УДК 681.3

ДОСТОВІРНІСТЬ НЕЧІТКОСТІ: ТЕОРІЯ ТА ЗАСТОСУВАННЯ

О.І. Провотар, О.О. Провотар

Розглядається підхід до знаходження достовірнісних оцінок нечітких подій в системах нечіткого логічного виведення. Такі системи використовуються для подання нечітких знань, зокрема, в експертних системах, системах розпізнавання образів, системах прогнозування, тощо. За допомогою громіздких інтелектуальних обчислень здійснюється видобування знань в таких системах, які потребують оцінки їх достовірності. Приводиться приклад застосування запропонованого підходу в експертних діагностичних системах з нечіткими моделями подання знань.
Ключові слова: нечітка подія, ймовірність, достовірність.

Рассматривается подход к нахождению достоверностных оценок нечетких событий в системах нечеткого логического вывода. Такие системы используются для представления нечетких знаний, в частности, в экспертных системах, системах распознавания образов, системах прогнозирования и тому подобное. С помощью громоздких интеллектуальных вычислений осуществляется извлечения знаний в таких системах, требующих оценки их достоверности. Приводится пример применения предложенного подхода в экспертных диагностических системах с нечеткими моделями представления знаний.
Ключевые слова: нечеткое событие, вероятность, достоверность.

An approach to finding a credible estimates of fuzzy events in fuzzy inference systems is considered. Such systems are used to represent fuzzy knowledge, in particular, in expert systems, pattern recognition systems, forecasting systems, and so on. The extraction of knowledge in such systems is carried out with the help of cumbersome intellectual calculations, and the knowledge obtained requires an assessment of their credibility. An example of application of the proposed approach in expert diagnostic system with fuzzy model of presentation of knowledge is given.
Key words: fuzzy event, probability, credibility.

Вступ

Відомо, що зручним інструментом для подання знань в інформаційних системах є нечіткі системи логічного виведення (нечіткі специфікації) [1–6], які будуються на основі ідей та методів індуктивної математики. Використання таких систем є досить зручним при роботі з нечіткими знаннями. Але в багатьох випадках виникає потреба не тільки в знаходженні нечітких знань, але і в оцінці їх достовірності. В роботі пропонується підхід, який дозволяє одержувати оцінки достовірності нечітких знань в нечітких системах логічного виведення.

Під нечіткою специфікацією задачі розуміють упорядковану множину нечітких інструкцій. Нечітку специфікацію задачі разом з алгоритмом, при виконанні якого одержуємо наближене (нечітке) розв’язання цієї задачі називатимемо нечіткою системою логічного виведення.

Нехай $x_1, x_2, \ldots, x_n$ вхідні лінгвістичні змінні і y – вихідна лінгвістична змінна [1–5, 8]. Упорядкована множина нечітких інструкцій має вигляд:

якщо $x_1 \in A_{11} \wedge \ldots \wedge x_n \in A_{1n}$ ***то*** $y \in B_1$,

якщо $x_1 \in A_{21} \wedge \ldots \wedge x_n \in A_{2n}$ ***то*** $y \in B_2$,

. . . .

якщо $x_1 \in A_{m1} \wedge \ldots \wedge x_n \in A_{mn}$ ***то*** $y \in B_m$,

де A_{ij} и B_i – нечіткі множини, символ “$\wedge$” інтерпретується як *t*-норма нечітких множин.

Алгоритм обчислення виходу такої специфікації при входах $A_1', \ldots, A_n'$ полягає у виконанні наступних кроків:

1. Обчислюємо рівні істинності правил:

$$\alpha_i = \min[\max(A_1'(x_1) \wedge A_{i1}(x_1), \ldots, \max(A_n'(x_n) \wedge A_{in}(x_n))] ;$$

