

СТВОРЕННЯ РЕКОМЕНДАЦІЙНОЇ СИСТЕМИ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ ЗАПИСУ НА ВИБІРКОВІ НАВЧАЛЬНІ ДИСЦИПЛІНИ

У статті розвинуто підходи до побудови рекомендаційної системи підтримки прийняття рішень на базі автоматизованої системи запису студентів на вибіркові навчальні курси, що була розроблена в Національному університеті «Києво-Могилянська академія». Сформульовано відповідні оптимізаційні задачі, при цьому основний акцент робиться на пошуку методів знаходження адекватних цільових функцій. Розглянуто можливість застосування евристичних методів, зокрема на базі PageRank-подібних методик та методу Сааті. Крім того, розглянуто задачу автоматизованого розрахунку кількості груп в умовах нечітко сформульованих м'яких обмежень.

Ключові слова: автоматизація, e-learning, рекомендаційна система, оптимізаційна задача, оптимальний вибір множин.

Вступ

У 2016–2017 роках у Національному університеті «Києво-Могилянська академія» (НаУКМА) була створена автоматизована система, яка забезпечує для студентів можливість запису на вибіркові навчальні дисципліни, а для кафедр і факультетів пропонує набір дисциплін, з яких здійснюватиметься вибір [2]. При цьому береться до уваги те, що, крім вибіркових, студенти обов'язково повинні прослухати певну кількість нормативних дисциплін, існують певні обмеження щодо кількості набраних кредитів, максимального тижневого навантаження, кількості груп та розподілу студентів по групах з урахуванням розкладу тощо.

Видається доцільним здійснити інтелектуалізацію розробленої системи та надати їй рис рекомендаційної системи підтримки прийняття рішень. Дослідження такого роду проводяться в низці університетів Європи та світу; як приклад можна навести рекомендаційну систему, розроблену в Університеті ім. Масарика (Брно, Чехія) [7].

Підхід, який розвивається в [7], базується передовсім на певних евристичних прийомах (роївий інтелект, прийняття рішень за схожістю і т. п.). Ми визнаємо корисність та перспективність такого підходу, але його можна посилити за рахунок розв'язання оптимізаційних задач, залучення апарату марковських процесів, теорії нечітких множин тощо.

Підхід, який залучає до розв'язання цієї задачі оптимізаційні постановки, в основних рисах розглядався в [2; 8]. При цьому основна складність

пов'язана навіть не стільки з розв'язанням власне оптимізаційних задач, скільки з належним підбором цільових функцій. Деякі міркування щодо можливого застосування окремих евристичних методів, зокрема навчання з підкріпленням, наведено в [2–5; 8; 9].

Цю статтю присвячено розвитку та конкретизації цього підходу в поєднанні із застосуванням евристичних методик. Основний акцент при цьому робиться на методиках пошуку цільових функцій та їхніх параметрів.

Вибір дисциплін як оптимізаційна задача

Математично задачу вибору оптимального набору дисциплін можна сформулювати таким чином.

Нехай задано трійку $(X, F(X), u)$, де X – база множин елементів, $F(X)$ – деяка система підмножин множини X , $f: X \rightarrow F(X)$ – деяка функція від множин. Тоді задача полягає в такому: знайти множину $A \in D \subseteq F(X)$ таку, що

$$u(A) \rightarrow \max,$$

D – деяка множина допустимих множин.

Змістовно для нашої задачі елементи множини X відповідають дисциплінам, що викладаються. Точніше, $X = \bigcup_{i=1}^q X^{(i)}$; $X^{(i)}$ – набір дисциплін, що читається i -ю кафедрою, $\forall i, j, i \neq j X^{(i)} \cap X^{(j)} = \emptyset$. Елементи $F(X)$ відповідають наборам дисциплін. Функція u інтерпретується як функція корисності набору дисциплін; ми вважаємо, що $f(\emptyset) = 0$.

Обмеження, які визначають множину D , залежать від специфіки задачі; для вибору оптимальних наборів дисциплін ці обмеження можуть полягати в тому, що сумарна кількість набраних залікових балів, а також тижневе навантаження не мають перевищувати встановлених величин.

