

Національна академія наук України
Інститут кібернетики імені В. М. Глушкова

На правах рукопису

БРАТУСЬ Олександр Васильович

УДК 681.3.16

СИСТЕМА ІНДУКТИВНОГО ВИВОДУ
НА ОСНОВІ МОДИФІКОВАНОГО
GUHA-МЕТОДУ

01.05.03 — математичне та програмне забезпечення
обчислювальних машин і систем

Автореферат дисертації на здобуття наукового ступеня
кандидата фізико-математичних наук

Київ — 1996



004.4
Дисертацією є рукопис.

Робота виконана в Інституті кібернетики ім. В. М. Глушкова НАН України та Науково-учбовому центрі прикладної інформатики.

Наукові керівники: доктор технічних наук **СТУКАЛО О. С.**,
доктор фізико-математичних наук,
ГУПАЛ А. М.

Офіційні опоненти: доктор фізико-математичних наук
КНОПОВ П. С.,
кандидат фізико-математичних
наук **РЕЗНИЧЕНКО В. А.**

Провідна організація: Київський університет
ім. Т. Шевченка.

Захист відбудеться «20» грудня 1996 р. о 11
год. на засіданні спеціалізованої вченої ради Д 01.39.02
при Інституті кібернетики ім. В. М. Глушкова НАН України
за адресою:

252022 Київ 22, проспект Академіка Глушкова, 40.

З дисертацією можна ознайомитися в науково-технічному
архіві інституту.

Автореферат розісланий «———» ————— 199 р.

учений секретар
спеціалізованої вченої ради

СИНЯВСЬКИЙ В. Ф.

Загальна характеристика роботи

Актуальність теми. Досвід створення та застосування інтелектуальних прикладних систем виявив особливу роль знань про предметну область. Традиційно придбання знань здійснюється за схемою "експерти-> інженери за знаннями-> база знань", яка складає до 90% вартості інтелектуальної системи. Автоматизація процесу придбання знань є важливою проблемою.

Мета роботи. Розробити процедуру індуктивного виводу для ефективного здобуття найбільш цінних знань із емпіричних даних будь-якого походження. Розробити процедуру для перевірки адекватності здобутих знань. На цій основі розробити систему для розв'язання широкого класу задач. Система повинна мати гарний показник "вартість - ефективність".

Загальна методика. В дисертації здійснена модифікація ЄУНА-методу. При дослідженні використовувався апарат математичної логіки, математичної статистики, обчислювальності функцій.

Наукова новизна. Розроблені критерій цінності знань, який відповідає принципу "все важливе", індуктивна процедура для отримання всіх "важливих" знань. Доведено ряд тверджень для ефективного отримання таких знань. Розроблена процедура абдуктивного виводу, наведені статистичні оцінки результатів абдуктивного виводу. Запропонована процедура автоматичного налагодження на клас задач з максимальною достовірністю виводу, що є можливою при цьому підході.

Практична цінність та впровадження. Розроблена система LOGIS. Практичне застосування системи здійснено у межах таких договорів:

1. "Створення системи імітаційного моделювання селекційного процесу комплексно-стійких сортів озимої пшениці". Замовник - Інститут захисту рослин Української академії аграрних наук (УАН), Миронівський інститут пшениці УАН.
2. "Розробка системи прогнозування вторинної структури білків". Замовник - Інститут молекулярної біології Національної академії наук України.
3. "Розробка імітаційної моделі функціонування системи коадаптації, визначення умов, які характеризують стан рівно-

В. Стефанік
АН України

ваги в агробіоценозах". Замовник - Інститут захисту рослин.

4. "Розробка системи прогнозування стійкості с/г культур за спектром гліадіна". Замовник - Інститут захисту рослин.

5. "Розробка принципів та методів планування створення екологічно-пластичних сортів зернових та круп'яних культур з використанням логічного імітаційного моделювання процесу селекції на ПЕСМ". Замовник - ДКНТ України.

6. "Розробка методів, алгоритмів та програм автоматизованого формування бази знань за клінікою і патогенезом ВІЛ". Замовник - Інститут кібернетики ім. В.М.Глушкова НАН України.

Публікації та апробації робіт. За темою дисертації опубліковано 16 робіт. Загальні положення доповідались та обговорювались на засіданні Координаційного комітету з обчислювальної техніки (ККОТ) по експертних системах у 1985 р.