2. Обчислюємо виходи кожного правила:

$$B_i'(y) = \min(\alpha_i, B_i(y));$$

3. Обчислюємо агрегатний вихід:

$$B'(y) = \max\,(B'_1(y), \ldots, B'_m(y))$$

Ймовірність нечіткої події. У випадку подання знань нечіткими системами логічного виведення важливим є питання достовірності агрегованого виходу таких систем. Один із підходів до розв'язання цієї задачі полягає у знаходженні ймовірнісних оцінок одержаних результатів [9]. Як відомо [10], щоб визначити ймовірність події A у просторі елементарних подій $\boldsymbol{X}$, вводиться поняття ймовірнісної міри. Це числова функція P, яка ставить у відповідність число $P(A)$ елементарній події A, причому:

$$0 \le P(A) \le 1,\ P(\boldsymbol{X}) = 1,\ P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)\,.$$

для будь-яких $A_1, A_2, \ldots$ таких, що $A_i \cap A_j = \varnothing$, при $i \ne j$.

Нечіткою подією A в просторі $\boldsymbol{X}$ будемо називати нечітку множину

$$A = \{(x, \mu_A(x)), x \in \boldsymbol{X}\},$$

де μ_A: $\boldsymbol{X} \to [0,1]$ – функція належності нечіткої множини A. Тоді ймовірність події A можна обчислити за формулою

$$P(A) = \sum_{\boldsymbol{x} \in \boldsymbol{A}} \mu_A(x) P(x).$$

Враховуючи це, можна обчислювати ймовірності будь-яких нечітких подій при заданій ймовірнісній мірі.

Умовною ймовірністю події A за умови виконання події B називають ймовірність події A, що обчислена з припущенням того, що відбулась подія B. Позначають таку умовну ймовірність наступним чином

$$P(A|B) \text{ або } P_B(A).$$

В загальному випадку знайти умовну ймовірність в класичному розумінні ймовірності досить просто і можна це зробити наступним чином. Нехай з n взаємовиключних та рівно ймовірних елементарних подій $A_1, A_2, \ldots A_n$

події A сприяє m елементарних подій,

події B сприяє k елементарних подій,

події AB сприяє r елементарних подій,

(зрозуміло, що $r \le k$, $r \le m$). Якщо подія B відбулась, то це означає, що настало одна з елементарних подій A_j, що сприяє події B. При цій умові події A сприяє лише r і тільки r елементарних подій A_j, що сприяють AB. Таким чином отримуємо

$$P(A\,|\,B) = \frac{r}{k} = \frac{r/n}{k/n} = \frac{P(AB)}{P(B)}.$$

Звідки, ймовірність одночасної появи двох залежних подій буде дорівнювати добутку ймовірності однієї з них на умовну ймовірність іншої, обчисленої за умови, що перша відбулась, тобто

$$P(AB) = P(A)P(B\,|\,A) = P(B)P(A\,|\,B).$$

Це твердження називають теоремою множення для умовних ймовірностей.

З умовною ймовірністю подій тісно пов'язано поняття *незалежності подій*. Кажуть, що подія A *незалежна* від події B, якщо має місце рівність

$$P(A|B) = P(A),$$

тобто якщо настання події B ніяким чином не змінює ймовірності настання події A. Властивість незалежності подій є взаємним, тобто якщо подія A незалежна від події B, то подія B також незалежна від події A і навпаки. В цьому легко можна переконатись, використовуючи теорему множення.

З теореми множення також можна отримати альтернативне означення незалежності подій, а саме, якщо A та B незалежні події то виконується наступна рівність

$$P(AB) = P(A)P(B),$$

і, навпаки, якщо виконується рівність (3) то події A та B незалежні.

Умовну ймовірність нечіткої події A за умови виконання нечіткої події B називають будемо визначати за Демпстером [7]. А саме, функція розподілу $P_{(A|B)}$ умовної ймовірності нечіткої події A за умови виконання нечіткої події B визначається через функцію розподілу $P_{(A,B)}$ бінарної ймовірності декартового добутку $A \times B$ та функцію розподілу P_B ймовірності нечіткої події B, за умови що вона не дорівнює нулю, тобто для будь-якої пари (x,y) з декартового добутку просторів $\boldsymbol{X} \times \boldsymbol{Y}$ виконується

$$Q_{(A\times B)}(x,y) = \begin{cases} \dfrac{P_{(A,B)}(x,y)}{P_B(y)}, & P_B(y) \neq 0 \\ 1, & P_B(y) = 0. \end{cases},$$

$$P_{(A|B)}(x,y) = \left\{ \dfrac{Q_{A\times B}(x,y)}{\sum\limits_{x,y} Q_{A\times B}(x,y)} \right. .$$

Враховуючи це, можна обчислювати умовні ймовірності будь-яких нечітких подій при заданій ймовірнісній мірі.