Функція корисності набору отримується шляхом комбінування окремих одноелементних множин:

$$u(\{x_1, \dots, x_n\}) = \text{Comb}(u(\{x_i\})),$$

конкретний вигляд функції Comb залежить від специфіки задачі.

Якщо функція u є лінійною, зокрема, якщо її можна подати у вигляді

$$u(\{x_1, \dots, x_r\}) = \sum_{i=1}^r c_i u(\{x_i\}),$$

ми отримуємо задачу про рюкзак. Але ця функція не обов'язково має бути лінійною, оскільки корисність вибору дисципліни може залежати від вибору інших дисциплін.

Крім того, функція u може бути задалегідь невідомою або відомою неточно.

Дворівнева оптимізаційна задача

Слід зазначити, що в задачі вибору оптимального набору дисциплін виокремлюється як мінімум два рівні. Спочатку методисти кафедр формують множини дисциплін, які рекомендуються для вибору, а потім студенти вибирають дисципліни з рекомендованих множин.

Для побудови формалізованого опису такої дворівневої задачі позначимо описану вище оптимізаційну задачу через

$$Z = Z((X, F(X), u), D),$$

а її розв'язок (знайдену оптимальну або, можливо, субоптимальну множину елементів) через

$$A = r(Z).$$

Тоді задача вибору оптимальної множини дисциплін для кожного конкретного студента запишеться у вигляді набору

$$Z((A = \bigcup_i A^{(i)}, F(A), u), D),$$

де

$$A^{(i)} = r(Z((X^{(i)}, F^{(i)}(X^{(i)}), u^{(i)}), D^{(i)})).$$

Слід звернути увагу на таке:

– обмеження для кафедр та для студентів можуть бути різними; зокрема, якщо обмеження

для студентів пов'язані з кількістю набраних кредитів та тижневим навантаженням, то для кафедр подібних обмежень узагалі може не бути;

- цільові функції в загальному випадку також можуть бути різними;
- можна вводити індивідуальні цільові функції для різних студентів.

Деякі задачі багатокритеріальної оптимізації при виборі навчальних дисциплін

Замість одного критерію оптимізації u можна розглядати цілу низку критеріїв; тоді задача стає багатокритеріальною. Формально таку задачу можна записати як

$$Z((X, F(X), \{L_k, k = 1, 2, \dots\}), D),$$

де L_k – окремі критерії, які потрібно максимізувати.

Ми розглядаємо, зокрема, такі системи критеріїв:

- максимізувати власне корисність обраної множини дисциплін, але мінімізувати сумарну складність засвоєння цих дисциплін;
- при залученні до розгляду онтології предметної галузі слід враховувати, що кожна дисципліна передбачає вивчення окремих тем, пов'язаних із вузлами онтології, і тоді кожний критерій L_k полягає в оптимізації знань, отриманих із відповідної теми.

Згортка критеріїв дозволяє перейти до однокритеріальної задачі

$$u(A) = \sum_k \lambda_k L_k(A) \rightarrow \max$$

при тих самих обмеженнях, і тоді коефіцієнти λ_k відображатимуть індивідуальні міри важливості кожної теми.

Автоматичне визначення кількості груп та розподіл на групи

Задача визначення кількості груп залежно від кількості студентів, які записалися на конкретні дисципліни, виявилася досить нетривіальною. В Академії діють норми, відповідно до яких встановлюється мінімальна кількість студентів у групі m та максимальна кількість M , методисти кафедр при роботі в системі автоматизованого запису можуть коригувати ці параметри в бік збільшення.

Оптимальна кількість студентів у групі не регламентується. Зрозуміло, що її зменшення призводить до збільшення кількості груп і відповідно – до зайвих фінансових витрат, а збільшення – до зниження якості викладання.

Крім того, бажано уникати ситуацій, коли виникає надлишок записаних студентів, які не

можуть бути розподілені по групах із дотриманням встановлених вимог. Такі ситуації є дуже небажаними, але гарантовано їх уникнути можна, лише якщо встановити M не меншим за $2m-1$. Зрозуміло, що це є неприйнятним.

