

УДК 004.896

## ГЕНЕРАЦІЯ МНОЖИНИ ЯДЕР ПРОДУКЦІЙНИХ ПРАВИЛ В СИСТЕМАХ УПРАВЛІННЯ ОЦІНЮВАННЯМ ЗНАТЬ

Турчак А.М., Єгорова О.В., Златкін А.А.

Черкаський державний технологічний університет

У статті проаналізована задача побудови системи управління оцінюванням знань. Розглядається спосіб добування знань з повнотекстових джерел. Запропоновано підхід до розв'язання проблеми автоматичної побудови продукційних правил з використанням модифікованого багатофакторного композиційного методу спрямованої оптимізації. Наведено структурні особливості та конструктивні елементи методу для деревоподібної форми представлення потенційних розв'язків. Модифікований метод призначено для формування бібліотеки методів побудови онтологій.

**Ключові слова:** онтологія, продукційні правила, еволюційні алгоритми.

**Постановка проблеми.** В останні роки спостерігається зростання кількості обсягів підготовки спеціалістів, проте їх компетентність, здатність адекватно відповідати на виклики сучасності та приймати правильні рішення піддається сумніву. Підготовка спеціалістів у вищих навчальних закладах, а також підвищення їх кваліфікації, на різних курсах закінчується процесом оцінювання знань. Аналітичний огляд методів контролю рівня професійної підготовки фахівців різного профілю свідчить про переважання традиційних методів контролю знань, зокрема, у формі заліків, іспитів або тестування. Разом з тим, таке оцінювання знань є неповним та має низький рівень об'єктивності, тому контроль знань доцільно здійснювати з використанням сучасних інформаційних технологій.

В більшості випадків в таких системах реалізовано жорсткий шаблон організації подачі навчального матеріалу або тестування. Як наслідок, системи, що розглядаються характеризуються відсутністю засобі перевірки розгорнутих природномовних робіт, творчих здібностей та логічного мислення. Наведений вище аспект є причиною розробки автоматизованих систем навчання та контролю знань орієнтованих на аналіз текстової інформації, представленої на природній мові; пошук, обробку та добування знань; інтеграцію знань із декількох предметних областей для забезпечення ефективності проведення досліджень міждисциплінарного характеру.

Створенню ефективної автоматизованої системи навчання та контролю знань повинна передувати розробка онтології предметної області. В свою чергу, розробка онтології предметної області можлива лише за умови існування бази даних, що містить всю необхідну інформацію про об'єкт або процес. Користувач такої системи визначає набір та структуру правил, а експерт – відповідну онтологію, створюючи тим самим фіксовану ієрархічну систему виведення. Для того, щоб мати можливість масштабувати системи автоматизованого навчання і контролю знань необхідно також створити онтологію засобів подання вхідної та вихідної інформації, а також програмних елементів.

Взаємодія експерта, аналітика та оператора відбувається в процесі спілкування. При цьому виникає ряд лінгвістичних проблем і задач, до яких належить: побудова словника предметної області, побудова номенклатури предметної області та поєднання онтологій.

**Аналіз останніх досліджень і публікацій.** Методи побудови онтологій умовно поділяють на дві групи. До першої групи належать традиційні методи природномовної обробки тексту, до другої – методи, орієнтовані безпосередньо на побудову онтологій.

До традиційних методів аналізу, що використовуються для побудови онтологій належать методи морфологічного та синтаксичного аналізу, а також метод виділення словосполучень. Детально суть перелічених методів викладено в роботах [1-3]. Безпосередню побудову онтологій в [8] пропонують здійснювати шляхом ситуаційного моделювання.

**Виділення не вирішених раніше частин загальної проблеми.** Онтологія є ієрархічною структурою скінченної множини понять, що описує задану предметну область, а отже, містить машинно-орієнтовані визначення основних понять предметної області і відношень між ними. Наведені властивості онтології забезпечують можливість повторного використання знань іншими людьми, базами даних та програмними додатками. При цьому значно підвищується ефективність використання як інтелектуальних систем, так і традиційних інформаційних систем. Побудова онтологій в діалоговому режимі з використанням існуючих інструментальних засобів вимагає значених витрат часу та ресурсів. Один із шляхів вирішення проблеми полягає у автоматичній будові онтологій на основі заданих повнотекстових джерел.

