

СЕРВІС АВТОМАТИЗОВАНОГО ЗАПИСУ НА ВИБІРКОВІ НАВЧАЛЬНІ ДИСЦИПЛІНИ В НАУКМА ТА МОЖЛИВІ НАПРЯМИ ЙОГО ІНТЕЛЕКТУАЛІЗАЦІЇ

Описано основні функціональні можливості автоматизованої системи запису студентів на вибіркові навчальні курси, що була розроблена в Національному університеті «Києво-Могилянська академія». Розглянуто можливості розвитку цієї системи в напрямі її інтеграції з іншими системами автоматизації бізнес-процесів університету, а також щодо набуття нею рис рекомендаційної системи.

Для надання рекомендацій щодо вибору дисциплін запропоновано три підходи: на основі розв'язання оптимізаційних задач, на основі аналізу мір близькості між анотаціями та на основі використання методик Data Mining і Web Usage Mining. Показано, що в найпростішому випадку автоматичне формування рекомендацій зводиться до розв'язання задачі про рюкзак із булевими змінними. Але врахування додаткових факторів призводить до нечітких і багатокритеріальних оптимізаційних задач, а також до необхідності використання методів машинного навчання.

Для вироблення рекомендацій на основі мір близькості запропоновано розвивати підхід на основі формалізації нечіткого поняття «середня відстань». Сформульовано основні напрями застосування методів Data Mining і Web Usage Mining.

Ключові слова: автоматизація, e-learning, рекомендаційна система, нечітка оптимізаційна задача, нечіткі міри близькості, Data Mining.

Вступ

Сьогодні ні в кого немає сумнівів, що система університетської освіти має зазнати кардинальних трансформацій, які значною мірою пов'язані з бурхливим розвитком інформаційно-комп'ютерних технологій та їхнім стрімким впровадженням у навчальний процес, а також у системи університетського менеджменту [10; 8 та ін.]. У [8] наведено певне узагальнення досвіду таких трансформацій в ІТ-освіті на прикладі Національного університету «Києво-Могилянська академія» (НаУКМА), але сказане значною мірою стосується й інших спеціальностей.

У цьому контексті видається винятково корисним посилення вибіркової компоненти навчального процесу, зокрема надання студентам можливості обирати, які дисципліни вони слухатимуть. Така практика є дуже поширеною в Європі, США та інших розвинених країнах, але не в Україні.

У НаУКМА запис студентів на вибіркові дисципліни було передбачено з самого початку становлення відродженої Академії, але до останнього часу він не був автоматизований і здійс-

нювався в ручному режимі. Тестування сервісу автоматизованого запису на вибіркові навчальні дисципліни було здійснено в червні–вересні 2016 р., і у березні 2017 р. планується провести автоматизований запис бакалаврів на вибіркові навчальні дисципліни з використанням цієї системи.

Крім суто технічних аспектів, подальший розвиток сервісу має передбачати розв'язання таких важливих взаємопов'язаних проблем:

- інтеграція сервісу з іншими системами, які використовують для автоматизації діяльності НаУКМА;
- інтелектуалізація сервісу, набуття ним рис рекомендаційної системи, яка могла б надавати рекомендації щодо доцільності вибору окремих навчальних дисциплін;
- акумуляція та аналіз даних про запис студентів на дисципліни;
- оцінка задоволеності студентів своїм вибором;
- стабілізація процесів запису та подальшої корекції.

Цю статтю присвячено обговоренню можливих шляхів до вирішення вказаних проблем.

Функціональність сервісу

Сервіс автоматизованого запису на вибіркові навчальні курси розроблявся у тісній взаємодії з зацікавленими підрозділами НаУКМА, і його організація добре узгоджується з загальними принципами університетського менеджменту в сфері організації навчального процесу. Зокрема, основні ролі користувачів системи відповідають рівням ієрархії керування, а саме:

- рівень студентів як споживачів навчальних послуг;
 - рівні надання навчальних послуг, а саме:
 - рівень кафедри;
 - рівень факультету;
 - рівень навчально-методичного відділу як центру керування навчальним процесом.
- Виокремлюються такі етапи бізнес-процесу:
- попередній етап, під час якого відбувається вибір дисциплін, які буде рекомендовано для запису;
 - етап запису, під час якого студенти мають можливість записуватися на вибіркові дисципліни;
 - етап корекції, під час якого на запис накладаються додаткові обмеження, визначається кількість груп, а студенти автоматично випишуються з дисциплін, які не відбулися, та повинні записатися на інші (дисципліна відбувається, якщо кількість записаних на неї студентів стає не меншою за деякий встановлений поріг);
 - етап бездіяльності, під час якого можна лише переглядати перелік дисциплін, результати запису тощо, ніякі бізнес-процеси на цьому етапі не відбуваються.

