

МЕТОД КІЛЬКІСНОГО ОЦІНЮВАННЯ ВІДПОВІДЕЙ У СИСТЕМАХ ТЕСТУВАННЯ ЗНАНЬ ВІДКРИТОГО ТИПУ

І.В. ЗАМАРУЄВА, Л.М. БАДЬОРІНА

Розглянуто метод оцінювання знань тих, кого навчають, шляхом порівняльного аналізу тексту відповіді із заданим еталонним текстом та визначенням їх релевантності, наведено практичні розрахунки. Особливу увагу приділено дослідженню лінгвістичної моделі подання знань на основі синонімії термінів предметної галузі.

ВСТУП

Одним із найважливіших критеріїв засвоєння навчального матеріалу є рівень володіння майбутнім фахівцем понятійним апаратом із відповідної навчальної дисципліни. Сьогодні оцінювання відповідей, написаних у довільній формі, тими, кого навчають, є виключно прерогативою викладача, що вносить певну долю суб'єктивізму. З іншого боку, розвиток сучасних інформаційних технологій надає нові можливості щодо моделей, методів і принципів побудови автоматизованих систем контролю знань для тих, хто навчається. Одним із шляхів підвищення якості засвоєння навчального матеріалу та об'єктивізації контролю знань тих, хто навчається, є розробка знання-орієнтованого підходу до автоматизованого контролю знань. Вони подані у вигляді довільного природномовного тексту (ПМТ) та базуються на методах розпізнавання, вилучення і формалізації знань.

Загальна принципова особливість задачі оцінювання знань полягає в тому, що предметом аналізу для викладача є саме зміст відповіді (тексту) або знання, які там містяться. Урахування цієї особливості обумовлює необхідність вирішення проблеми моделювання процесу розуміння людиною текстової інформації. Основними компонентами знань із погляду їх формалізованого подання є поняття, відношення між поняттями, характеристики понять і відношень, а також модальності цих характеристик [1].

Отже, сутність знання-орієнтованого підходу до автоматизації задачі оцінювання знань за довільним текстом полягає в тому, що текст має бути підданий лінгвістичній обробці, спрямованій на розпізнавання в тексті основних компонент знань і встановлення логіко-семантичних зв'язків між ними. Подання знань у вигляді пов'язаної визначеними відношеннями сукупності понять із відповідними характеристиками та іншими атрибутами будемо називати поняттєвою структурою (ПС) змісту вхідного тексту.

Аналіз стану теоретичних робіт у галузі комп'ютерної лінгвістики та штучного інтелекту свідчить про те, що сьогодні відсутній теоретичний апарат, реалізувавши який можна створити програмні засоби автоматизації формування ПС. Водночас, створено достатній теоретичний доробок у цій галузі, який є основою для його узагальнення і подальшого розвитку

в напрямі автоматизації розпізнавання та вилучення знань, які містяться в ПМТ, їх формалізації в інтересах вирішення різноманітних задач, у тому числі й задачі оцінювання знань тих, хто навчається, за відповідями, представленими в ПМТ.

ОЦІНЮВАННЯ ТЕКСТОВИХ ВІДПОВІДЕЙ

Знання — орієнтований підхід до автоматизації оцінювання знань тих, хто навчається, за текстовими відповідями передбачає наявність засобів приведення еталонних відповідей, представлених природною мовою, і відповідей тих, хто навчається, до формалізованого подання у вигляді моделі знань про предметну галузь (ПС). Кожна модель відповіді тих, хто навчається, порівнюється на відповідність із еталонною моделлю. Як було сказано вище, ПС може містити різноманітні логіко-семантичні відношення між поняттями, по кожному з яких необхідно встановлювати ступінь відповідності. Введемо такі обмеження на оцінювання відповідності відповідей еталонному зразку:

- розглядаємо в якості відповідей лише визначення (тлумачення) термінів і понять із певної навчальної дисципліни;
- на множині відношень, які задано для термінів та понять із відповідної навчальної дисципліни, виділимо тільки родо-видові відношення та відношення синонімії.