Тому було прийнято рішення після завершення основного запису ще раз перерахувати ліміти, але цей процес здійснювався вручну і викликав низку незручностей. Тому актуальною є задача автоматизованого розрахунку кількості груп; ми пропонуємо підхід до її розв'язання, що базується на мінімізації небажаних ситуацій. Підхід до оптимізації можна описати як такий, що ґрунтується на фазифікації обмежень. У загальних рисах цей процес можна охарактеризувати таким чином: обмеження задаються як нечіткі величини; розв'язується оптимізаційна задача при різних рівнях задоволення обмежень; результати порівнюються. Після завершення процесу проводиться дефазифікація, внаслідок якої отримуються чіткі рекомендації щодо лімітів та оптимальної кількості груп. Деякі міркування щодо цього наведено в [9].

Можуть розв'язуватися також задачі, пов'язані з автоматичним розподілом студентів на групи так, щоб мінімізувалася кількість збігів у розкладі, а також з автоматичним складанням самого розкладу. Для цього видається перспективним застосування генетичних алгоритмів, які добре зарекомендували себе при розв'язанні подібних задач [1].

Створення бази знань про взаємозв'язки між дисциплінами

Для побудови рекомендацій щодо вибору навчальних дисциплін видається доцільним мати базу знань про взаємозв'язки між окремими дисциплінами. Видається доцільним побудувати таку базу знань на основі продукційних правил «якщо-то». Фрагмент такої бази знань у нотації, подібній до прологівської, може мати вигляд:

```

r e c o m m e n d ( X , S , y e s ) :
- k n o w l e d g e ( S , p r o g r a m m i n g , s o l i d ) ,
a d v a n c e d ( X , y e s ) .
r e c o m m e n d ( X , S , y e s ) :
- k n o w l e d g e ( S , p r o g r a m m i n g , w e a k ) ,
a d v a n c e d ( X , n o ) .
a d v a n c e d ( p a s c a l , n o ) .
a d v a n c e d ( j a v a , y e s ) .

```

Для низки дисциплін, які викладаються в Академії, встановлено вимоги щодо необхідності попереднього прослуховування інших

дисциплін; ці вимоги можуть бути безпосередньо включені до бази знань. Але можна піти далі і залежно від складності дисципліни та рівня знань студента давати рекомендації з різною мірою сили та категоричності.

Застосування агентно-орієнтованого підходу

Видається перспективним розвиток системи в напрямку агентних технологій. Ідея полягає в тому, що для кожного студента (або типу студентів) можна створити інтелектуального програмного агента, який мав би у своєму розпорядженні набір правил щодо того, як діяти в тих чи інших ситуаціях. Тоді такий агент при зміні ситуації (наприклад, після завершення формування списків дисциплін, які пропонуються до запису) міг би здійснити попередній вибір дисциплін, результати якого можуть бути скориговані студентом.

Евристичні методи

Характерною особливістю проблеми є те, що цільові функції, в тому числі корисності окремих дисциплін, а також функції комбінування корисностей множин дисциплін на основі корисностей окремих дисциплін можуть бути апріорно невідомими. Тому видається доцільним застосування різного роду евристик, спрямованих на апріорну оцінку окремих варіантів, а також їх порівняльний аналіз. Мова може йти про застосування методу Сааті та PageRank-подібних методик. Зокрема, можна розглядати граф станів, вузли якого в найпростішому випадку відповідають навчальним дисциплінам, та методу отримання перехідних імовірностей з урахуванням того, які дисципліни мають бути вивчені попередньо.

При цьому розглядаємо такі типи зв'язків:

- одинична залежність: для вивчення дисципліни A студент має попередньо вивчити дисципліну B ;
- І-зв'язки: для вивчення A треба вивчити $i B$, $i C$;
- АБО-зв'язки: для вивчення A треба вивчити або B , або C .

Матриця перехідних імовірностей будується на основі тих чи інших породжуючих моделей. Зокрема, як базова може розглядатися така породжуюча модель:

1. Агент починає навігацію з випадково вибраного вузла серед тих, які ще не розглядалися.
2. З імовірністю α агент переходить з вузла X до одного з вузлів, від яких X залежить, а з

імовірністю $(1-\alpha)$ – до випадковим чином вибраного іншого вузла.