Функцію розподілу бінарної ймовірності декартового добутку $A \times B$ будемо обчислювати за формулою

$$P_{(A,B)}(x,y) = \min(P_A(x), P_B(y)).$$

Маючи метод обчислення умовної ймовірності, можна обчислити ймовірність хвороби при заданій симптоматиці.

Приклад. Нехай $\boldsymbol{X}_1$ = {5, 10}, $\boldsymbol{X}_2$ = {5, 10}, $\boldsymbol{X}_3$ = {36, 37, 38, 39, 40} – простори для визначення значень лінгвістичних змінних

x_1 = "Кашель" = {"*слабкий* (*К*)", "*помірний* (*К*)", "*сильний* (*К*)"},

x_2 = "Нежить" = {"*слабкий* (*Н*)", "*помірний* (*Н*)", "*сильний* (*Н*)"},

x_3 = "Температура" = {"*нормальна*", "*підвищена*", *висока*", "*дуже висока*"}

відповідно.

Визначимо елементи цих множин:

"Кашель": "*слабкий* (*К*)" = 1/5; "*помірний* (*К*)" = 0.5/5 + 0.5/10; "*сильний* (*К*)" = 1/10.

"Нежить": "*слабкий* (*Н*)" = 1/5; "*помірний* (*Н*)" = 0.5/5 + 0.5/10; "*сильний* (*Н*)" = 1/10.

"Температура": "*нормальна*" = 1/36 + 0.5/37; "*підвищена*" = 1/37 + 0.5/38; "*висока*" = 1/38 + 0.5/39; "дуже *висока*" = 0.5/39 + 1/40.

Нехай $\boldsymbol{Y}$ = {Грип, ГРЗ, Ангіна, Запалення легенів} – простір для визначення значень лінгвістичної змінної y. Тоді залежність хвороби пацієнта від його симптомів може бути описана наступною системою специфікацій:

якщо x_1 є "*слабкий* (*К*)" $\wedge$ x_2 є "*слабкий* (*Н*)" $\wedge$ x_3 є "*підвищена*" ***то*** *y* є "0.5/Грип +0.5/ОРЗ +0.4/Ангіна + 0.8/Запалення легенів";

якщо x_1 є "*слабкий* (*К*)" $\wedge$ x_2 є "*помірний* (*Н*)" $\wedge$ x_3 є "*висока*" ***то*** *у* є "0.8/Грип +0.7/ОРЗ +0.8/Ангіна + 0.3/Запалення легенів";

якщо x_1 є "*слабкий* (*К*)" $\wedge$ x_2 є "*помірний* (*Н*)" $\wedge$ x_3 є "дуже *висока*" ***то*** *у* є "0.9/Грип +0.7/ОРЗ +0.8/Ангіна + 0.2/Запалення легенів".

Якщо на вхід x_1 цієї системи специфікацій подати величину A_1' = 1/5 + 0.5/10, на вхід x_2 – величину A_2' = 1/5 + 0.5/10, на вхід x_3 – величину A_3' = 1/38, то у відповідності з алгоритмом виконання системи нечітких специфікацій одержимо нечітке розв'язання задачі

B' = 0.5/Грип + 0.5/ГРЗ + 0.5/Ангіна + 0.5/Запалення легенів.

