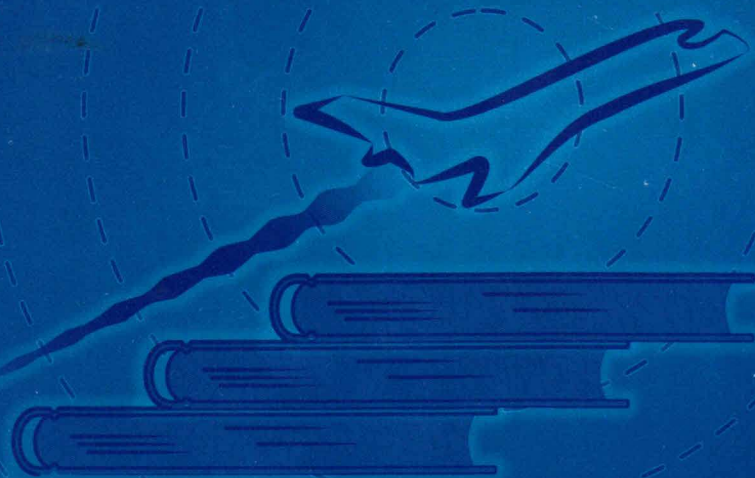


МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ АВІАЦІЙНИЙ УНІВЕРСИТЕТ



Наука і молодь

Збірник наукових праць

прикладна серія
випуск 5

Київ 2005 рік

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Національний авіаційний університет

НАУКА І МОЛОДЬ

ПРИКЛАДНА СЕРІЯ

ЗБІРНИК НАУКОВИХ ПРАЦЬ

Випуск 5

Київ 2005

АНАЛІЗ ІЄРАРХІЧНО СТРУКТУРОВАНИХ ІНФОРМАЦІЙНИХ МАСИВІВ В УМОВАХ НЕВИЗНАЧЕНОСТІ

В статті розглядається критичний вплив невизначеності, пов'язаної з неповністю вихідних даних на функцію належності структурованого об'єкта масиву порівняння. Показана залежність функції належності від кількості невизначених параметрів.

Постановка проблеми. Сучасний розвиток науки та галузей народного господарства пов'язаний з необхідністю прийняття рішень з урахуванням все більшої кількості чинників, що на них впливають.

Одним із шляхів прийняття ефективних рішень є вибір з масиву інформації або бази знань про існуючі плани рішень чи моделей об'єктів тих, що найбільш відповідають наявному об'єкту або ситуації, що склалася. Проте досить часто, особливо при аналізі динамічних систем або даних, отриманих емпіричним шляхом, вихідна інформація для прийняття рішень може бути неповною, фрагментованою або опосередкованою, причому строки прийняття рішення не дозволяють встановити невизначені параметри більш точно або це встановлення взагалі не уявляється можливим.

Разом з тим, прийняття зважених рішень або вибір операцій підтримки цілеспрямованих дій на основі отриманих результатів обробки даних повинно враховувати усі особливості та типи даних, що містяться в умовах поставлених завдань. Інакше кажучи, при прийнятті рішень в умовах недостатньої вихідної інформації потрібно вміти оцінити критичний вплив, який може спричинити невизначеність на область найбільш прийнятних рішень, визначених на основі обробки вихідної інформації.

Аналіз останніх досліджень. Прийняття рішень в умовах невизначеності в широкому розумінні цього поняття має розгалужене прикладне підґрунтя [4,7]. Останнім часом широке застосування знаходить метод обробки інформаційних масивів шляхом створення інтелектуальних систем, що використовують спеціалізовані алгоритми на основі м'яких обчислень і дозволяють формалізувати інформацію або знання про об'єкт дослідження та невизначеність, що міститься у його описанні. Як математичний інструмент інтелектуальні системи використовують теоретико-множинний підхід, теорію нечітких множин, математичну логіку та інші розділи дискретної математики [6].

Задачі, пов'язані з розпізнаванням та класифікацією структурованих об'єктів при невизначеності, пов'язаної з розпливчастими умовами, описанні об'єкта вектором належностей або полем імовірностей широко представлені і розглянуті: [2,3,5] та ін. Наближений підхід до визначення впливу невизначеності, пов'язаної з відсутністю даних, окреслений у роботі [1].

Постановка завдання. Для наочності розглядатимемо множини даних, що містять певні характерні ознаки яких-небудь явищ або об'єктів, тобто їх *прояви*. Прояви у множинах даних розподілені по визначеним групам та підгрупам, які підпорядковані ієрархічній структурі, і утворюють *класифікатор* даних.