(м.Баку) - кер. Г.С.Поспелов; на засіданні ККОТ по експертних системах у 1986 р. (м.Ташкент) - кер. Г.С.Поспелов; на республіканському семінарі "Математичне забезпечення автоматизованих систем обробки даних та пакетів програм" Наукової ради АН УРСР з проблеми "Кібернетика" в 1986 р. (м.Київ) - кер. І.В. Сергієнко.

Структура та обсяг роботи. Дисертація складається із вступу, трьох глав, висновку та списку літератури, який містить 51 найменування.

Зміст роботи

У вступі обґрунтовується актуальність теми дисертації, формулюється ціль дослідження та наводиться короткий зміст дисертації за главами.

Досвід побудування та використання інтелектуальних систем виявив особливу роль знань. Здобуття знань є трудомістким процесом, який дорого коштує. Треба створити процедуру автоматичного здобуття знань із емпіричних даних. Процедура повинна мати показник "вартість - ефективність", який задовольняє користувача.

У першій главі аналізуються вимоги, яким повинна відповідати процедура автоматичного здобуття знань. Виділяються такі основні вимоги:

автоматичне породження всіх знань в деякій мові;
статистична оцінка кожного знання;

визначення "суттєвого" знання;

відбір усіх "суттєвих" знань;

інтерпретація усіх "суттєвих" знань.

Таким чином, принцип автоматичного придбання знань формулюється як "все суттєве". Далі аналізуються методи автоматичного придбання знань, які існують. Зазначається, що один із методів, відомий як GUNA-метод, найбільш повно відповідає наведеним вимогам. Наводиться коротке викладення GUNA-методу.

У другій главі доводиться, що рішення в розумінні GUNA-методу не відповідає принципу "все суттєве" в частці "все". Вводиться поняття "критичний базис". Доводиться, що результат роботи індуктивної процедури у вигляді критичного базису відповідає принципу "все суттєве". Доводяться ствердження по оптимізації побудови критичного базису. Обумовлюється процедура інтерпретації критичного базису.

Прийнято обчислення в квантором асоціації Фішера, тобто кванторні формули інтерпретуються за допомогою точного критерію перевірки на незалежність двох ознак (критерій Фішера).

Моделі обчислення:

$M = \langle Mod, f_1, \dots, f_n \rangle$, де

Mod - несуча множина, кінцева і непуста;

$f_i : Mod \rightarrow V_i$ ($i=1, n$) - унарні відображення моделі;

$V_i = \{1, 2, \dots, k_i, Un\}$ - множина значень;

Un - невизначене значення.

Мова обчислення складається з таких символів:

F_1, \dots, F_n - функціональні символи;

x - виділена змінна;

$\&, (K)$ - логічні операції;

$\sim\alpha$ - квантор Фішера.

Літерал обчислення має вигляд

$(K)F_1(x)$, де K є непуста підмножина V_1 .

Інтерпретація літерала: для довільного об'єкта o в Mod

$|(K)F_1(x)|[o] = 1$, якщо $f_1(o)$ належить K ,

$|(K)F_1(x)|[o] = Un$, якщо $f_1(o) = Un$,

$|(K)F_1(x)|[o] = 0$, якщо $f_1(o)$ не належить K .

Відкриті виділені формули мають вигляд

$P(x) = \&(K_i)F_1(x)$, I не містить однакових елементів.

1

Інтерпретація формули: для довільного об'єкта o в Mod

- $|P(x)|\{o\} = 0$, якщо хоча б один літерал $P(x)$ неістинний на об'єкті o ,
- $|P(x)|\{o\} = U_n$, якщо хоча б один літерал $P(x)$ невизначений на об'єкті o та нема неістинних літералів,
- $|P(x)|\{o\} = 1$, якщо всі літерали $P(x)$ істинні на об'єкті o .

В подальшому визначення змінної x та моделі M будемо опускати.

Визначення підформули. Нехай $P_1 = \&(K_i)F_i$, $P_2 = \&(N_j)F_j$, тоді

P_1 є підформулою P_2 , якщо I - підмножина J , K_i - підмножина N_j для довільного i з I . Кванторна формула має вигляд $(\sim\alpha)(P_1, P_2)$ або $P_1 \sim \alpha P_2$, де P_1 і P_2 - відкриті виділені формули обчислення. Інтерпретація кванторної формули:

$|P_1 \sim \alpha P_2| = 1$ тоді і тільки тоді, коли

$$\text{Fish}(P_1, P_2) = \sum_{i=s}^{\min(r, k)} q(i, r, k, m) < \alpha, \text{ де}$$

$0 < \alpha \leq 0.5$ - рівень прийняття гіпотези;

$s = \text{card}\{o : |P_1 \& P_2|\{o\} = 1\}$;

$r = \text{card}\{o : |P_1|\{o\} = 1\}$;

$k = \text{card}\{o : |P_2|\{o\} = 1\}$;

$m = \text{card}(\text{Mod})$ - загальна кількість об'єктів Mod ;

$q(i, r, k, m) = r!k!(m-r)!(m-k)!/m!i!(r-i)!(k-i)!(m+i-r-k)!$.

