

УДК 519.816

МЕТОД ЗМІСТОВОЇ ІДЕНТИФІКАЦІЇ ОБ'ЄКТІВ БАЗ ЗНАНЬ СИСТЕМ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ

О.В. Андрійчук

Інститут проблем реєстрації інформації НАН України

e-mail: andreychuck@ukr.net

Вступ

Згідно з [1] основними вимогами до подачі знань є однорідність представлення, простота розуміння, несуперечливість та повнота. Граф ієрархії цілей відповідає вище зазначеним вимогам, значить він є однією з моделей подачі знань. Тому ієрархія цілей систем підтримки прийняття рішень (СППР) – це база знань (БЗ), а кожна ціль ієрархії – це об'єкт БЗ.

В процесі визначення та ідентифікації цілей ієрархії експертами в основному покладаються на чіткість та однозначність їхніх формулювань. Але при досить великих розмірах БЗ існує можливість помилкового введення однакових за змістом цілей. При об'єднанні декількох ієрархій, що сформульовані експертними групами різної спеціалізації, можлива ситуація, коли різними формулюваннями описується одна й та ж сама ціль. Все це може погіршити адекватність побудованої БЗ моделі предметної області.

В доповіді про прогнозний граф для прогнозування і керування науковими дослідженнями [2] В.М. Глушков, серед наукових проблем математичної обробки, відзначав важливе питання ототожнення близьких за значенням проблем, яке на той час до кінця не вміли вирішувати. Воно полягало у визначенні семантично подібних науково-технічних проблем, оскільки потрібно завжди ототожнювати близькі проблеми, які вирішуються більш менш схожими засобами. З іншого боку існують тонкі приклади, коли проблеми зовсім мало відрізняються в постановці і начебто їх треба ототожнювати, а насправді за методами розв'язання вони цілком різні, і тому їх треба розвести (не ототожнювати). Розв'язання цього питання було актуальним для ефективного використання методики прогнозного графа і щоб запобігти розростанню графу до нескінченності. Проблеми мали відрізнитись якісно, а не тільки за деякими параметрами. Передбачалося використання спеціального органу експертів, типу інституту наукової інформації з кваліфікованими людьми, який би міг розв'язувати проблеми такого роду, а у випадку неможливості для нього, – для остаточного винесення судження потрібно було звертатися до спеціалістів більш високої кваліфікації.

Суть підходу

Для розв'язку вищезазначених задач передбачається використання методу змістової ідентифікації об'єктів баз знань СППР. Змістова ідентифікація полягає у визначенні чи однакові два об'єкта БЗ за змістом. В основі методу лежить підхід аналогічний до застосовуваного при оцінці компетентності експертів [3]. Суть підходу полягає в наступному: для відображення змістової сутності, кожній цілі ієрархії співставляється множина ключових слів (КС) з відповідними коефіцієнтами важливості. Ця множина має мінімальну потужність для збереження унікальності змісту цілі. Коефіцієнти важливості КС визначаються експертним шляхом і представляють собою деякі дійсні числа. Змістова ідентифікація відбувається шляхом порівняння множин КС та їхніх коефіцієнтів важливості.

Постановка задачі

Задача змістової ідентифікації об'єктів баз знань СППР може зводитись до наступної загальної постановки.

Дано: дві цілі G_1 та G_2 з відповідними кортежами ключових слів, що мають відповідні нормовані вагові коефіцієнти w_{1i} та w_{2i} , причому $0 \leq w_{1i} \leq 1$, $0 \leq w_{2i} \leq 1$, $i = \overline{1, n}$, $\sum_{i=1}^n w_{1i} = 1$, $\sum_{i=1}^n w_{2i} = 1$, де n – кількість КС у кортежах (поки що розглядаємо випадок, коли кількість КС у кортежах однакова); симетрична матриця подібності КС S , де $S = \|s_{ij}\|_{i=\overline{1, m}; j=\overline{1, m}}$, $0 \leq s_{ij} \leq 1$ & $s_{ij} = s_{ji}$, $i = \overline{1, m}$, $j = \overline{1, m}$, де m – кількість КС; відносне відхилення d , де $0 < d \leq 1$.

Потрібно визначити: чи є дані цілі подібними в межах заданого відносного відхилення d ?

Суть методу

Розглянемо суть методу. При формулюванні цілі експертом, для характеристики її змістовної сутності та забезпечення однозначної відповідності сутності уявленням експерта, задається множина КС з відповідними їм коефіцієнтами важливості. Ця множина КС може бути задана експертами, сформульована виходячи з нормативних документів або базуватись на раніше побудованих ієрархіях. Множина КС об'єкту БЗ має бути мінімальної потужності і при цьому містити всю необхідну інформацію для його ідентифікації. Важливість КС для об'єкта БЗ визначаються експертним шляхом. Вагові коефіцієнти КС можуть визначатись за матрицею парних порівнянь, наприклад, методом власного вектору [4], «трикутник», «квадрат» [3], комбінаторним [5, 6] чи іншим методом обробки експертної інформації. Будується загальна для проблемної області матриця подібності КС на основі порівняння всіх КС між собою за семантичною подібністю.