Отже, треба знайти ймовірність хвороби

B' = 0.5/Грип + 0.5/ГРЗ + 0.5/Ангіна + 0.5/Запалення легенів

при симптомах A_1' = 1/5 + 0.5/10, A_2' = 1/5 + 0.5/10, A_3' = 1/38 відповідно. Крім того, нехай розподіли ймовірностей у просторах $\boldsymbol{X}_1$ = {5, 10}, $\boldsymbol{X}_2$ = {5, 10}, $\boldsymbol{X}_3$ = {36, 37, 38, 39, 40}, $\boldsymbol{Y}$ = {Грип, ГРЗ, Ангіна, Запалення легенів} задаються як

"Кашель": $P_{X_1}(5) = 0.4,\ P_{X_1}(10) = 0.6;$

"Насморк": $P_{X_2}(5) = 0.4,\ P_{X_2}(10) = 0.6;$

"Температура": $P_{X_3}(36) = 0.3,\ P_{X_3}(37) = 0.3,\ P_{X_3}(38) = 0.2,\ P_{X_3}(39) = 0.1,\ P_{X_3}(40) = 0.1;$

"Хвороба": P_Y(Грип) = 0.5, P_Y(ГРЗ) = 0.3, P_Y(Ангіна) = 0.1, P_Y(Запалення легенів) = 0.1.

Спочатку обчислимо ймовірності гіпотез – нечітких специфікацій логічного виведення. Для прикладу перетворимо першу гіпотезу

H_1 = ***якщо*** x_1 є "*слабкий* (*К*)" $\wedge$ x_2 є "*слабкий* (*Н*)" $\wedge$ x_3 є "*підвищена*" ***то*** *у* є "0.5/Грип +0.5/ГРЗ +0.4/Ангіна + 0.8/Запалення легенів"

до вигляду

H_1 = $\neg$(x_1 є "*слабкий* (*К*)") $\vee$ $\neg$(x_2 є "*слабкий* (*Н*)") $\vee$ $\neg$(x_3 є "*підвищена*") $\vee$ *у* є "0.5/Грип +0.5/ОРЗ +0.4/Ангіна + 0.8/Запалення легенів".

Знаходимо відповідні доповнення і одержуємо нечіткі множини:

$\neg$(x_1 є "*слабкий* (*К*)") = 1/10;

$\neg$(x_2 є "*слабкий* (*Н*)") = 1/10;

$\neg$(x_3 є "*підвищена*") = 1/36 + 0.5/38 + 1/39 + 1/40.

Далі обчислюємо ймовірності нечітких множин-подій:

Р ($\neg$(x_1 є "*слабкий* (*К*)")) = 0.6·1 = 0.6;

Р ($\neg$(x_1 є "*слабкий* (*Н*)")) = 0.6·1 = 0.6;

Р ($\neg$(x_3 є "*підвищена*")) = 0.3 + 0.1 + 0.1 + 0.1 = 0.6;

Р("0.5/Грип +0.5/ГРЗ +0.4/Ангіна + 0.8/Запалення легенів") = 0.25 + 0.15 + 0.04 + 0.08 = 0.52.

Тоді ймовірність першої гіпотези H_1 дорівнює:

$$P(H_1) = 0.58.$$

Аналогічно обчислюємо ймовірності гіпотез H_2 і H_3. У випадку цих гіпотез будемо мати

$$P(H_2) = 0.5675,\ P(H_3) = 0.6775.$$

На наступному кроці обчислимо умовні ймовірності $P(B/H_1)$, $P(B/H_2)$, $P(B/H_3)$. Алгоритм обчислення умовної ймовірності $P(B/H_i)$ полягає у виконанні наступних кроків:

1. Обчислюємо функцію розподілу бінарної ймовірності $P_{(B,H_i)}$:

$$P_{(B,H_i)}(x_1, \ldots, x_n, y) = \begin{matrix} \min[\max(P_{X_1}(x_1)\cdot\mu_{A_1'}(x_1), \ldots, P_{X_n}(x_n)\cdot\mu_{A_n'}(x_n),\ P_Y(y)\cdot\mu_{B'}(y)), \\ \max(P_{X_1}(x_1)\cdot\mu_{A_{i1}}(x_1), \ldots, P_{X_n}(x_n)\cdot\mu_{A_{in}}(x_n),\ P_Y(y)\cdot\mu_{B_i}(y))] \end{matrix}.$$

2. Обчислюємо ймовірнісну функцію декартового добутку за формулою

$$Q_{(B\times H_i)}(x_1,\ldots,x_n,y) = \begin{cases} \dfrac{P_{(B,H_i)}(x_1,\ldots,x_n,y)}{P_B(y)}, & P_B(y)\neq 0 \\ 1, & P_B(y)=0. \end{cases}$$