Автоматичне добування знань з науково-технічних текстів передбачає не тільки виявлення термінів, але і добування знань про них. Для цього необхідно розпізнати в тексті семантичні відношення між термінами, оскільки вони задають семантичну структуру термінології.

Для задач, що належать до слабкоструктурованих предметних областей, майже відсутні об'єктивні моделі їх подання. В більшості випадків, такі задачі розв'язуються експертами. Автоматична побудова правил розв'язання задачі на основі конструктивних знань експерта дозволяє сформулювати бібліотеку декларативних довговічних та масштабованих методів. Як правило, експертам складно формалізувати правила, які вони використовують для розв'язання задач через евристичний характер знань.

**Мета статті.** Метою даної роботи є розробка модифікованого багатофакторного методу композиційного подолання невизначеності для генера-

ції ядер продукційних правил призначених для природномовної обробки технічних текстів.

**Виклад основного матеріалу.** Вважатимемо, що знання експертів набувають вигляду продукційних правил. Продукцією називається правило виду «Якщо А, то В». Продукції можуть описувати імплікації (з істинності А випливає істинність В), задавати умовні операції (якщо справедлива умова А, то виконати дію В), бути праведлими підстановки (змінити А на В).

Представлення знань у вигляді сукупності продукцій називається продукційною моделлю. Під антицидентом А і консеквентом В розуміють деяку множину фактів. Кожний факт має структуру простої ядерної конструкції мови ситуаційного моделювання, яку формально подають як  $xRy$  де  $x, y$  – терміни,  $R$  – семантичні відношення.

Для розв'язання задачі генерації ядер продукційних правил модифікуємо метод композиційного подолання невизначеності наведений в [4]. В основу даної модифікації покладемо наступні положення:

а. Ядро продукційного правила подамо у вигляді потенційного розв'язку.

б. Потенційним розв'язком вважатимемо сукупність індивідуальних програм певного формату. Потенційний розв'язок має форму деревоподібної структури.

с. Програма потенційного розв'язку складається з динамічного набору елементів.

д. Початкова сукупність потенційних розв'язків генерується випадковим чином, причому елементи програми вибирають з області допустимих значень відповідного шаблону.

е. Кожна нова популяція створюється за допомогою оператора мутації з урахуванням міри впевненості в тому, що розв'язок представник є близьким до предметної області та важливості шаблонів і зв'язків між ними.

ф. Потенційний розв'язок не повинен містити повторюваних програм.

г. Розв'язком задачі генерації ядер продукційних правил є сукупність потенційних розв'язків, що задовольняють наведені вище обмеження.

h. Функція пристосованості обчислюється для індивідуальної програми, потенційного розв'язку та сукупності потенційних розв'язків.

В [5] структуру потенційного розв'язку подають таким шаблоном судження:

$$PP = \langle \text{ЯКЩО}, (PX, PR, PY), \langle \text{логічна зв'язка} \rangle, \dots, \langle \text{логічна зв'язка} \rangle, (PX, PR, PY) \rangle, \\ TO, (PX, PR, PY), \langle \text{логічна зв'язка} \rangle, \dots, \langle \text{логічна зв'язка} \rangle, (PX, PR, PY) \rangle,$$

де ЯКЩО, ТО – службові слова;  $PX, PR, PY$  – шаблони компонентів  $x, R, y$  простої ядерної конструкції; І, АБО – логічні зв'язки.

В наведеній простій ядерній конструкції шаблони  $PX, PY$  складаються з двох частин: терм методу та позначення терму. Шаблон  $PY$ , на відміну від шаблону  $PX$ , може містити текстові, символні або числові константи. Для кожного шаблону виділяють два типи терм методів – базові та власні. Крім того, базовий або власний терм може містити терм-супутники терміна. Шаблон  $PR$  складається з дієслова та терм-супутника семантичного відношення.

Для того, щоб на шаблонах компонентів можна було будувати метрику, розширимо цю модель потенційного розв'язку за рахунок введення двох

скалярних величин – ваги важливості компонентів шаблону ядерної конструкції та зв'язків [6, 7]. Коefіцієнт важливості компонентів шаблону ядерної конструкції (зв'язку) – це чисельна міра, яка характеризує значущість певного терм-методу (зв'язку) у конкретному продукційному правилі і динамічно змінюється за певними правилами у процесі генерації ядер продукційного правила.