Нехай ми маємо деяку множину даних, в якій містяться дані що описують певне явище або об'єкт дослідження A_K , і розподілені згідно обраного класифікатору по групам розподілу: $A_K \{a_1; a_2; a_3; \dots; a_X\}$. X - кількість груп розподілу. Число проявів у кожній з груп розподілу довільне і скінчене:

$$a_1 \{a_{11}; a_{12}; a_{13} \dots a_{1\alpha_1}\}, a_2 \{a_{21}; a_{22}; a_{23} \dots a_{2\alpha_2}\}, a_3 \{a_{31}; a_{32}; a_{33} \dots a_{3\alpha_3}\}, \dots, a_X \{a_{X1}; a_{X2}; a_{X3} \dots a_{X\alpha_X}\}$$

При матричному представленні, кожен з елементів виду a_{ij} являє собою одномісну компонент підмножину, що може містити, або не містити прояв явища $A_K : A_K \{a_{ij}\}$.

Розглянемо тепер іншу множину даних, що також описує певне явище або об'єкт G_N , причому всі його прояви є чітко визначені на основі багаторазових прямих спостережень, вимірювань, експериментів та інших достовірних даних, тобто є детермінованими. Прояви в G_N розподілені довільним чином, але відносно тих самих груп розподілу: $G_N \{a_1; a_2; a_3; \dots; a_X\}$, $G_N \{a_{ij}\}$. При

дослідженні множина явища або об'єкта $A_K \{a_{ij}\}$ підлягає класифікації відносно множин $G_N \{a_{ij}\}$. Інакше кажучи, кожна з множин $G_N \{a_{ij}\}$ є гіпотезою-множиною при розпізнаванні або ототожненні явища-множини $A_K \{a_{ij}\}$. Об'єднання сукупності множин $G = \bigcup_{N=1}^L G_N$ де L – кількість гіпотез, утворює основний масив порівняння.

Уведемо невизначеність. Припустимо, що деякі з груп розподілу $\{a_1; a_2; a_3; \dots; a_X\}$ множин A_K містять невідоме число проявів, причому число одномісних компонент-множин у кожній з цих груп та саме число груп є відомим. Відповідно, найменшою структурною одиницею невизначеності для будь-якої множини буде група невизначеності a_{Ui} .

$A_{KU} = \bigcup_{i=1}^{\theta_K} a_{Ui}$ – підмножина, що містить усі групи невизначених проявів для даного явища.

θ_K – кількість невизначених груп у A_K .

Задача полягає у тому, щоб порівняти послідовно множину проявів явища або об'єкта A_K з кожною із множин основного масиву порівняння $G = \{G_1; G_2; G_3; \dots; G_L\}$ і обрати з цього масиву множини, що найбільше відповідають A_K , урахувавши при цьому вплив невизначеності.

Виклад основного матеріалу дослідження. У даній роботі розглядатимемо випадок присутності невизначеності, коли не відомі ані імовірнісні характеристики невідомих проявів, ані вектори належностей до кожної з гіпотез основного масиву. Вагові характеристики для кожного типу проявів у групах вважаються однаковими.

Функцію належності [5] множини окремого, на початковому етапі досліджень не ототожненого явища або об'єкту $A_K \{a_{ij}\}$ множині гіпотези $G_N \{a_{ij}\}$, що характеризує її пріоритетність при застосуванні для ототожнення $A_K \{a_{ij}\}$ названо застосовністю даної гіпотези [1]:

$$\mu_{G_N}(A_K) = P_N = \begin{cases} (A'_K \triangleleft G'_N) \rightarrow (P_N = S'_N / G'_N), \\ (A'_K \triangleright G'_N) \rightarrow (P_N = S'_N / A'_K), \\ (A'_K = G'_N) \rightarrow \{(P_N = S'_N / G'_N) \vee (P_N = S'_N / A'_K)\}; \end{cases} \quad (1)$$

$$P_N = \begin{cases} (0,1], A_K \cap G_N = \overline{\emptyset}, \\ 0, A_K \cap G_N = \emptyset. \end{cases}$$

Раніше, з умов мінімізації застосовності, було уведено спрощене урахування невизначеності шляхом заміни груп розподілу на одномісні підмножини, в припущенні, що вони містять хоча б по одному прояву, відмінному, проте від проявів основного масиву [1]. При такому урахуванні невизначеності застосовність штучно знижується при виконанні умови

$$(\theta_K + A'_{KU}) > G'_N \rightarrow (P_{Na.u.} = S'_N / (\theta_K + A'_{KU})).$$

Таке урахування невизначеності припустиме при первинному аналізі даних, але не враховує максимальне значення, яке може приймати застосовність під впливом невизначеності та залежить від превалюючої кількості одномісних множин проявів, що містить кожна з груп невизначеності, тобто не застосовне для довільних масивів даних.