На попередньому етапі на рівні кафедри здійснюється вибір дисциплін, які рекомендуватимуть для запису. Приміром, нехай є q кафедр. Якщо позначити через $D^{(k)}$ множину дисциплін, яку читає k -та кафедра, то в результаті попереднього етапу має бути сформовано набір множин $R^{(k)}$, $k = 1, \dots, q$, де $R^{(k)} \subseteq D^{(k)}$.

На етапі запису для кожного j -го студента має бути сформовано набір дисциплін $V^{(j)}$, такий що

$$\forall v_{ij} \in V^{(j)} \exists R^{(k)} : v_{ij} \in R^{(k)}. \quad (1)$$

На етапі корекції набори множин $V^{(j)}$ та $R^{(k)}$ коригуються так, що обмеження (1) залишається в силі та з'являється додаткове обмеження: для будь-якого елемента будь-якої множини $R^{(k)}$ кількість студентів, які записалися на відповідну дисципліну, не має бути меншим за встановлений для цієї дисципліни поріг. Крім того, з кожної i -ю дисципліною пов'язана кількість заліко-

вих балів w_i , і ще одне обмеження полягає в тому, що для кожного студента сума залікових балів не має перевищувати встановленого порогу W . Аналогічні обмеження встановлюють на максимальне тижневе навантаження в кожному семестрі.

Проблема інтеграції

Навіть якби була можливість комплексної та всеохопної автоматизації всіх бізнес-процесів, НаУКМА не міг би реально наблизитися до такої автоматизації, особливо з огляду на умови, які склалися в країні в 2015–2016 рр. Тому домінує підхід, за якого автоматизуються окремі фрагменти університетської діяльності, і відповідні системи мають між собою певним чином обмінюватися даними та взаємодіяти.

На момент початку розробки сервісу автоматизованого запису на вибіркові навчальні дисципліни в Університеті функціонувала автоматизована система «Оптіма», в якій, зокрема, зберігаються дані про навчальні дисципліни та про студентів. Відповідно дані з «Оптіми» завантажуються до проміжної бази даних, а потім результати запису передаються до «Оптіми».

Перспективний напрям розвитку сервісу може бути пов'язаний із його інтеграцією з сервером дистанційної освіти distedu.ukma.kiev.ua, побудованим на базі MOODLE. Зокрема, на основі результатів автоматизованого запису можна було б записувати студентів на електронні навчальні курси в середовищі distedu, отримувати від distedu дані про проходження студентами електронних курсів та про їхні оцінки з окремих видів робіт тощо.

Проблема вироблення рекомендацій

Ще один перспективний напрям у розвитку сервісу автоматизованого запису на вибіркові навчальні дисципліни полягає в його інтелектуалізації. Зокрема, йдеться про набуття рис рекомендаційної системи, яка надавала б рекомендації кафедрам щодо формування переліку навчальних дисциплін, які рекомендуються до запису, а студентам – щодо вибору дисциплін, на які їм варто записуватися.

У рамках наведеної вище формалізації система має надавати кожній k -й кафедрі рекомендації щодо формування множини $R^{(k)}$, а кожному j -му студентові – щодо вибору множини $V^{(j)}$, яка задовольняє базовому обмеженню (1), а також обмеженням на максимальну кількість залікових балів і максимальне тижневе навантаження.

Сформульовану проблему можна розв'язати на основі різних підходів, зокрема:

- побудови деякої оптимізаційної задачі;
- аналізу мір близькості;
- застосування методик Data Mining та Web Usage Mining.

Розглянемо ці підходи більш детально.

Оптимізаційний підхід до вироблення рекомендацій

Введемо таке позначення: $x_{ij} = 1$, якщо j -й студент обирає i -ту дисципліну; 0 – якщо не обирає.