Початковою інформацією для генерації множини ядер продукції з використанням багатofакторного композиційного методу спрямованої оптимізації буде: множина базових термів  $T_b$ ; множина коефіцієнтів важливості базових термів  $W_b$ ; множина власних термів методу  $T_i$ ; множина коефіцієнтів важливості власних термів  $W_i$ ; множина терм-супутників терму  $T_{S_{xy}}$ ; множина коефіцієнтів важливості терм-супутників терму  $W_{S_{xy}}$ ; множина типів констант  $C_T$ ; множина допустимих значень змінних  $V$ ; множина допустимих операцій виразу  $O$ ; множина дієслів  $D$  на яких базуються семантичні відношення  $R$ ; множина терм-супутників  $T_{SR}$  семантичного відношення  $R$ ; множина коефіцієнтів важливості терм-супутників  $W_{SR}$  семантичного відношення  $R$ ; множина семантичних відношень  $R = \{(v, t) | v \in V, t \in T_{SR}\}$ ; множина коефіцієнтів важливості  $W_R = \{w_v | v \in V, t \in T_{SR}\}$  семантичних відношень  $R$ ; множина терм-ознак  $T_C$ .

Як зазначалося раніше, програма потенційного розв'язку складається з динамічної множини елементів. Подамо ці елементи у десятковому представлення таким чином: елемент 1 – ключове слово або логічна зв'язка; елемент 2 – коefіцієнт важливості ключового слова або логічної зв'язки; елемент 3 – тип терму (складова  $x$ ); елемент 4 – базовий або власний терм, або терм-супутник  $t \in T_{S_{xy}}$ , або терм-ознака (складова  $x$ ); елемент 5 – коefіцієнт важливості базового або власного терма, або терм-супутника (складова  $x$ ); елемент 6 – тип значення терма (складова  $x$ ); елемент 7 – позначення терма або вираз (складова  $x$ ); елемент 8 – дієслово (складова  $R$ ); елемент 9 – коefіцієнт важливості дієслова (складова  $R$ ); елемент 10 – терм-супутник (складова  $R$ ); елемент 11 – коefіцієнт важливості терм-супутника (складова  $R$ ); елемент 12 – тип терма (складова  $y$ ); елемент 13 – базовий або власний терм, або терм-супутник  $t \in T_{S_{xy}}$ , або терм-ознака (складова  $y$ ); елемент 14 – коefіцієнт важливості базового або власного терма, або терм-супутника  $t \in T_{S_{xy}}$ , або терм-ознаки (складова  $y$ ); елемент 15 – тип значення терма (складова  $y$ ); елемент 16 – позначення терма або вираз для терма, або константа (складова  $y$ ).

Модифікований з метою розв'язання задачі генерації ядер продукційних правил багатofакторний композиційний метод спрямованої оптимізації міститиме такі кроки:

Крок 0. Встановити лічильник ітерацій  $t = 0$ .

Крок 1. Визначити початкову кількість потенційних розв'язків  $\zeta = 1, \rho$  та створити рівномірно розподілену на  $\Omega$  представницьку популяцію потенційних розв'язків  $x_1^t, x_2^t, \dots, x_\rho^t$ , де  $\Omega$  – деякий компакт.

Крок 3. Визначити критерій  $E$  завершення пошуку розв'язку.

Крок 4. Обчислити придатність індивідуальної програми, потенційного розв'язку та популяції потенційних розв'язків.

Придатність індивідуальної програми  $f_i^t$  обчислюємо за виразом  $f_i^t = \sum_{j=1}^C (M - |C_{(i,j)} - T_j|)_b$  де  $M$  –

коefficient мутації,  $C_{(i,j)}$  – значення, повернене індивідуальною програмою  $i$  для значення придатності  $j$  (поза випадками придатності  $C_q$ ),  $T_j$  – цілкове значення для значення придатності  $C_q$ .

Крок 5. Доки критерій  $E$  не виконується:

Крок 5.1. Пронормувати значення  $f'_c$  так, щоб  $f^{nh} \in [0;1]$ ,  $\sum f^{nh} = 1$ .