Повне урахування невизначеності будемо реалізовувати через пошук максимальних та мінімальних значень, які може приймати застосовність при всіх можливих випадках розкриття невідомих проявів для фіксованої кількості груп невизначеності.

1. Знайдемо максимальну застосовність, що може виникати при урахуванні невизначеності. Для цього розкладемо: $A_K = A_{KU} \cup A_{KU}$; $G_N = G_{NU} \cup G_{NU}$;

G_{NU} – підмножина гіпотези G_N , відповідна за групами класифікаційного розподілу підмножині явища A_{KU} . Сума проявів по групах для множин A'_K та G'_N , визначається як:

$G'_N = \sum_{n=1}^X g_n$; $A'_K = \sum_{n=1}^X t_n$ де g_n і t_n є числа компонентів по групах розподілу у відповідних у множинах.

Чисельно можна записати: $A'_K = A'_{K\bar{U}} + A'_{KU}$; $G'_N = G'_{N\bar{U}} + G'_{NU}$.

Для визначення впливу невизначеності на максимальне значення застосовності уведемо підмножину m_U – варіант заповнення підмножини A_{KU} проявами, що містить найбільшу їхню кількість з усіх реально можливих комбінацій проявів в усіх класифікаційних групах, що містять невизначеність. m_U визначається через відповідні класифікаційні групи розподілу з максимальним заповненням виходячи з властивостей класифікаційних груп розподілу проявів та реальних чинників, що впливають на A_K та G_N :

$m_{U_i} = \max(G_{1U}\{a_{U_i}\}; G_{2U}\{a_{U_i}\}; \dots; G_{LU}\{a_{U_i}\})$; $m_U = \bigcup_{i=1}^{\theta_K} m_{U_i}$. Чисельно можна записати:

$m'_{U_i} = \max(G_{1U}\{g_{U_i}\}; G_{2U}\{g_{U_i}\}; \dots; G_{LU}\{g_{U_i}\})$; $m'_U = \sum_{i=1}^{\theta_K} m'_{U_i}$. Очевидно, що $m'_U \geq G'_{NU}$.

Множина спільних проявів для A_K та G_N визначається як перетин цих множин:

$$S_N\{a_1; a_2; a_3; \dots; a_X\} = G_N \cap A_K = G_N\{a_1; a_2; a_3; \dots; a_X\} \cap A_K\{a_1; a_2; a_3; \dots; a_X\}; S'_N = \sum_{n=1}^X s_n;$$

Відповідно, множину спільних проявів можна представити як $S_N = (A_{KU} \cup A_{K\bar{U}}) \cap (G_{NU} \cup G_{N\bar{U}})$. Максимальна застосовність досягатиметься за умови максимуму перетину, тобто коли $m_U = G_{NU}$:

$$S_{N\max} = (m_U \cup A_{K\bar{U}}) \cap (G_{NU} \cup G_{N\bar{U}}) = (G_{NU} \cup A_{K\bar{U}}) \cap (G_{NU} \cup G_{N\bar{U}});$$

застосувавши принцип дистрибутивності, отримаємо [6]:

$$S_{N\max} = (A_{K\bar{U}} \cap G_{N\bar{U}}) \cup G_{NU} = (A_{K\bar{U}} \cap G_{N\bar{U}}) \cup G_{NU} = S_{N\bar{U}} \cup G_{NU}.$$

Тоді з системи (1), максимальна застосовність визначатиметься як

$$\begin{cases} (A'_{K\bar{U}} > G'_{N\bar{U}}) \rightarrow P_{N\max} = (S'_{N\bar{U}} + G'_{NU}) / (A'_{K\bar{U}} + G'_{NU}), \\ (A'_{K\bar{U}} < G'_{N\bar{U}}) \rightarrow P_{N\max} = (S'_{N\bar{U}} + G'_{NU}) / (G'_{N\bar{U}} + G'_{NU}), \\ (A'_{K\bar{U}} = G'_{N\bar{U}}) \rightarrow P_{N\max} = \{(S'_{N\bar{U}} + G'_{NU}) / (A'_{K\bar{U}} + G'_{NU})\} \vee \{(S'_{N\bar{U}} + G'_{NU}) / (G'_{N\bar{U}} + G'_{NU})\}; \end{cases} \quad (2)$$

$P_{N\max}$ – максимальне значення застосовності P_N при даній невизначеності m_U .