У найпростішому випадку можна вважати, що вибори студентів не залежать один від одного. Тоді достатньо розглядати змінні $x_j = 1$, якщо студент обирає i -ту дисципліну; 0 – якщо не обирає.

Як і раніше, w_i – «вага» i -ї дисципліни, кількість кредитів за неї. Для простоти пояснення в рамках цієї статті ми розглядаємо лише обмеження, пов'язані з кількістю залікових балів; обмеження на тижневе навантаження можна описувати аналогічно.

Крім того, нехай c_i – міра корисності i -ї дисципліни.

Тоді рекомендації щодо вибору оптимальної множини дисциплін можуть бути отримані на основі розв'язання такої оптимізаційної задачі:

$$\begin{aligned} \sum_i c_i x_i &\rightarrow \max, \\ \sum_i w_i x_i &\leq W, \\ x_i &\in \{0,1\} \end{aligned} \quad (2)$$

По суті, це задача про рюкзак із булевими змінними. Якщо потрібно, щоб сумарна вага не перевищувала сумарного порогу, а точно дорівнювала йому, виникає інша задача:

$$\begin{aligned} \sum_i c_i x_i &\rightarrow \max, \\ \sum_i w_i x_i &= W, \\ x_i &\in \{0,1\} \end{aligned} \quad (3)$$

В Університеті прийнято вимогу, відповідно до якої сума набраних кредитів W має дорівнювати деякій величині приблизно, тобто задана неточно, нечітко. Аналогічно, коефіцієнти c_i теж можуть бути задані неточно. Відповідно до цього W та c_i природно можуть бути представлені як нечіткі величини. Тоді (3) перетворюється на нечітку оптимізаційну задачу. Слід зазначити, що на сучасному етапі методи розв'язання нечітких оптимізаційних задач досліджені недостатньо та інтенсивно розвиваються; деякі підходи до цього розглядалися в [6; 4].

Множина рекомендованих дисциплін $R = \bigcup_{k=1}^q R^{(k)}$ є неоднорідною. Студент під час навчання має засвоїти певний набір понять, які можна згрупувати за тематичними кластерами; позначимо кількість цих кластерів через M . Кожна дисципліна робить свій внесок у засвоєння понять з тих чи тих кластерів.

Позначимо через $c_i^{(l)}$ міру корисності i -ї дисципліни для l -го кластера. Тоді набір дисциплін, які система може рекомендувати студентові, може бути отриманий на основі розв'язання такої багатокритеріальної задачі:

$$\begin{aligned} \sum_i c_i^{(l)} x_i &\rightarrow \max, l = 1, \dots, L, \\ \sum_i w_i x_i &\leq M, \\ x_i &\in \{0,1\} \end{aligned} \quad (4)$$

Один із найуживаніших підходів до розв'язання задачі (4) полягає в згортці критеріїв, тобто в переході до однокритеріальної задачі.

$$\begin{aligned} \sum_{l=1}^L \alpha_l \sum_i c_i^{(l)} x_i &\rightarrow \max, \\ \sum_i w_i x_i &\leq M, \\ \sum_{l=1}^L \alpha_l &= 1, \\ x_i &\in \{0,1\} \\ \alpha_l &\in [0,1] \end{aligned} \quad (5)$$

Тут α_l – задані коефіцієнти, які змістовно означають міру важливості l -го кластера.

І коефіцієнти $c_i^{(l)}$, і коефіцієнти α_l можуть бути задалегідь невідомими. Тоді вони мають визначатися під час навчання на основі накопичення досвіду, і перспективним тут видається застосування методів навчання з підкріпленням, зокрема марковських процесів прийняття рішень [9].

Аналіз мір близькості

Відомим підходом до формування рекомендацій щодо вибору об'єктів є використання мір близькості. Зокрема, як міру схожості дисциплін можна розглядати міру близькості між анотаціями в рамках моделі «термін–документ» [5; 7; 1].