Визначення поняття — у самому широкому розумінні є логічна операція, в процесі якої розкривається зміст поняття. В основі правил еталонних визначень термінів і понять покладено 7 правил, що вивчаються формальною логікою [2]:

- поняття визначається через родові й видові відмінності;
- визначення повинно мати такий же вимір, що і поняття, тобто обсяг значення поняття, що визначається, і поняття, за допомогою якого здійснюється визначення, мають відповідати одне одному;
- видовою відмінністю має бути ознака або група ознак, які притаманні лише цьому поняттю, і відсутні в інших поняттях, що відносяться до того ж родового поняття;
- визначення не має містити поняття, що визначається, не має визначатися через поняття, яке стає зрозумілим лише через поняття, що визначається;
- визначення не має бути тільки заперечним, оскільки заперечення вказує на відсутні ознаки і не дає суттєвих ознак, що характеризують це поняття;
- визначення не має бути суперечливим з точки зору формальної логіки;
- визначення має бути ясним, чітким і не містити подвійного тлумачення.

Нехай S — множина всіх еталонних визначень понять і термінів із відповідної навчальної дисципліни, поданих у вигляді природно-мовного тексту й укладених за визначеними вище правилами. Множина S є кінцевою й неупорядкованою: $S = \{s_i : 1 \leq i \leq n\}$, де s_i — визначення терміна; n — ціле число.

Сукупність відповідей тих, хто навчається, виражено також природною мовою, позначимо як множину T . Ця множина є підмножиною множини S та має всі її властивості: $T \subset S$; $T = \{t_i : 1 \leq i \leq m\}$, де m — ціле число; $m \leq n$.

Кожна відповідь з множини T може містити терміни і поняття, які пов'язані родо-видовими відношеннями, або відношеннями синонімії з поняттями відповідної еталонної відповіді множини S . Взаємозв'язок термінів і понять у заданій предметній галузі (навчальній дисципліні) представимо у вигляді тезаурусу. Тезаурус — словник, що відбиває семантичні відношення між поняттями в певній предметній галузі й призначений для пошуку заданого слова за його смисловими зв'язками з іншими словами [3].

Структура тезаурусу, як правило, включає такі відношення:

поняття := <рід-вид> <частина-ціле> <синоніми> <антоніми> <асоціації>.

Відношення рід-вид дозволяє включити у пошукове поле більш абстрактні або конкретні поняття. Відношення частина-ціле включає у пошукове поле частини цілого об'єкту. Відношення синонімії й антонімії дозволяє здійснювати пошук синонімів й антонімів. Відношення асоціації різноманітні та індивідуальні за своєю природою і вказують на контекстну залежність пошукового поняття.

Відповідь того, хто навчається, визначається певною структурою понять і термінів, тобто ПС. З урахуванням визначених обмежень кожне поняття в тлумачній частині може описуватися через синоніми.

Елемент e , відносно якого утворюється множина (тобто синонімічний ряд) D_e , назовемо базовим термом, інші елементи множини D_e (слова-синоніми) назовемо залежними термами. Необхідно встановити відповідність між термами еталонного визначення і термами відповіді, спираючись на поняття синонімічної відповідності термів, використовуючи тезаурус. І це, в свою чергу, дасть змогу обчислити показник релевантності та визначити співвідношення базових і еталонних термів.

Таким чином, еталонне визначення слід розглядати як сукупність базових термів e , а відповідь як сукупність термів t , для кожного необхідно знайти відповідний базовий терм [4].

Якщо A — множина термів еталонного визначення, B — множина термів відповіді, то формалізоване подання еталонного визначення і відповіді буде мати такий вигляд:

$$A = \{e_1, e_2, \dots, e_i, 1 \leq i \leq N\},$$

де N — кількість термів еталонного визначення;

$$B = \{t_1, t_2, \dots, t_i, 1 \leq i \leq M\},$$

де M — кількість термів відповіді.