$|P_1 \sim \alpha P_2| = 0$, якщо $\text{Fish}(P_1, P_2) > \alpha$.

Значення $\text{Fish}(P_1, P_2) = \alpha$ називається критичним рівнем формули (гіпотези) $P_1 \sim \alpha P_2$.

GUHA-метод для емпіричної моделі M , формули P_1 та значення α з інтервалу $(0, 0.5]$ видає рішення у вигляді базису та доповнення.

Визначення базисної формули. Для даних (M, P_1, α) формула P базисна, якщо

$$|P_1 \sim \alpha P| = 1 \quad (\text{Fish}(P_1, P) < \alpha),$$

$$|P_1 \sim \alpha Q| = 0 \quad (\text{Fish}(P_1, Q) > \alpha),$$

де Q - будь-яка непуста підформула P .

Виділення базису повинно відповідати виділенню важливих, суттєвих знань з дотриманням принципу "все суттєве". Покажемо, що базис у визначенні GUHA-методу не відповідає даному

принципу в частині " все ".

Нехай для (M, P_1, α) P - базисна формула, тоді

а) $Fish(P_1, P) < \alpha$

б) $Fish(P_1, Q) > \alpha$, де Q - будь-яка непуста підформула P .

Виберемо Q^* - підформулу P , таку, що $Fish(P_1, Q^*) = \alpha^* < 0.5$.

Тоді для (M, P_1, α^*) формула P не буде базисною, оскільки пункт б) для Q^* не буде виконаний, а саме вираз

$Fish(P_1, Q^*) = \alpha^* > \alpha$ невірний.

Визначимо, що згідно з б) $\alpha^* > \alpha$. Таким чином, базис для $(M, P_1, \alpha^* > \alpha)$ не містить формулу P , яка була базисною для більш суворого критерію $\alpha < \alpha^*$. З послабленням критерію формула P стала " несуттєвою " гіпотезою, хоча факт послаблення критерію ніяк не повинен вести до втрати гіпотез, які були " суттєвими " при більш суворому критерії. Базис повинен містити " всі суттєві " гіпотези незалежно від завдання значення критерію. Критерій повинен здійснювати лише функцію граничного довірчого рівня за оцінкою гіпотез. Таким чином, базис у визначенні ЄУНА-методу не відповідає основному принципу індукції в частині " все ". Модифікуємо поняття базису.

Визначення критичної базисної формули.

Для даних (M, P_1, α) формула P є критична базисна формула, якщо

$Fish(P_1, P) = \alpha^* < \alpha$,

$Fish(P_1, Q) > \alpha^*$,

де Q - будь-яка непуста підформула формули P , α^* - критичний рівень формули $P_1 \sim \alpha P$. Зрозуміло, що критична базисна формула є базисною у визначенні ЄУНА-методу, якщо замість α взяти α^* .

Визначення критичного базису. Для даних (M, P_1, α) множина критичних базисних формул називається критичним базисом та позначається $\{P_i\}$.

Твердження 1. Нехай для (M, P_1, α) формула P є критичною базисною формулою. Тоді P - критична базисна формула для (M, P_1, α^*) , де α^* належить інтервалу $[\alpha, 0.5]$.

Доведення.

а) $Fish(P_1, P) = \alpha^* < \alpha^*$, оскільки $\alpha^* < \alpha$, $\alpha < \alpha^*$,

б) $Fish(P_1, Q) > \alpha^*$ (за визначенням Q - підформула P).

Твердження показує, що критичний базис задовольняє основному принципу індукції в частині " все ". Для повної відповідності принципу " неістинне " знання (множина доповнення) виключається з рішення. Дійсно, базис, згідно з теорією ЄУНА-

методу, визначає всю множину рішень. Таким чином, інтерпретація рішення обов'язково починається в інтерпретації базису, та лише потім залучаються знання з множини доповнення. Постановка задачі абдуктивного виводу дозволяє обмежитись знаннями критичного базису.

Таким чином, рішення у вигляді критичного базису відповідає основним вимогам задачі здобуття знань - "все суттєве".

Визначення критичного базису в порівнянні з ГУНА-методом вимагає більше часу на його побудову. Проблема часової оптимізації стає суттєвою. Наступні твердження дозволяють вирішити цю проблему.