3. При аналізі залежностей розрахункових переходних імовірностей визначається типом залежності:

- за наявності одиничної залежності X від Y агент переходить до вузла Y з імовірністю α ;

- за наявності I-залежності X від Y та Z агент переходить до кожного з цих вузлів з імовірністю $\alpha/2$. Якщо вузол, скажімо, Y , уже розглядався, перехід до Y відбувається з імовірністю 0 , а до Z – з імовірністю α ;

- за наявності АБО-залежності X від Y та Z агент переходить до кожного з цих вузлів з імовірністю $\alpha/2$. Якщо хоча б один із вузлів Y, Z уже розглядався, перехід до будь-якого з них відбувається з імовірністю 0 .

Очевидним розвитком цієї методики може бути введення кількісних оцінок, а також аналіз імовірностей повернення на залежні вузли.

Ми також розвиваємо підхід, який базується на побудові моделі задачі у вигляді деякої графової моделі, для якої характеристики вузлів певним чином залежать від характеристик суміжних вузлів. Зокрема, можна вводити певні логічні рівні, і опис елементів (цільових функцій, обмежень, інших характеристик) на одному рівні утворюється на основі опису елементів попереднього рівня.

Для рекомендаційної системи, що розглядається, ми розглядаємо такі рівні:

- рівень онтології – теми та окремі питання, які мають бути вивчені;

- рівень альтернатив, кожна альтернатива розглядається як рекомендована підмножина базового набору дисциплін;

- рівень окремих дисциплін.

Зв'язки між рівнями можливо описати таким чином.

Міри важливості вибору окремих дисциплін для вивчення окремих тем як вузлів онтології задамо у вигляді матриці $M=(m_{ij})$, m_{ij} – важливість i -ї дисципліни для j -ї теми.

Нехай будь-яка k -та альтернатива полягає у виборі дисциплін, що утворюють множину L_k . Логічно прийняти, що для кожної окремої теми корисність набору не перевищує максимальної корисності окремої дисципліни, і тоді комбінована корисність k -ї альтернативи для досягнення j -го критерію запишеться як

$$A_{kj} = \text{Comb}(m_{ij}) = \max_{i \in L_k} m_{ij}. \quad (1)$$

Нехай c_{kj} – комбінована корисність, ρ_j – індивідуальна корисність j -го критерію, r_k – корисність

k -ї альтернативи. Тоді можна розглядати співвідношення

$$r_k = \sum_{j \in L_k} c_{kj} \rho_j. \quad (2)$$

Тепер, коли побудована система співвідношень (1)-(2), ми можемо використовувати наявні знання про окремі коефіцієнти цих співвідношень для отримання інших коефіцієнтів на основі відповідних прямих та обернених задач. Зокрема, для співвідношення (2) коефіцієнти r_k (корисності множин елементів) можуть бути оцінені на основі побудови матриці парних порівнянь із подальшим застосуванням методу Сааті [6].

Висновки

У статті проаналізовано можливості створення рекомендаційної системи підтримки прийняття рішень для автоматизації запису на вибіркові навчальні дисципліни, яка б рекомендувала факультетам та кафедрам дисципліни, які варто пропонувати до запису, а студентам – дисципліни, на які варто записуватися.

У своїй основі задача вибору оптимального набору дисциплін може бути математично описана як задача вибору оптимальної підмножини деякої базової множини елементів. З урахуванням специфіки предметної галузі формулюється дворівнева оптимізаційна задача, а також можливі багатокритеріальні задачі. Одна з таких постановок залучає до розгляду онтологію тем, пов'язаних із дисциплінами. При цьому основна проблема полягає в необхідності уточнення самих цільових функцій, які в загальному випадку можуть бути неадитивними відносно корисностей базових варіантів та відповідно нелінійними. З урахуванням цього формулюються певні пропозиції щодо формування функцій корисності множин варіантів на основі корисностей окремих варіантів; при цьому залучається до розгляду онтологія тем, пов'язаних із дисциплінами.

Обговорюється можливість врахування зв'язків між дисциплінами на основі побудови певної продукційної системи правил. Для утвореного таким чином I-АБО-графа пропонується методика знаходження кількісних оцінок на основі PageRank-подібного аналізу переходів. Інша евристика, яка пропонується в статті, пов'язана з побудовою деякого графа залежностей та аналізу співвідношень між його вузлами. При цьому характеристики одних вузлів можуть бути оцінені на основі парних порівнянь і методу Сааті, а інших, пов'язаних із ними – на основі системи рівнянь (2).