3. Обчислюємо функцію розподілу умовної ймовірності за формулою

$$P_{(B|H_1)}(x_1,\ldots,x_n,y) = \left\{ \frac{Q_{B\times H_1}(x_1,\ldots,x_n,y)}{\sum\limits_{x_1,\ldots,x_n,y} Q_{B\times H_1}(x_1,\ldots,x_n,y)} \right..$$

Обчислимо, для прикладу, значення

$$P_{(B,H_1)}(5,5,36,\text{Грип}),\ Q_{(B\times H_i)}(5,5,36,\text{Грип}) \text{ та } P_{(B|H_1)}(5,5,36,\text{Грип}).$$

Маємо:

$$P_{(B,H_1)}(5,5,36,\text{Грип}) = \min[\max(P_{X_1}(5)\cdot\mu_{A_1'}(5),\ P_{X_2}(5)\cdot\mu_{A_2'}(5),\ P_{X_3}(36)\cdot\mu_{A_3'}(36),\ P_Y(\text{Грип}))\cdot\mu_{B'}(\text{Грип})),$$

$$\max(P_{X_1}(5)\cdot\mu_{A_{11}}(5),\ P_{X_2}(5)\cdot\mu_{A_{12}}(5), P_{X_3}(36)\cdot\mu_{A_{13}}(36),\ P_Y(\text{Грип})\cdot\mu_{B_1}(\text{Грип}))] =$$

$$= \min[\max(0.4, 0.4, 0, 0.25), \max(0.4, 0.4, 0, 0.25)] = 0.4.$$

$$Q_{(B\times H_1)}(5,5,36,\text{Грип}) = P_{(B,H_1)}(5,5,36,\text{Грип}) / P_B(y) = 0.8.$$

$$P_{(B/H_1)}(5,5,36,\text{Грип}) =$$
$$= Q_{(B\circ H_1)}(5,5,36,\text{Грип}) / \sum_{x_1,\ldots,x_n,y} Q_{B\times H_1}(5,5,36,\text{Грип}) = 0.8/190 = 8/1900.$$

Аналогічно обчислюються значення функцій розподілу бінарної ймовірності, ймовірнісної функції декартового добутку та функції розподілу умовної ймовірності для інших значень аргументів.

Наступним кроком є обчислення декартових добутків $A_1' \times A_2' \times A_3' \times B'$ та $A_{11}\times A_{12}\times A_{13}\times B_1$ з подальшою їх агрегацією.

Після цього можна обчислити умовну ймовірність $P(B/H_1)$. А саме,

$$P(B/H_1) = \frac{131}{1425}.$$

Для обчислення ймовірності $P(B/H_2)$ знаходимо декартовий добуток $A_{21}\times A_{22}\times A_{23}\times B_2$ і обчислюємо умовну ймовірність

$$P(B/H_2)=\frac{77}{950}.$$

Для обчислення ймовірності $P(B/H_3)$ знаходимо декартовий добуток $A_{31}\times A_{32}\times A_{33}\times B_3$ і обчислюємо умовну ймовірність

$$P(B/H_3)=\frac{122}{950}.$$

Далі, використовуючи аналог формули повної ймовірності

$$P(B)=\sum_{i=1}^{n}P(H_i)P(B/H_i)$$

можна обчислити ймовірність події B, тобто ймовірність того, що вихід системи нечіткого логічного виведення є B' при входах A_1', A_2', A_3'. Отже, будемо мати:

$$P(B)=\sum_{i=1}^{3}P(H_i)P(B/H_i)=0.58\cdot\frac{131}{1425}+0.5675\cdot\frac{77}{950}+0.6775\cdot\frac{122}{950}\approx 0.2.$$

Висновки

Таким чином, запропонований в статті алгоритм, дозволяє обчислювати ймовірнісні оцінки для різних нечітких подій. Зрозуміло, що такі оцінки дуже важко інтерпретувати в категоріях частотних характеристик. Тому, для таких ймовірнісних оцінок нечітких подій пропонується ввести інший термін – достовірність. Отже, всі ймовірнісні оцінки нечітких подій, про які йдеться в даній статті, є не що інше як характеристика достовірності цих подій.