Твердження про хибність сукупності формул.

Нехай M - модель, α - значення критерію, $P1$ і P - формули.

Якщо для формули $P1 \sim \alpha P$ в моделі M виконано

$$q(\min(r, k), r, k, m) > \alpha, \text{ тоді для будь-якої формули } Q,$$

що містить формулу P , формула $P1 \sim \alpha Q$ хибна в моделі M .

Виятком є $\{ k > r, q(r, r, r, m) < \alpha \}$.

Лема. Функція $q(\min(r, k), r, k, m)$ при $k > r$ монотонно зростає по k , а при $k < r$ - монотонно знижується по k .

Висновок з леми:

$$\min_k q(\min(r, k), r, k, m) - q(r, r, r, m).$$

Доведення твердження. Нехай формула $P1 \sim \alpha P$ характеризується в моделі M числами r, k . При створенні відкритих формул вигляду Q число r не змінюється, а число k може лише зменшитися. Згідно з умовами, $q(\min(r, k), r, k, m) > \alpha$. Це означає, що $\text{Fish}(P1, P) > \alpha$, тобто формула $P1 \sim \alpha P$ хибна в моделі M . Нехай $k < r$, тоді при створенні формул вигляду Q значення $q(\min(r, k), r, k, m)$ для формул $P1 \sim \alpha Q$, згідно з лемою, буде зростати, тобто всі формули вигляду $P1 \sim \alpha Q$ хибні в моделі M . Нехай $k > r$, тоді при створенні формул вигляду Q значення $q(\min(r, k), r, k, m)$ для формул $P1 \sim \alpha Q$, згідно з лемою, буде зменшуватися. Згідно з висновком леми це значення не може бути меншим, ніж значення $q(r, r, r, m)$, тобто, якщо $q(r, r, r, m) > \alpha$, всі формули у вигляді $P1 \sim \alpha Q$ хибні в моделі M . Якщо $q(r, r, r, m) < \alpha$, ніяких висновків не робиться, тобто умова $\{ k > r, q(r, r, r, m) < \alpha \}$ є виключенням. Твердження доведено.

Твердження про необхідні умови для критичних формул.

Для (M, P_1, α) формула P є критичною базисною формулою, якщо непушта множина

$\{ \alpha : |P_1|[\alpha] - 0, |P|[\alpha] - 0, |Q|[\alpha] - 1 \text{ або } \cup n \}$, де

Q - будь-яка непушта підформула P , що відрізняється від P одним літералом.

Доведення. За визначенням, P є критичною базисною формулою, якщо $\text{Fish}(P_1, Q) > \alpha^*$, де Q - будь-яка непушта підформула P , α^* - критичний рівень формули $P_1 \sim \alpha P$.

Якщо вилучити від P один літерал, значення Fish зміниться тоді і тільки тоді, коли непушта множина

$\{ \alpha : |P|[\alpha] - 0, |Q|[\alpha] - 1 \}$, або непушта множина

$\{ \alpha : |P|[\alpha] - \cup n, |Q|[\alpha] - 1 \}$, де $Q \in P$ без літералу.

Значення Fish збільшується, якщо при вилученні літералу від P кількість об'єктів, де гіпотеза $P_1 \sim \alpha Q$ спростовується, зростає за рахунок об'єктів, де гіпотеза $P_1 \sim \alpha P$ підтверджується. Така множина є тільки

$\{ \alpha : |P_1|[\alpha] - 0, |P|[\alpha] - 0, |Q|[\alpha] - 1 \text{ або } \cup n \}$.

Твердження доведено.

Обчислювальна складність вихідного GUNA-методу є поліноміальною.

Автор запрограмував вихідний та модифікований методи. Порівняльні випробування програм на даних, які представляють складні біологічні системи, дали ефективність оптимізації за часом не нижче двох порядків.

Процедура дедуктивного виводу

Нехай для емпіричної моделі M задана формула $P_1 = \&L_1$, яка не виконується на жодному об'єкті M (P_1 описує унікальну ситуацію), та задано рівень α . Для отримання знань про унікальну ситуацію вона розбивається на фрагменти, про які можливо здобути знання за допомогою процедури індуктивного виводу. В даному випадку окремий фрагмент описується одним літералом з P_1 , оскільки окремий літерал виконується на більшій кількості об'єктів моделі M , ніж кон'юнкція з двох та більше літералів, і за таким фрагментом можна здобути найбільш повні знання. Таким чином, на вхід індуктивної процедури подаються M, L_i , рівень α та визначається критичний базис $\{L_i\}$. Далі формується набір знань про всі фрагменти унікальної ситуації. Оскільки P_1 є кон'юнкція літералів, для отримання знань про

P1 потрібно виконати операцію кон'юнкції всіх $\{L_i\}$. Ця операція здійснюється за допомогою процедури дедуктивного виводу.