З системи (2) видно, що максимальне значення застосовності з урахуванням невизначеності не залежить власне ні від кількості груп невизначеності, ні від кількості одномісних компонент множин, що містяться у цих групах, а лише від кількості проявів у відповідних їм групах множини гіпотези G_{NU} . Можна також зробити висновок, що $P_{N\max} = 1$ тоді і тільки тоді, коли $A_{K\bar{U}} = G_{N\bar{U}}$.

2. Визначимо тепер мінімальну застосовність з урахуванням невизначеності.

Мінімальному випадку буде відповідати заповнення підмножини невизначеності усіма проявами з максимального варіанту заповнення, за винятком проявів, що містяться у відповідній підмножині гіпотези G_{NU} . Сукупність таких проявів можна записати через різницю множин

$A_{KU} = m_U \setminus G_{NU}$. Тоді мінімальний перетин множин A_K та G_N буде:

$$S_{N\min} = (A_{KU} \cup A_{K\bar{U}}) \cap (G_{NU} \cup G_{N\bar{U}}) = \{(m_U \setminus G_{NU}) \cup A_{K\bar{U}}\} \cap (G_{NU} \cup G_{N\bar{U}});$$

$$S_{N\min} = \{(m_U \setminus G_{NU}) \cup A_{K\bar{U}}\} \cap G_N.$$

Підставивши в (1) чисельні значення проявів у відповідних підмножинах, отримаємо:

$$\begin{cases} (A'_{K\bar{U}} + (m'_U - G'_{N\bar{U}}) < (G'_{N\bar{U}} + G'_{NU}) \rightarrow (P_{N\min} = \frac{S'_{N\bar{U}}}{G'_{N\bar{U}} + G'_{NU}}), \\ (A'_{K\bar{U}} + (m'_U - G'_{N\bar{U}}) > (G'_{N\bar{U}} + G'_{NU}) \rightarrow (P_{N\min} = \frac{S'_{N\bar{U}}}{(A'_{K\bar{U}} + (m'_U - G'_{N\bar{U}}))}); \end{cases} \quad (3)$$

Критичний вплив невизначеності на застосовність окремої детермінованої гіпотези до вибірки множини явища можна проілюструвати графіком залежності $P = f(m'_U)$ (рис. 1). Крива $P_{Nau.}$ показує наближене урахування невизначеності шляхом мінімізації застосовності. Відрізок $(0; m'_U = G'_{NU})$ по осі m'_U охоплює значення, які не може приймати невизначеність, що випливає з умови $m'_U \geq G'_{NU}$.

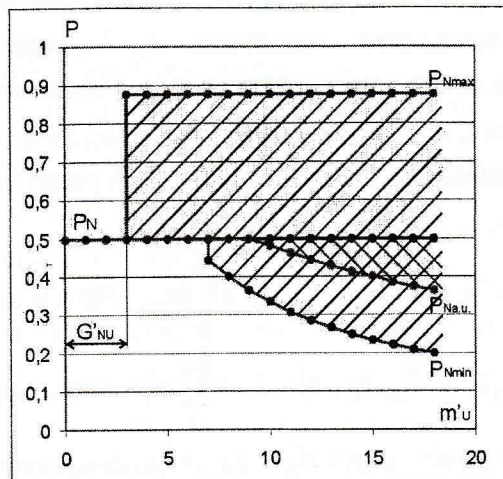


Рис. 1. Залежність функції належності від кількості проявів у підмножині невизначеності

Висновки. Таким чином, область можливих дискретних значень P_N при зростанні числа невизначених проявів являє собою півплощину, що виходить з точки $(P_N; m'_U = G'_{NU})$ і окреслена кривими P_{Nmax} та P_{Nmin} причому P_{Nmax} приймає фіксоване значення, а P_{Nmin} по мірі збільшення числа можливих проявів в підмножині невизначеності асимптотично наближається до 0.

Експериментально підтверджено, що всі інші можливі дискретні значення застосовності будуть знаходитися тільки у межах визначеної області, а значить, справляти передбачуваний вплив на результати прийняття рішень на основі неповних вихідних даних. Слід також зазначити, що встановлені залежності (1), (2), та (3) дійсні для даних як якісного, так і для кількісного характеру.

На основі розробленої методики створено пробну модель експертної системи, що дозволяє класифікувати невідомі літальні апарати та атмосферні явища по характерним проявам. Перші результати обробки тематичних інформаційних повідомлень за допомогою створеної системи показали високу чутливість до вхідних даних та значне скорочення часових та трудових витрат в порівнянні з існуючими схемами обробки повідомлень.

Перспективним є вирішення задачі урахування невизначеності, пов'язаної з неповністю інформації із залученням систем уведення вагових коефіцієнтів ознак для визначення сукупності моделей рішень або об'єктів, більш адекватних критеріям пріоритетності.