Крок 5.2. Сформувати матрицю попарних порівнянь Саати  $S$  таким чином. Серед нормованих значень функції знаходимо мінімальне  $f^{nh}$ , розбиваємо відрізок  $[0;1]$  на 10 інтервалів:  $[0;0,1)$ ,  $[0,1;0,2)$ , ...,  $[0,9;1]$ . Тоді для всіх  $h \in \{1,2,\dots,\rho\}$ , якщо  $f^{nh} \in [0,1k;0,1+0,1k)$  і  $f^{nh} \in [0,1l;0,1+0,1l)$ , де  $k,l \in \{0,1,\dots,9\}$ , то  $s_{ch} = l-k+1$ . Інші елементи матриці  $S$  розраховуються так:  $S_{rq} = S_{cq}/S_{cr}$ .

Крок 5.3. Розрахувати власні числа матриці  $S$  і для максимального власного числа  $a_{max}$  знаходимо відповідний власний вектор  $w$ . Значення  $w_c$  вказують на міру впевненості в тому, що потенційні розв'язки  $x_1^t, x_2^t, \dots, x_p^t$  належить до предметної області.

Крок 5.4. Генерація нащадків потенційних розв'язків і формування нової популяції потенційних розв'язків. В методі композиційного подолання невизначеності [4] нащадків пропонують одержувати з використанням оператора мутації. Для деревоподібних структур автори генетичного програмування пропонують використовувати вузловий, відтинаючий або зростаючий тип оператора мутації.

Ми вважаємо, що для ефективного пошуку розв'язку при реалізації оператора мутації необхідно враховувати міру впевненості в тому, що потенційний розв'язок належить до предметної області. Це дозволить більш детально дослідити область пошуку. При цьому виникає декілька гіпотез:

- ймовірність застосування оператора мутації до базової вершини або відношення між вершинами зменшується із збільшенням значення ваги базової вершини та відношення між вершинами;
- чим меншим є значення  $w_c$ , тим більшими повинні бути сегменти дерева, що піддаються мутації, що дозволить розширити область пошуку в околі кращого розв'язку, а в області іс-

нування потенційного розв'язку, що найменше відповідає предметній області, окіл пошуку буде максимально звуженим;

– більше значення  $w_c$  є причиною детального дослідження окремих елементів деревовидних структур та зв'язків між ними, що дозволить детально дослідити окіл існування кращого розв'язку.

Для визначення ваг базових ознак, які задовольняють описані вище припущення, в [8] пропонується один із двох способів:

а. арифметичний, який полягає в обчисленні відношення різниці  $(k+1)$  рівня дерева та рівня, на якому знаходиться терм, до суми всіх рівнів гілки:

$$\omega_i = \frac{k+1-i}{\sum_{j=1}^k j};$$

б. геометричний, який полягає в обчисленні суми геометричної прогресії:

$$\omega_i = \frac{2^{k-i}}{2^k - 1}.$$

Ваги важливості зв'язків наводяться в [9].

Крок 5.5. На попередньому кроці виконано генерацію  $s$ -ї потенційних розв'язків. Знаходимо відповідні значення придатності потенційних розв'язків  $f'_c$ . За цими значеннями, а також за значеннями  $f'_c$  визначаємо  $s$  кращих розв'язків  $x_1^{t+1}, x_2^{t+1}, \dots, x_p^{t+1}$ , збільшуємо лічильник ітерацій  $t = t+1$  і переходимо на крок 4.

**Висновки і пропозиції.** Вдосконалено шаблон судження ядра продукційного правила. Побудова такої структури здійснена за адаптацією термів та зв'язків. З цією метою у шаблон судження, який задає потенційний розв'язок, додано дві скалярні величини (важливості термів та зв'язків), які використовуються для обчислення відстаней. Розглянуто способи задання початкових коефіцієнтів важливості понять та відношень. Модифіковано багатофакторний композиційний метод спрямованої оптимізації для розв'язання задачі генерації ядер продукційних правил. Проведені дослідження свідчать на користь модифікованого методу. Його верифікація на відомих тестових наборах свідчить про ефективність запропонованих ідей та технік.