Дедуктивний вивід дає множину формул, отриманих внаслідок незалежного перехрестя всіх критичних базисів $\{L_i\}$.

Процедура абдуктивного виводу

Нехай предметна область представлена навчаючою вибіркою та задано набір значень деяких параметрів предметної області. Необхідно вивести набір значень решти параметрів. Формально процедура абдуктивного виводу описується таким чином. Для даної моделі M та відкритої виділеної формули $P1$ (літерали сформовані з підмножини заданих параметрів) потрібно вивести $P2$ (літерали сформовані з підмножини параметрів, які необхідно визначити).

Опишемо процедуру виводу $P2$ по заданих $P1$, M , α . Для даної емпіричної моделі M , формули $P1 = \&L_i$ (i пробігає непусту множину I) та рівня α процедура абдуктивного виводу виконує індукційний крок. Для кожного літералу L_i будується критичний базис $\{L_i\}$ ($L_i \sim \alpha\{L_i\}$), який містить знання про визначаючі параметри;

дедукційний крок. Застосовується дедуктивне правило виводу: якщо $L_i \sim \alpha\{L_i\}$, то $\&\{L_i\}$. Формальне правило абдуктивного виводу

$$\&L_i, L_i \sim \alpha\{L_i\}$$

$$\&\{L_i\}$$

Іншими словами, якщо задана формула $P1 = \&L_i$ та для кожного літералу L_i знайдено критичний базис $\{L_i\}$ на рівні α , то формула $P2$ належить $\&\{L_i\}$.

Достовірність процедури абдуктивного виводу

Процедура абдуктивного виводу будує множину $\&\{L_i\}$. Кожна формула з $\&\{L_i\}$ є непуста кон'юнкція формул, кожна з яких належить тільки одній множині $\{L_i\}$. Оцінивши ймовірність виводу кожної формули з $\&\{L_i\}$, можна обмежити множину формул, що виводяться.

Для оцінки ймовірності абдуктивного виводу звернемося за результатами з математичної статистики в частині визначення критерію одночасного виконання сукупності гіпотез. Нехай для перевірки k гіпотез є k критеріїв. Нехай відповідні статистики стохастично незалежні. Потрібно визначити критерій одночасного

виконання всіх k гіпотез. Один з підходів до об'єднання незалежних критеріїв ґрунтується на тому факті, що якщо i -й критерій має розмір α_1 , то ймовірність відхилення хоча б однієї з гіпотез за умови, що всі вони правдиві, дорівнює

$$1 - \prod_{i=1}^k (1 - \alpha_1). \quad (1)$$

Це співвідношення залишається наближено правдивим також для залежних критеріїв. Щодо абдуктивного виводу ці результати мають таке значення: нехай $L_1 \sim \alpha_1 Q_1$ - формули, істинні в моделі M , де α_1 є критичний рівень, тоді формула $\&Q_1$ оцінюється по співвідношенню (1). Для впорядкування формул, які виводяться по значенню оцінки, достатньо впорядкувати формули критичного базису по значенню критичного рівня.

Достовірність процедури абдуктивного виводу оцінюється таким чином. Нехай для даної емпіричної таблиці (модель M) визначено набір заданих (формула P_1) та набір визначаючих (формула P_2) параметрів, тобто визначено клас задач користувача. Підставою до прийняття результатів роботи абдуктивної процедури служить ступінь з'ясування кожного спостереження вихідної емпіричної таблиці. В нашому випадку процедурі виводу пропонуються всі спостереження (а саме частина, яка відповідає P_1) та вихідна емпірична таблиця (модель M) без даного спостереження. Користувач задає обмеження на кількість формул, які виводяться. Результат роботи процедури виводу порівнюється з усунутих спостереженням (з частиною, яка відповідає P_2) та підраховується число вірно виведених параметрів. У підсумках визначається достовірність виводу.

У третій главі дисертації описується система LOGIS. Отримані результати були використані при створенні програмної системи LOGIS. Аббревіатура LOGIS утворена з LOGIC та STATISTIC. Оскільки предметна область представляється даними, система створена в середовищі однієї з СУБД, а саме СУБД FOXPRO для IBM PC. Всі процедури, які реалізують індуктивний, дедуктивний, абдуктивний виводи, були переписані мовою прикладного програмування FOXPRO. LOGIS реалізована у вигляді систем меню та дозволяє користувачу ініціалізувати предметну область, редагувати інформацію про предметну область, вирішувати відповідні задачі та отримувати звіти.