Для ідентифікації подібних об'єктів БЗ, спочатку виконується пошук об'єктів, в яких співпадають множини КС, потім – пошук об'єктів з множинами КС, що збігаються неповністю, тобто коли деякі КС замінені на синоніми або подібні за змістом КС, відповідно до загальної матриці подібності КС. При цьому нові вагові коефіцієнти заміненних ключових слів v_i визначаються як $v_i = w_i s_{ij}$, а вагові коефіцієнти інших КС перенормовуються. Потім серед знайденого виділяються об'єкти, для яких співпадає кортеж КС (ранжирування КС за коефіцієнтами важливості). Після цього перевіряються умова відносного відхилення коефіцієнтів впливу: $\max_i |w_{1i} - w_{2i}| \leq d$. Далі в автоматичному режимі виділяються об'єкти, що відповідають заданому допустимому відносному відхиленню коефіцієнта впливу КС, а в напівавтоматичному режимі – виділяються об'єкти шляхом підбору значення цього відносного відхилення.

Слід зазначити, що всі КС об'єктів БЗ містяться у відповідній семантичній мережі, яка використовується для побудови загальної матриці подібності. Оскільки кількість КС у об'єкта БЗ може бути значна, то з урахуванням психофізіологічних обмежень людини [7], проводиться кластеризація множини КС на групи для порівняння не більше ніж по 9 КС. Також слід відмітити, що метод змістової ідентифікації доцільно використовувати при досить великих розмірах БЗ.

Випадок різних КС у кортежах різної довжини

Розглянемо випадок різних КС у кортежах різної довжини. При цьому, на відміну від попереднього випадку, цілі G_1 та G_2 мають відповідні кортежі КС з відповідними нормованими ваговими коефіцієнтами w_{1i} та w_{2j} , де $0 \leq w_{1i} \leq 1$, $0 \leq w_{2j} \leq 1$, $i = \overline{1, p}$, $j = \overline{1, r}$, $\sum_{i=1}^p w_{1i} = 1$, $\sum_{j=1}^r w_{2j} = 1$, де p – кількість КС у кортежі цілі G_1 , r – кількість КС у кортежі цілі G_2 , причому нехай нумерація буде такою, щоб перші n КС у обох кортежах були ті ж самі. Перенормуємо перші n вагових коефіцієнтів обох кортежів наступним

чином: $\sum_{i=1}^n w'_{1i} = 1$, $\sum_{i=1}^n w'_{2i} = 1$, де $w'_{1i} = w_{1i}k_1$, $w'_{2i} = w_{2i}k_2$, $i = \overline{1, n}$, $k_1 = \frac{1}{\sum_{i=1}^n w_{1i}}$, $k_2 = \frac{1}{\sum_{i=1}^n w_{2i}}$. Тоді

умова відносного відхилення коефіцієнтів впливу перевіряється наступним чином:

$\max_i |w'_{1i} - w'_{2i}| + k_1 \sum_{i=n+1}^p w_{1i} + k_2 \sum_{i=n+1}^r w_{2i} \leq d$, оскільки додавати потрібно величини однакової розмірності.

Інші можливі варіанти використання методу

Також існують інші можливі варіанти використання методу змістової ідентифікації. При встановленні впливів в ієрархії цілей, коли експертові потрібно вибрати всі підцілі, що впливають на деяку ціль, метод може використовуватись для пошуку формулювань цілей з відомою експертові семантикою, але з невідомим формулюванням. В раніше побудованих ієрархіях можна знаходити помилково введені однакові за семантикою цілі, а також здійснювати пошук відсутніх, але необхідних цілей ієрархії. Також можливе використання методу як додаткового інструменту для аналізу адекватності БЗ побудованої моделі предметної області.

Накопичені в процесі роботи методу дані (а саме: КС та матриця подібності КС) можна використовувати при урахуванні компетентності експертів, а також для формування експертних груп у відповідній предметній області.

Висновки

Запропоновано підхід та метод змістової ідентифікації об'єктів баз знань систем підтримки прийняття рішень за відповідними кортежами ключових слів, важливість яких визначається експертним шляхом. Використання цього підходу та методу дозволяє запобігти помилковому пропущенню однакових за змістом цілей при введенні нової цілі та при об'єднанні ієрархій цілей, сформованих різними експертними групами. Запропонований метод також може використовуватись для пошуку цілей ієрархії з невідомим формулюванням.

Література

1. Таран Т.А. Штучний інтелект. Теорія і застосування. Навчальний посібник [Текст] / Т.А.Таран, Д.А.Зубов. – Луганськ: Вид-во СНУ ім. В. Даля, 2006. – 240 с.
2. Глушков В.М. Прогнозування і керування науковими дослідженнями [Електрон. ресурс] / В.М. Глушков // Наукові доповіді на засіданнях Президії НАН України 1970-2001 рр; ІПРІ НАН України. - Київ, 2008. – 1 DVD.
3. Тоценко В.Г. Методы и системы поддержки принятия решений. Алгоритмический аспект [Текст] / В.Г.Тоценко – К.: Наукова думка, 2002. – 382 с.
4. Саати Т. Принятие решений. Метод анализа иерархий [Текст] / Т. Саати. – Москва: Радио и связь, 1993. – 278 с.
5. Циганок В.В. Комбінаторний алгоритм парних порівнянь зі зворотним зв'язком з експертом [Текст] / В.В. Циганок // Реєстрація, зберігання і обробка даних. – 2000. – Т.2, №2. – С.92-102.
6. Tsyganok V.V. Investigation of the aggregation effectiveness of expert estimates obtained by the pairwise comparison method. Mathematical and Computer Modelling, – v.52. – №3-4, August 2010, pp.538-544.
7. Миллер Дж. Магическое число сем плюс или минус два. О некоторых пределах нашей способности перерабатывать информацию [Текст] / Дж. Миллер – Инженерная психология. – М.: Прогресс, 1964. С. 192–225. – 696 с.