## Список літератури:

1. Найханова Л.В. Автоматическое реферирование научного текста на основе использования онтологического тезауруса [Текст] / Л.В. Найханова, С.В. Машанова // Теоретические и прикладные вопросы современных информационных технологий: мат-лы Всерос. науч.-техн. конф. – Улан-Удэ: Изд-во ВСГТУ, 2008. – С. 130-134.
2. Снитюк В.Е. Интеллектуальное управление оценением знаний [Текст]: монография / В.Е. Снитюк, К.Н. Юрченко. – Черкасы, 2013. – 262 с.
3. Комп'ютерні онтології та їх використання у навчальному процесі. Теорія і практика [Текст]: монографія / С.О. Довгий, В.Ю. Велічко, Л.С. Глоба, О.Є. Стрижак., Т.І. Андрущенко, С.А. Гальченко, А.В. Гончар, К.Д. Гуляєв, В.М. Кудряк, К.В. Ляшук, О.В. Палагін, М.Г. Петренко, М.А. Попова, В.І. Сидоренко, О.О. Слюсаренко, Д.В. Стус, М.Ю. Терновой. – К.: Інститут обдарованої дитини, 2013. – 310 с.
4. Снитюк В.Е. Композиционное преодоление неопределенности в задачах нелинейной многофакторной оптимизации [Текст] / В.Е. Снитюк // Донецк: Искусственный интеллект. – 2004. – № 4. – С. 207-210.
5. Найханова Л.В. Генерация множества ядер продукционных правил в задаче автоматического построения библиотеки декларативных методом [Текст] / Л.В. Найханова // Информационные технологии. – 2008. – № 10. – С. 37-42.
6. Интеллектуальные системы, базовані на онтологіях [Текст] / Д.Г. Досин, В.В. Литвин, Ю.В. Нікольський, В.В. Пасічник. – Львів: Цивілізація, 2009. – 414 с.
7. Литвин В.В. Мультиагентні системи підтримки прийняття рішень, що базуються на прецедентах та використовують адаптивні онтології [Текст] / В.В. Литвин // Радіоелектроніка. Інформатика. Управління. – 2009. – № 2 (21). – С. 120-126.
8. Литвин В.В. Інтеллектуальні агенти пошуку релевантних прецедентів на основі адаптивних онтологій [Текст] / В.В. Литвин // Математичні машини і системи. – 2011. – № 3. – С. 66-72.
9. Найханова Л.В. Технология создания методов автоматического построения онтологий с применением генетического и автоматного программирования [Текст]: монография / Найханова Л.В. – Улан-Удэ: Издательство БНЦ СО РАН, 2008. – 244 с.

**Турчак А.М., Егорова О.В., Златкин А.А.**

Черкасский государственный технологический университет

## **ГЕНЕРАЦИЯ МНОЖЕСТВА ЯДЕР ПРОДУКЦИОННЫХ ПРАВИЛ В СИСТЕМАХ УПРАВЛЕНИЯ ОЦЕНИВАНИЕМ ЗНАНИЙ**

### **Аннотация**

В статье проанализирована задача построения системы управления оцениванием знаний. Рассматривается способ получения знаний из полнотекстовых источников. Предложен подход к решению проблемы автоматического построения продукционных правил с использованием модифицированного многофакторного композиционного метода направленной оптимизации. Приведены структурные особенности и конструктивные элементы метода для древовидной формы представления потенциальных решений. Модифицированный метод предназначен для формирования библиотеки методов построения онтологий.

**Ключевые слова:** онтология, продукционные правила, эволюционные алгоритмы.

**Turchak A.M., Yegorova O.V., Zlatkin A.A.**

Cherkasy State Technological University

## **GENERATION OF MULTIPLE CORES PRODUCTION RULES IN A KNOWLEDGE EVALUATION CONTROL SYSTEM**

### **Summary**

In this paper the problem of building a knowledge evaluation system was analyzed. Method for obtaining knowledge from the full-text sources has been considered. The approach to solve problem of automatic building of production rules by the use of the modified method of multivariate composite directed optimization was offered. The structural features and structural elements of this method for potential solutions as trees were suggested. The modified method is meant for forming library of constructing ontologies methods.

**Keywords:** ontology, production rules, evolutionary algorithms.