Риси системи LOGIS :

1. Динамічне вилучення знань. Реалізація швидкої процедури індуктивного виводу дозволяє добувати знання динамічно під потреби конкретної задачі і вилучає проблему створення і ведення бази знань.
2. Ефективна актуалізація знань. Знання добуваються безпосередньо з даних. Логічна копія даних, на якій працює процедура індуктивного виводу, визначається самими даними і значеннями шкал та формується автоматично кожного разу при виборі предметної області. Таким чином, актуалізація даних та шкал автоматично веде до актуалізації знань.
3. Автоматичне настроювання на розв'язування класу задач.
4. Незалежність від експертів. Бази Даних розповсюджені в багатьох галузях людської діяльності. Після введення даних LOGIS автоматично налагоджується на розв'язування класу задач та виконує певні, обґрунтовані виводи, тобто має добрий показник відношення вартість/ефективність.

**Постановка та розв'язання практичних задач
Класифікація ситуацій**

Постановка задачі. Предметна область представлена таблицею емпіричних даних (модель M). Множина параметрів, які задані (формула P_2), вміщує всі параметри, крім виділеного (формула P_1). Треба визначити значення виділеного параметру (від 1 до k).

Розв'язання задачі. Спочатку визначається, належить задана ситуація першому класу чи ні.

Нехай для даних M і P_2 відомо значення P_1 . Задамо на вхід індуктивної процедури значення $\langle M, (1)P_2, 0.5 \rangle$ та $\langle M, (^1)P_2, 0.5 \rangle$. Отримаємо два критичних базиси $\{(1)P_2\}$ та $\{(^1)P_2\}$. Кожний базис є набір формул, кожна формула має свій критерій. Розмістимо критерії всіх формул обох базисів в порядку зростання від α_1 до α_N . Вибір α_i ($i=1, N$) визначає $\{(1)P_2\}$ та $\{(^1)P_2\}$, тобто визначає результат порівняння оцінок $\{(1)P_2\}$ та $\{(^1)P_2\}$. Оскільки значення P_1 відомо, кожне значення α_i ($i=1, N$) визначає дійсний або хибний результат класифікації. Об'єднання сусідніх значень α_i надає систему інтервалів значень α . Кожен інтервал зачинений зліва і відчинений справа. Вибір будь-якого α з одного інтервалу дає однаковий результат класифікації (дійсний або хибний). Діхотоміч-

ний підхід надає однакову цінність інтервалам дійсного та хибного результатів. Якщо знати, що α належить інтервалу хибної класифікації, тоді отриманий результат треба змінити на протилежний. Оптимальне значення α для класифікації "перший - не перший" визначається наступним чином. Усунемо із емпіричної таблиці перший рядок та здійснимо класифікацію цього рядка. Отримаємо систему інтервалів однакових результатів класифікації. Таким чином, отримаємо системи інтервалів для всіх рядків таблиці. Якщо вибрати будь-яке значення α , можна дізнатися, скільки рядків таблиці класифіковані вірно, а скільки невірно. Якщо перебрати всі значення лівих меж інтервалів, то знайдемо значення α , що надає максимальну достовірність класифікації "перший - не перший". Далі здійснюється класифікація "другий - не другий" . . . "к не к".

Приклад практичного застосування. Процедура використана в задачі прогнозування вторинної структури білків. Білок є послідовність будь-якої довжини, що складається з 20 літер (амінокислотні залишки). Визначити вторинну структуру білка - це означає класифікувати кожний амінокислотний залишок в залежності від сусідніх амінокислотних залишків. Є багато методів для розв'язання цієї задачі. Найкраща достовірність - 70.4%. Застосування запропонованої процедури надає достовірність 70.8%.

Задача абдуктивного виводу

Постановка задачі. Предметна область представлена таблицею емпіричних даних (Модель М). Множина заданих параметрів (формула P1) та множина визначаючих параметрів (формула P2) являють собою клас задач виводу P2 по P1.

Приклад практичного застосування. На основі запропонованої процедури створено АРМ селекціонера озимої пшениці. Навчаюча вибірка - результат схрещування материнської та батьківської форм озимої пшениці отримано в Миронівському інституті пшениці. В рамках АРМ вирішуються такі задачі:

за заданими значеннями материнської і батьківської форм вивести значення гібриду;

за заданими значеннями гібриду і однієї з батьківських форм вивести значення другої батьківської форми;

за заданими значеннями гібриду вивести значення батьківських форм.