Література

1. Provotar O. Fuzzy Systems of Logical Inference and Their Application. Proceedings of 24-th International Workshop CS&P, 2015. Rzeszow, Poland, September 28-30. 2015. Vol. 2. P. 111–120.
2. Рутковская Д., Пилиньский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы. Москва: Телеком, 2006. 382 с.
3. Zadeh L.A. Fuzzy Sets. *Information and Control*. 1965. Vol. 8. P. 338–353.
4. Провотар А.И., Лапко А.В. О некоторых подходах к вычислению неопределенностей. *Проблеми програмування*. 2010. № 2–3. С. 22–27.
5. Zadeh L.A. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets ana Systems*. 1978. Vol. 1. P. 3–28.
6. Гупал А.М., Сергиенко И.В. Оптимальные процедуры распознавания. Киев: Наукова думка, 2008. 232 с.
7. Vejnarová J. Conditional Independence Relations in Possibility Theory. Int. J. *Uncertainty, Fuzziness and Knowledge-Based Systems*. 2000. N 8. P. 253–269.
8. Джексон П. Введение в экспертные системы. Москва: Вильямс, 2001. 624 с.
9. Zadeh L.A. Probability Measures of Fuzzy Events. *Journal of Mathematical Analysis and Applications*. 1968. Vol. 10. P. 421–427.
10. Гнеденко Б. Курс теории вероятностей. Москва: Едиториал УРСС, 2005. 448 с. (Учебник. Изд. 8-е, испр. и доп.).

References

1. Provotar O. Fuzzy Systems of Logical Inference and Their Application. Proceedings of 24-th International Workshop CS&P, 2015. Rzeszow, Poland, September 28-30. 2015. Vol. 2. P. 111–120.
2. Rutkovskaya D., Pilinsky M., Rutkowski L. Neural networks, genetic algorithms and fuzzy systems. Moscow: Telecom, 2006. 382 p.
3. Zadeh L.A. Fuzzy Sets. *Information and Control*. 1965. Vol. 8. P. 338–353
4. Provotar A.I., Lapko A.V. On some approaches to the calculation of uncertainties. *Problems of programming*. 2010. N 2–3. P. 22–27.
5. Zadeh L.A. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*. 1978. Vol. 1. P. 3–28.
6. Gupal A.M., Sergienko I.V. Optimal recognition procedures. Kiev: Naukova Dumka, 2008. 232 p.
7. Vejnarová J. Conditional Independence Relations in Possibility Theory. Int. J. *Uncertainty, Fuzziness and Knowledge-Based Systems*. 2000. N 8. P. 253–269.
8. Jackson P. Introduction to expert systems. Moscow: Williams, 2001. 624 p.
9. Zadeh L.A. Probability Measures of Fuzzy Events. *Journal of Mathematical Analysis and Applications*. 1968. Vol. 10. P. 421–427.
10. Gnedenko B. Course of the theory of probability. Moscow: Editors of the URSS, 2005. 448 p. (Textbook, edition 8th, corrected and supplemented.).

Про авторів:

Провотар Олександр Іванович,
доктор фізико-математичних наук, професор,
професор Київського національного університету імені Тараса Шевченка,
Professor of Rzeszow University.
Кількість наукових публікацій в українських виданнях – 100.
Кількість наукових публікацій в зарубіжних виданнях – 30.
Індекс Хірша – 4.
http://orcid.org/0000-0002-6556-3264,

Провотар Олександр Олександрович,
аспірант факультету комп’ютерних наук та кібернетики
Київського національного університету імені Тараса Шевченка.
Кількість наукових публікацій в українських виданнях – 7.
Кількість наукових публікацій в зарубіжних виданнях – 2.
http://orcid.org/0000-0001-7983-4996.

Місце роботи авторів:

Київський національний університет імені Тараса Шевченка,
03187, Київ-187, Проспект Академіка Глушкова, 2, к. 6.
Тел.: (044) 259 0511.
Факс: (044) 259 7044.
E-mail: aprowata@unicyb.kiev.ua,
aprovata@gmail.com