Розглянуто також задачу автоматизованого розрахунку кількості груп за нечітких м'яких обмежень та задачу автоматизованого формування груп.

Список літератури

1. Глибовець М. М. Еволюційні алгоритми / М. М. Глибовець, Н. М. Гулаєва. – Київ : НАУКМА, 2013. – 828 с.
2. Горборуков В. В. Сервіс автоматизованого запису на вибіркові навчальні дисципліни в НАУКМА та можливі напрями його інтелектуалізації / В. В. Горборуков, О. В. Олецький // Наукові записки НАУКМА. – 2016. – Т. 190 : Комп'ютерні науки. – С. 52–56.
3. Олецький О. В. Про оптимізацію структури веб-порталу на основі марковських процесів прийняття рішень / О. В. Олецький // Вісник Київського національного університету імені Тараса Шевченка. Серія Фізико-математичні науки. – 2013. – Спецвипуск. – С. 134–137.
4. Олецький О. В. Про підхід до автоматичного формування рекомендацій для відвідувачів веб-порталу на основі теорії нечітких множин / О. В. Олецький // Наукові записки НАУКМА. – 2015. – Т. 177 : Комп'ютерні науки. – С. 37–40.
5. Олецький О. В. Про деякі задачі прийняття рішень при побудові рекомендаційної системи вибору навчальних матеріалів / О. В. Олецький // Теоретичні та прикладні аспекти побудови програмних систем : матеріали міжнародної конференції ТАAPSD'2016. Київ, 5–9 грудня 2016 р. – Київ : [б. в.], 2016. – С. 188–192.
6. Черноруцкий И. Г. Методы принятия решений / И. Г. Черноруцкий. – Санкт-Петербург : БХВ-Петербург, 2005. – 616 с.
7. Bydzovska H. Course enrollment recommender system / H. Bydzovska // Proceedings of the 9th International Conference of Educational Data Mining. – 2016. – P. 312–317.
8. Oletsky O. A non-linear optimization task for forming optimal sets of selective disciplines / O. Oletsky // Обчислювальний інтелект (результати, проблеми, перспективи) : матеріали IV Міжнародної науково-практичної конференції, 16–18 травня 2017 р., Київ, Україна. – Київ, 2017. – С. 180–181.
9. Oletsky O. On establishing fuzzy constraints for the problem of optimizing amount of groups for selective educational courses / O. Oletsky // Problems of decision making under uncertainties (PDMU-2017). May 10–13, 2017. Mukachevo, Ukraine. – 2017. – P. 94.

V. Gorborkov, O. Oletsky

THE MAIN APPROACHES TO DEVELOPING A RECOMMENDATION DECISION MAKING SUPPORT SYSTEM FOR ENROLLMENT IN SELECTIVE DISCIPLINES

The paper regards approaches to developing a recommendation decision making supporting system for enrollment in selective educational disciplines. Reciprocal optimization problems including multi-level and multi-criteria problems of choosing sets of disciplines are formulated. Hereby it is taken into account that optimization criteria and goal functions might be defined not well and therefore should be estimated, e.g. by taking into account the available experience. These functions may be non-additive and non-linear, so combining functions of getting usefulness of sets by usefulness of single disciplines should be involved. Some approaches to getting such combining functions are proposed in the paper. Possible ways of taking into account possible connections between disciplines (e.g. if one discipline should rely upon another one and therefore should precede it) which form the framework of a curricula template are also regarded. A possible fragment of such a graph of relations in a Prolog-like notation is given. The ontology of topics covered by disciplines is considered. A possibility of applying heuristic approaches, such as PageRank-like techniques and the Saati method based on pair comparisons, has been explored. As a result, we regard a possible generative model for getting transitional probabilities and some equations for getting unknown coefficients. The problem of automated calculating of amount of groups and forming groups of students under soft and fuzzy constraints is regarded as well.

Keywords: automation, e-learning, recommendation system, optimization task, optimal selection of sets.

Матеріал надійшов 12.10.2017