Економічна ефективність застосування АРМ щодо виведення одного сорту озимої пшениці становить 171 тис. доларів. Рахунки виконані фахівцями Інституту захисту рослин.

Задача виводу покращених якостей

Постановка задачі. Нехай в N спостережень M параметрів предметної області та виділені цільові параметри. Треба визначити умови досягнення ситуації в комплексом поліпшених значень цільових параметрів.

Розв'язання задачі. Для цього необхідно знайти причини зміни значень цільових параметрів в залежності від змін значень інших параметрів. Перетворимо таблицю спостережень в таблицю з відповідною семантикою.

Розглянемо результат порівняння будь-яких двох спостережень предметної області (i -го та j -го). Домовимося, що спостереження з меншим номером має семантику "було", а спостереження з більшим номером має семантику "стало". Тоді результат порівняння є спостереження з семантикою "було \rightarrow стало". Порівнюючи кожне спостереження з усіма наступними (за номером спостереженнями), отримуємо повну таблицю порівнянь між спостережень предметної області. Таблиця має зміст: "Зміни нецільових параметрів визивають зміни цільових параметрів". Якщо параметр кількісної природи, результат порівняння є результатом віднімання від значення "стало" значення "було". Параметр шкалюється з виділенням шкал категорії "зменшилось" та категорії "збільшилось". Якщо параметр якісний, результат порівняння кодується " k_1-k_2 ", де k_1 і k_2 - номери шкал параметра в даному порівнянні. Параметр шкалюється з виділенням категорій або по частині "було", або по частині "стало", або "було-стало". Далі розв'язується задача абдуктивного виводу. Користувач задає вхідну ситуацію, вказавши шкали зміни цільових параметрів, тим самим визначає задачу "Які зміни нецільових параметрів треба утворити, щоб отримати завдання зміни цільових параметрів?". Процедура виводить обмежену кількість альтернатив зміни нецільових параметрів. Користувач вказує спостереження, до якого потрібно застосувати виведені зміни. Обчислюються абсолютні значення нецільових параметрів, які повинні призвести до комплексного покращення цільових параметрів.

Приклад практичного застосування. Ця процедура була використана для визначення складу гербіциду в високою ефективністю. Для трьох компонентів гербіциду (метафос, деїс, фосфамід) була надана таблиця експериментів. Налаштування процедури проведено з достовірністю 83%. Отримано 9 альтернатив, які надали 6 формул суміші. Польові випробування дали прогнозовану ефективність для двох формул.

Прогноз розвитку динамічної системи

Постановка задачі. Нехай є N спостережень M параметрів предметної області, зроблених з однаковим інтервалом часу. Треба спрогнозувати $N+1$ спостереження.

Розв'язання задачі. Перетворюємо вихідну таблицю спостережень в таблицю темпів, тобто від кожного i рядка віднімаємо $i-1$ рядок ($i = 1, N$). Кожний (i, j) елемент таблиці темпів має семантику: "Змінення значення j -го параметра при переході системи від $(i-1)$ -го спостереження до i -го спостереження". Таблиця темпів має $N-1$ рядків. Усунаємо від цієї таблиці перший рядок. Отримаємо таблицю в $N-2$ рядків. Приєднаємо до вихідної таблиці темпів в $N-1$ рядків таблицю темпів в $N-2$ рядків. Кожний i -ий рядок подвійної таблиці темпів має семантику " $\Delta_i S \rightarrow \Delta_{i+1} S$ ", де $\Delta_i S$ - змінення значень параметрів системи при переході від $(i-1)$ -го спостереження до i -го спостереження ($\Delta_{i+1} S$ - аналогічно). Останній рядок подвійної таблиці темпів містить половину елементів, які створюють вхід до процедури абдуктивного виводу разом з подвійною таблицею темпів. Отримаємо темпи змінення значень параметрів системи при переході системи від N -го спостереження до $(N+1)$ -го спостереження.

Приклад практичного застосування. В 1986 р. за замовленням Управління інформації Ради Міністрів СРСР була розроблена та впроваджена Діалогова система прогнозування розвитку динамічних систем.

У висновку дисертації наведені підсумки роботи.

Основні результати роботи.

Запропоновано та теоретично обгрунтовано поняття "критичний базис" для отримання найбільш важливих гіпотез.

Доведені твердження щодо ефективної побудови формул в критичного базису.