У результаті ми можемо отримати одне з таких співвідношень між множинами A і B .

1. $A = B$ — відповідь того, хто навчається, повністю збігається з еталонною відповіддю.
2. $A \subset B$ — відповідь того, хто навчається, містить всі терми з еталонної відповіді, а також додаткові терми.

3. $B \subset A$ — відповідь того, хто навчається, частково відповідає еталонній відповіді, у ній відсутні деякі базові терми.

4. $A \cap B = \emptyset$ — відповідь того, хто навчається повністю не відповідає еталонній відповіді.

5. $A \cap B \neq \emptyset$ — еталонна відповідь і поточна відповідь спільні терми [4].

Продемонструємо викладене вище на наступному прикладі. Припустимо, що ми маємо еталонне визначення: «Програма — опис алгоритму розв’язання задачі, заданий на мові обчислювальної машини» [3].

В еталонному визначенні виділено ключові базові терми, які відповідають умовам необхідності та достатності правильної відповіді для тих, хто навчається. Інші поняття є додатковими. Вони також можуть мати синонімічні ряди, які не враховуються під час кількісного оцінювання відповіді того, хто навчається. Тобто, для правильної відповіді визначається два необхідних і достатніх поняття, які за правилами побудови тлумачної частини терміну «програма» формують його унікальні відмінні ознаки. Для цих базових термів у тезаурісу можна побудувати такий синонімічний ряд:

Алгоритм := {сукупність правил; послідовність операцій; сукупність дій};

Мова обчислювальної машини := {мова програмування; штучна мова; машинна мова; формальна мова, мова ЕОМ}.

Позначимо через **A1** множину, що визначає синонімічний ряд для поняття «алгоритм», і через **A2** — синонімічний ряд для поняття «мова обчислювальної машини».

Тоді формалізоване подання еталонної відповіді буде мати такий вигляд:

Програма :=

опис {представлен}

A1 \subset {алгоритм; сукупн+правил; послідовн+операцій; сукупн+дій}
розв’язання {вирішення; обчислення}
задач

\wedge
зadan {представлен; опис}

на **A2** \subset {мов+обчислювальн+машин; мов+програмування; машин+мов; формальн+мов; мов+ЕОМ}.

У цьому прикладі поняття, через які відбувається тлумачення, подано як пошукові образи через знак «+» поєднуються слова, які складають термін для заданої навчальної дисципліни, а логічна операція \wedge вказує на обов’язкову наявність двох базових термів. Інші відношення в силу введених раніше обмежень пропущені.

Це представлення є основою для порівняння з поточними відповідями тих, хто навчається. При цьому будемо вважати, терми a (з еталонної відповіді) і b (відповіді того, хто навчається) збігаються повністю, якщо для b знайдено хоча б один пошуковий образ із синонімічного ряду терму a . Тобто, для двох термів a та b можна визначити функцію $f(a, b)$. Вона характеризує показник повноти тлумачення терміну через поняття, що його описують по відношенню синонімії, та приймає значення 1, якщо $b \subset A$, а також приймає значення 0, якщо $A \cap b = \emptyset$.