Список літератури

1. Білик А.С., Порівняння масивів якісних даних на прикладі не ототожнених явищ // 36. наук. праць IV Міжн. наук. конф. „Політ”, – К.: НАУ, 2004.
2. Вятчин Д.А., Об индексации нечеткой иерархии методами нечеткой классификации на основе понятия распределения // Мат. междунар. науч. сем. „Интеллектуальный анализ информации”, – К.: НТУУ «КПИ», Просвіта, 2004. – С. 69-77.
3. Обработка информации и принятие решений в условиях неопределенности: Сб. науч. ст. // АН КиргССР, Ин-т автоматизации, – Фрунзе: Илим, 1980.
4. Ходаков В.Е., Граб М.В., Классификация ситуаций в задаче планирования управления лесным пожаром // 36. наук. статей конф. „Проблеми моделювання та прогнозування надзвичайних ситуацій”, – К.: КНУБіА, Чорнобильінтерінформ, 2002. – С. 3-5.
5. Шлепаков Л.Н., Системы с базами данных по решению задач распознавания и классификации информационных сообщений // Интеллектуализация систем обработки информ. сообщений: Сб. науч. тр., – К.: НАНУ, Ин-т матем., 1995. – С. 11-38.
6. Шаранов О.Д., Дискретный анализ: навч.-метод. посібник, – К.: КНЕУ, 2002.
7. Принятие решений в условиях неопределенности: Межвуз. науч. сб., – Уфа: Уфимский гос. авиац. техн. ун-т, 2000.

У збірнику вміщено статті, в яких викладено результати наукових, експериментальних і теоретичних досліджень молодих учених. Присвячений актуальним проблемам сучасності в галузі науки і техніки. Розрахований на науковців та інженерно-технічних працівників.

ГОЛОВНИЙ РЕДАКТОР

Бабак В.П., ректор, доктор технічних наук, професор,
член-кореспондент Національної академії наук України,
заслужений діяч науки і техніки України,
лауреат Державної премії України в галузі науки і техніки

ЗАСТУПНИК ГОЛОВНОГО РЕДАКТОРА

Харченко В.П., проректор з наукової роботи, доктор технічних наук,
професор, лауреат Державної премії України в галузі науки і техніки

ВІДПОВІДАЛЬНИЙ СЕКРЕТАР

Сябрюк Є.М., голова студентського науково-технічного товариства

РЕДАКЦІЙНА КОЛЕГІЯ

Технічні науки

Дмитрієв С.О., доктор технічних наук, професор
Іванов С.В., доктор хімічних наук, професор
Конахович Г.Ф., доктор технічних наук, професор
Азарсков В.М., доктор технічних наук, професор
Сидоров М.О., доктор технічних наук, професор
Кулик М.С., доктор технічних наук, професор
Прокопенко І.Г., доктор технічних наук, професор
Запорожець О.І., доктор технічних наук, професор
Литвиненко О.Є., доктор технічних наук, професор

Фізико-математичні науки

Азнакаєв Е.Г., доктор фізико-математичних наук, професор
Байса Д.Ф., доктор фізико-математичних наук, професор
Денисюк В.П., доктор фізико-математичних наук, професор
Куліш В.В., доктор фізико-математичних наук, професор
Поліщук А.П., доктор фізико-математичних наук, професор

Друкується за постановою вченої ради Національного авіаційного університету (протокол № 10 від 21.12.2005).

Зареєстровано Державним комітетом телебачення і радіомовлення України.

Свідectво про державну реєстрацію друкованого засобу масової інформації КВ № 7799 від 28.08.2003.

Адреса редакційної колегії: 03058, Київ-58, проспект Космонавта Комарова 1, тел. 408-53-33.

© Національний авіаційний університет, 2005

Наукове видання

НАУКА І МОЛОДЬ
ПРИКЛАДНА СЕРІЯ
ЗБІРНИК НАУКОВИХ ПРАЦЬ

Випуск 5

В авторській редакції

Комп'ютерна верстка: В.В. Мішкур, Н.О. Тонкаль

Дизайн: І.В. Муштіна, О.В. Шитіков

Підп. до друку 12.07.06. Формат 60х84/8. Папір офс.
Офс. друк. Ум. друк арк. 18,13. Обл.-вид. арк. 19,5.
Тираж 100 пр. Замовлення № 153-1. Вид. № 12/II.

Видавництво НАУ.

03058. Київ-58, проспект Космонавта Комарова, 1.

Свідоцтво про внесення до Державного реєстру ДК № 977 від 05.07.2002.