Теоретично обгрунтована процедура абдуктивного виводу, отримані статистичні оцінки результатів абдуктивного виводу.

Розроблена процедура виявлення оптимального значення параметра абдуктивного виводу, яке надає максимальну достовірність виводу.

Сформульовано клас задач користувача, створена програмна система LOGIS для розв'язування такого роду задач.

Наведені приклади застосування системи LOGIS в розв'язанні складних біологічних задач, зокрема в моделюванні процесу селекції озимої пшениці та прогнозуванні структури білків.

Основні положення дисертації опубліковані в таких працях:

1. Братусь А.В. Преобразование эмпирической информации в систему логических высказываний// Докл. АН УССР. Сер. А.-1987.-№5.-С.72 - 74.

2. Братусь А.В. Об одной модели пакетов прикладных программ сложной структуры// Машинные методы обработки информации для задач управления.-Киев:Ин-т кибернетики им.В.М.Глушкова АН УССР, 1986.-С.17 - 21.

3. Братусь А.В. Имитационное моделирование систем на основе логических выводов// Докл. АН УССР. Сер. А.-1987.-№10.-С.62 - 64.

4. Сергиенко И.В., Гупал А.М., Братусь А.В. LOGIS-система, реализующая статистический абдуктивный вывод на эмпирических данных// Кибернетика.-1995.-№ 3.-С.160 - 173.

5. Вагис А.Г., Братусь А.В., Василенко В.И. Реализация экспертной системы на основе логической имитации.-Киев, 1987.-22с.-(Препр./АН УССР, Ин-т кибернетики им. В.М.Глушкова,57).

6. Вагис А.Г., Братусь А.В., Василенко В.И. Экспертная система на основе образцов и логической имитации// Тез. докл. XII семинара "Прикладная информатика автоматизированных систем проектирования, управления, программированной эксплуатации".-Калининград, 1987.-С.123.

7. Lesovoi M., Smelyanets V., Aifaoui A., Vasilenko V, Bratus A. Logical simulation of processes at forming of winter wheat germplasm with complex resistance to insect pests and pathogens (P.1)// Arch. Phytopath. Pflanz.- 1994.-29.-P.153 - 164.

8. Lesovoi M., Smelyanets V., Aifaoui A., Vasilenko V, Bratus A. Logical simulation of processes at forming of

winter wheat germplasm with complex resistance to insect pests and pathogens (P.2)// Ibid.-P.263 - 283.

9. Chaschin N.A., Malchenko S.Z., Bratus A.V. A new method for the prediction of secondary protein structures with using of selfstudying expert system// Bioinformatics in the 90's Final programme and abstracts, November 20-22,1991.- Maastricht (Netherlands),1991.- P.13.

10. Chaschin N.A., Malchenko S.Z., Bratus A.V. The Design of Model Proteins with the predefined structure. Protein engineering and beyond// Miami Short Reports.-1993.- 3.-P.10.

Внесок автора дисертаційної роботи в спільних публікаціях:

4 - обґрунтовано поняття "критичний базис", розроблена процедура індуктивного виводу, отримані статистичні оцінки абдуктивного виводу;

5,6 - доведені твердження щодо ефективної побудови критичного базису;

7,8,9,10 - програмна обробка та інтерпретація даних, наданих замовником.

Братусь А.В. Система індуктивного виводу на основі модифікованого GUHA-метода. Дисертація на соискание ученой степени кандидата физико-математических наук по специальности 01.05.03 - математическое и программное обеспечение вычислительных машин и систем, Институт кибернетики им. В.М. Глушкова НАН Украины, Киев, 1996.

Предлагается процедура логического вывода, состоящая из процедур индуктивного и дедуктивного вывода. Приводятся статистические оценки результатов вывода. Описываются примеры практического применения процедуры.

The logical inference procedure, composed from inductive and deductive inference procedures is suggested. Probability evaluation of inference result is given. Practical examples are considered.

Ключові слова:

емпіричні дані, гіпотеза, індуктивний вивід, абдуктивний вивід, інтерпретація результатів, достовірність виводу.

Лид. до друку 02.10.06. Формат 60x84/16 ІапІр пич. № 1.
Оіс. друк. Ум. друк. арк. 0,93. Ум. фарбо-відб. 1,05.
Обл.-вид. арк. 1,0. Зам. 469. Тираж 100 пргм.
Редакційно-видавничий відділ з поліграфічною дільницею
Інституту кібернетики Імені В.М.Глушкова НАН України
252022 Київ 22, проспект Академіка Глушкова, 40

437439

AV 36.175

7