- Изследвайки работата с тестови системи при оценяване на студенти, могат да се оптимизират техниките за избор на тестови елементи и да се предприеме подходящ вид намеса в дейността на обучаемия.

- Изследвайки данните за потребителите да се идентифицират типове обучавани, на които да се предлага подходящо продължение на обучението. Използвайки техниката „класификация“ може да се идентифицира типа на студентите, които желаят да се запишат и завършат различни специалности и магистърски курсове. Когато се уточни кои студенти ще се привличат или участват в класирането може да се определи вероятността те да завършат съответната специалност в университета.

- Критичен период по време на обучението е между втората и третата година от следването. Ако студентите отпаднат или се прехвърлят в друга специалност и това може да бъде предвидено, то тяхното академично развитие може да бъде подпомогнато посредством допълнителни консултации, курсове за допълване на знанията или дори създаването на интересни интерактивни занимания с цел по-лесно продължаване на обучението. Тези техники могат да се използват за изясняване защо една група от студенти отпада от обучението, или кои са най-важните фактори за тяхната успеваемост по време на обучението.

- Използвайки класификационна функция може да се определи профила на добрите преподаватели. По-доброто разбиране на поведението им, позволява на организацията да разработи бизнес план за повишаване на тяхната удовлетвореност.

- Основна дейност на едно висше училище е настоящето търсене на студенти, които да се обучават в него. Подходящ модел на този процес може да отговори на въпроса коя група от потенциални студенти е най-желателна за обучение. За целта могат да се прилагат техниките “регресия” и “класификация”.

Заклучение

Интегрирането на системите за електронно обучение със средства за извличане на знания от данните е необходимо за процеса на персонализиране на курсове за дистанционно обучение. На базата на получените резултати могат да се въведат допълнителни мерки за анализ и промяна на обучаващите курсове и критериите за анализ. Това от своя страна е път към повишаване на качеството на обучението във висшето училище.

Литература

1. Atanassov Krassimir, On Generalized Nets Theory, “Prof. M. Drinov” Academic Publishing House, Sofia, 2007

2. Atanassov, Krassimir. Generalized Nets, World Scientific. Singapore, New Jersey, London, 1991
3. Chattamvelli, Data Mining Methods, Narosa Book Distributors, Pvt, Ltd, 2008.
4. Mark Hornick, Erik Marcade, Sunil Venkayala, Java data mining: strategy, standard, and practice, A practical Guide for Architecture, Design, and Implementation, 2006.
5. Pete Chapman, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer, Rüdiger Wirth, CRISP-DM 1.0, Step-by-step data mining guide, SPSS Inc. (USA) and OHRA Verzekeringen en Bank Groep B.V (The Netherlands), 2000.
6. Sumathi, S., S.N. Sivanandam, Introduction to Data Mining Principles and its Applications, Studies in Computational Intelligence, Springer, Vol. 29, 2006.
7. Orozova, D., E. Sotirova, P. Chountas, Generalized Net Model of the Knowledge Discovery in Medical Databases, International Journal BIO Automation, 2009, Volume 13/4, pp.281-288.
8. Orozova, D., K. Atanassov, Generalized Net Model of the Process of Selection and Usage of an Intelligent e-Learning System, book No 5, vol. 65, 2012, of the journal Bulgarian Academy of Sciences, pp. 591-598.
9. Сотирова, Е., Д. Орозова, Прилагане на data mining техники в електронното обучение, Международна конференция "Предизвикателства пред висшето образование и научните изследвания в условията на криза", 25-26 юни 2010, Бургас, стр. 201-205.

APPLYING DATA MINING TOOLS IN E-LEARNING

Daniela Orozova

Abstract: *Some applications of data mining tools in e-learning environments are considered. A Generalized Net (GN) model of the applying data mining tools is constructed. It provides facilities to explore student's behavior, analyzing resulting data and evaluation of the effectively of studies in the e-learning environment.*