Тоді для заданого еталонного зразка необхідну і достатню умови правильної відповіді, яку можна записати таким чином:

$$f(a, b) = 1, \text{ якщо } ((\mathbf{A1} \cap \mathbf{B1}) = \emptyset \wedge (\mathbf{A2} \cap \mathbf{B2}) = \emptyset). \quad (1)$$

Оскільки, відповіді тих, хто навчається, можуть бути неповними, тобто не всі ключові поняття, що описують термін, наявні в тлумачній частині, то у випадку відсутності у відповіді хоча б одного ключового терму значення $f(a, b) = 0$. У той же час при кількісному оцінюванні того, хто навчається, є потреба оцінити повноту відповіді за кожним ключовим термом тлумачної частини. Для цього введемо інтервал оцінювання $[0, 1]$. Функція $f(a, b)$ приймає значення 0 або 1 для кожного ключового терму, що входить до еталонного зразка, загальна сума балів за відповідь визначається за формулою:

$$K = \frac{\sum_{i=1}^n f(a_i, b_i)}{N}, \quad (2)$$

де N — загальна кількість ключових термів тлумачної частини, що підлягають оцінюванню.

Нехай для визначеного еталонного зразка ми маємо таку фактичну відповідь того, хто навчається: «Програма — послідовність операцій над даними, необхідними для обробки інформації та реалізації заданого алгоритму».

Ця відповідь приводиться до формалізованого подання. При цьому формуються пошукові образи кожного слова, що входить до тлумачної частини терміну, і порівнюються з елементами множини $\mathbf{A1}$ (визначає синонімічний ряд для терму «алгоритм»), і $\mathbf{A2}$ (визначає синонімічний ряд для терму «мова обчислювальної машини»). Інші слова також можуть перевірятися із синонімічними рядами термів, які не є ключовими для тлумачення терміну «програма», але вони не враховуються під час оцінювання відповіді того, хто навчається. Після необхідних перетворень формалізоване подання відповіді того, хто навчається буде мати такий вигляд:

Програма :=
 $\mathbf{A1}$ {послідовн+операцій}
над дан
необхідн
обробк/обробок
інформац
реалізац
 $\mathbf{A1}$ {алгоритм}.

Із наведеного прикладу видно, що поняття тлумачної частини терміну «програма» збігаються тільки з множиною $\mathbf{A1}$ еталонного зразка. Причому у відповіді знайдено 2 еквіваленти, оскільки з формули (1) випливає, що вона приймає значення 1, якщо знайдено хоча б один відповідник, тому згортання всіх знайдених відповідників з однієї множини дає значення 1, тобто $f(a_1, b) = 1, f(a_2, b) = 0$.

Кількісна оцінка обчислюється за формулою (2): $K = 0,5$.

Таким чином, якщо привести інтервал $[0, 1]$ до десятибальної шкали оцінювання, то ця відповідь буде мати оцінку 5.

ВИСНОВКИ

Запропоновано метод кількісного оцінювання відповідей тих, хто навчається, який дозволяє визначити і надати кількісну оцінку семантичній відповідності між текстовими відповідями й еталонним визначенням. Сутність терміну полягає у визначенні так званих опорних понять у тлумачній частині еталонної відповіді, пошуку і встановлення відповідності цих понять у відповідях тих, хто навчається, та приведення отриманого результату до кількісної шкали оцінювання. Особливістю методу є урахування тезаурусних відношень між поняттями тлумачної частини, зокрема, рід-вид, частина-ціле, причина-наслідок, а також відношення синонімії.

Це дає змогу певним чином об'єктивізувати процес оцінювання рівня володіння тими, хто навчається, та понятійним апаратом із заданої предметної галузі.

ЛИТЕРАТУРА

1. *Замаруєва І.В.* Комп'ютерна модель розуміння природно-мовної текстової інформації // Проблемы программирования. — 1999. — № 2. — С. 96–102.
2. *Кондаков Н.И.* Логический словарь-справочник. — М.: Наука, 1975. — 719 с.
3. *Словарь по кибернетике: Св. 2000 ст.* / Под ред. В.С. Михалевича. — 2-е изд. — Киев: Гл. ред. УСЭ им. М.П. Бажана, 1989. — 751 с.
4. *Бадьоріна Л.М.* Метод оцінювання довільних відповідей у комп'ютерних системах тестування знань // Математичні машини і системи. — 2006. — № 4. — С. 138–144.

Надійшла 09.06.2010