У [8] розвивається підхід, відповідно до якого найбільш рекомендовані об'єкти не повинні бути ні надто близькими до уже вибраних, ні надто віддаленими. Справді, найбільш близькі об'єкти будуть надто схожими на вже обрані, і існує значний ризик того, що відповідна інформація вже буде знайома студентові. Іншими словами, функція залежності міри перспективності від міри близькості має спочатку зростати, дося-

гати максимуму в районі середніх значень мір близькості, і потім знову спадати. Для формалізації таких мір залежності в [8] залучається апарат теорії нечітких множин на основі використання нечітких правил типу «якщо відстань СЕРЕДНЯ, релевантність ВИСОКА».

Можливі застосування Data Mining та Web Usage Mining для аналізу ходу запису та його результатів

Накопичення та аналіз інформації про результати запису має дати змогу говорити про виявлення певних закономірностей на основі застосування методів Data Mining та Web Usage Mining, які нині інтенсивно розвиваються [2; 11]. Зокрема, можна розглядати такі практично корисні задачі (перелік не є вичерпним):

- виявлення того, студенти яких спеціальностей найчастіше обирають дисципліни з інших спеціальностей;
- виявлення, які види поведінки студентів найчастіше спричиняють проблемні ситуації, які вимагають ручного втручання на етапі корекції;
- вироблення рекомендацій щодо вибору дисциплін за подібністю на основі того, які дисципліни найчастіше обирали схожі студенти зі схожими профілями поведінки;
- виявлення закономірностей, пов'язаних із ситуаціями, коли студент спочатку обирає дисципліну, а потім відмовляється від неї;
- виявлення закономірностей, пов'язаних із ситуаціями, коли студент починає не зі свого факультету, а з іншого;
- аналіз ситуацій, коли дисципліни не відбуваються, з метою зменшення ризику таких ситуацій.

Корекція правил вироблення рекомендацій може здійснюватися на основі таких факторів:

- наскільки часто студент слідує за рекомендаціями, наданими системою, або відмовляється від них;
- наскільки студент задоволений своїм вибором.

Проблема аналізу задоволеності студентів

І для навчання з підкріпленням, і для виявлення закономірностей на основі Data Mining важливим є аналіз того, наскільки студент задоволений рекомендаціями системи та своїм вибором, оскільки міру задоволеності можна безпосередньо пов'язати з винагородами за здійснені

дії в рамках марковських моделей прийняття рішень. Розвиток сервісу автоматизованого запису на вибіркові дисципліни передбачає два основні шляхи оцінювання винагород:

- на основі безпосереднього опитування;
- на основі аналізу успішності студентів з обраних дисциплін.

Проблема стабілізації запису та корекції

У процесі автоматизації запису студентів на вибіркові дисципліни постала проблема, пов'язана з формуванням груп на етапі корекції. В рамках наявної процедури визначення кількості сформованих груп і груп резерву операції запису та виписування можуть або спричинити потенційно нескінченний цикл зміни кількості груп і, відповідно, потенційно нескінченний період корекції, або ж призвести до того, що студентів не вдасться записати навіть за наявності такої можливості. З огляду на це довелося передбачити серйозне обмеження: заборонити студентам виписуватися зі сформованих груп. Зрозуміло, що це є небажаним. Тому для розвитку сервісу важливими є розробка та імплементація робастних алгоритмів формування груп, які б не спричиняли таких проблем.

Висновки

Можливості сервісу для автоматизованого запису студентів на вибіркові дисципліни, створеного в Національному університеті «Києво-Могилянська академія», можуть бути суттєво розширені, якщо надати йому рис рекомендаційної системи. Така система могла б давати рекомендації кафедрам під час вибору дисциплін, які пропонуються для запису, а також студентам під час запису.

Визначено три підходи до формування рекомендацій:

1. Розв'язання певних оптимізаційних задач. Як базову розглядають задачу про рюкзак із булевими змінними. Оскільки цільова функція та обмеження можуть бути визначені не до кінця і задані неточно, нечітко, розглядають нечітку задачу про рюкзак. Врахування розподілу дисциплін за темами призводить до багатокритеріальної оптимізаційної задачі. Крім того, коефіцієнти цих задач можуть бути заздалегідь невідомі, і тоді вони мають визначитися шляхом машинного навчання.

2. Аналіз мір близькості між дисциплінами, зокрема текстових мір подібності анотацій.

Пропонується розвивати підхід на основі формалізації нечіткого поняття «середня відстань».

3. Використання методів Data Mining та Web Usage Mining; наводяться приклади практич-

них задач, для яких такий аналіз міг би бути корисним.

Розвиток цих напрямів є предметом подальших досліджень.

Список літератури

1. Анісімов А. В. Пошук інформації / А. В. Анісімов, А. М. Глибовець, М. М. Глибовець, А. С. Шабінський. – К. : НАУКМА, 2015. – 284 с.
2. Барсегян А. А. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP / А. А. Барсегян, М. С. Куприянов, В. В. Степаненко, И. И. Холод. – СПб. : БХВ-Петербург, 2007. – 384 с.
3. Бублик В. В. Моделі трансформації інформаційної освіти в контексті руху до інформаційного суспільства: досвід факультету інформатики НАУКМА / В. В. Бублик, М. М. Глибовець, О. В. Олецкий // Наукові праці. Науково-методичний журнал. Т. 71. Вип. 58. Педагогічні науки. – Миколаїв : Вид-во МДГУ ім. П. Могили, 2007. – С. 60–64.
4. Ивохин Е. В. О подходах к решению транспортной задачи с нечеткими ресурсами / Е. В. Ивохин, А. Субхи Камл Баррак // Проблемы управления и информатики. – 2014. – № 5. – С. 47–62.
5. Ландэ Д. В. Поиск знаний в Интернет / Д. В. Ландэ. – М. : Изд. дом «Вильямс», 2005. – 272 с.
6. Лю Б. Теория и практика неопределенного программирования / Б. Лю. – М. : БИНОМ. Лаборатория знаний, 2005. – 416 с.
7. Маннинг К. Д. Введение в информационный поиск / К. Д. Маннинг, П. Рагхаван, Х. М. Шютце. – ООО «И. Д. Вильямс», 2011. – 528 с.
8. Олецкий О. В. Про підхід до автоматичного формування рекомендацій для відвідувачів веб-порталу на основі теорії нечітких множин / О. В. Олецкий // Наукові записки НАУКМА. – Т. 177 : Комп'ютерні науки. – К., 2015. – С. 37–40.
9. Рассел С. Искусственный интеллект: современный подход / С. Рассел, П. Норвиг. – М. : Изд. дом «Вильямс», 2006. – 1408 с.
10. Сергієнко І. В. Програмні засоби створення і супроводу розподіленого навчального середовища / І. В. Сергієнко, М. М. Глибовець, С. С. Гороховський, А. М. Глибовець. – К. : НАУКМА; Аграр Медіа Груп, 2012. – 710 с.
11. Bing L. Web Data Mining: Exploring Hyperlinks, Contents and Usage Data. Springer / L. Bing. – 2011. – 642 p.

V. Gorborkov, O. Oletsky

A SERVICE OF AUTOMATED ENROLLMENT TO OPTIONAL DISCIPLINES AND POSSIBLE WAYS OF MAKING IT MORE INTELLIGENT

The paper describes the main features of the system of automated enrollment to optional disciplines developed in the National University of "Kyiv-Mohyla Academy" and its place in the overall environment for automating learning processes in the University. The authors regard the main trends of integrating this system to other services and of transforming it to the recommendation system.

To form recommendations, three approaches are proposed: on the basis of solving the optimization task, on the basis of analyzing similarities between annotations, and on the basis of applying methods of Data Mining and Web Usage Mining.

Within the optimization approach, the problem, in the simplest case, can be formulated as a knapsack problem with boolean variables. But in more complex cases, some fuzzy and multi-criteria optimization problems are formulated. The multi-criteria problem arises if we regard different topics and the domain of knowledge addressed by courses. As the goal function and constraints of these tasks can be unknown, the use of machine learning methods might be needed.

It should be taken into account that the most appropriate recommendations elaborated by the recommendation system on the basis of evaluating similarity measures should neither be very similar to the already completed courses nor very distant from them. So an approach involving fuzzy concept of average distance is developed.

The main directions of applying Data Mining and Web Usage Mining such as detecting problem situations, analyzing cases of choosing disciplines from other faculties, reducing risks of failures in choosing disciplines, etc., are distinguished.

Keywords: automation, e-learning, recommendation system, fuzzy optimization task, fuzzy similarity measures, Data Mining.

Матеріал надійшов 15.10.2016