



УКРАЇНА

(19) UA (11) 28751 (13) U
(51) МПК (2006)
G06F 17/17
G06F 17/18
G06F 17/30

МІНІСТЕРСТВО ОСВІТИ
І НАУКИ УКРАЇНИ

ДЕРЖАВНИЙ ДЕПАРТАМЕНТ
ІНТЕЛЕКТУАЛЬНОЇ
ВЛАСНОСТІ

ОПИС ДО ПАТЕНТУ НА КОРИСНУ МОДЕЛЬ

видається під
відповідальність
власника
патенту

(54) ПРИСТРІЙ ДЛЯ ОБРОБКИ СЛАБОСТРУКТУРОВАНИХ ДАНИХ НА ОСНОВІ МЕРЕЖ БАЙЄСА

1

2

(21) u200707506

(22) 04.07.2007

(24) 25.12.2007

(72) ТЕРЕНТЬЄВ ОЛЕКСАНДР МИКОЛАЙОВИЧ,
UA, БІДЮК ПЕТРО ІВАНОВИЧ, UA, КОРШЕВНЮК
ЛЕВ ОЛЕКСАНДРОВИЧ, UA

(73) НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
УКРАЇНИ "КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ
ІНСТИТУТ", UA

(56)

(57) 1. Пристрій для обробки слабоструктурованих даних на основі мереж Байєса, що містить послідовно з'єднані шинами даних блок вводу даних, блок побудови мережі Байєса, блок пам'яті зберігання побудованої мережі Байєса, блок виводу результатів, зв'язані з блоком керування, який **відрізняється** тим, що додатково містить блок агрегування даних, вхід якого з'єднаний з виходом блока вводу даних, вихід з'єднаний з входом блока побудови мережі Байєса.

2. Пристрій за п. 1, який **відрізняється** тим, що додатково містить послідовно з'єднані шинами даних блок побудови таблиці умовних

імовірностей, вхід якого з'єднаний з виходом блока агрегування даних, блок пам'яті зберігання таблиці умовних імовірностей та блок пам'яті формування імовірнісного виводу, вхід якого з'єднаний з виходом блока пам'яті зберігання побудованої мережі Байєса, вихід з'єднаний з входом блока виводу результатів.

3. Пристрій за пп. 1, 2, який **відрізняється** тим, що додатково містить блок перевірки множини інстанційованих значень, вхід якого з'єднаний з виходом блока вводу даних, та послідовно з'єднані шинами даних блок побудови матриці емпіричних значень сумісного розподілу ймовірностей всієї мережі Байєса, вхід якого з'єднаний з виходом блока агрегування даних, блок пам'яті зберігання матриці емпіричних значень сумісного розподілу ймовірностей всієї мережі Байєса та блок пам'яті формування імовірнісного виводу, вхід якого з'єднаний з виходом блока пам'яті зберігання побудованої мережі Байєса і з виходом блока вводу даних, вихід з'єднаний з входом блока виводу результатів.

Корисна модель відноситься до обчислювальної техніки, і може бути застосована для перетворення вхідних статистичних даних у вихідний сигнал, що визначає поточний та прогнозний стан досліджуваного процесу чи системи.

В світі постійно накопичуються великі об'єми інформації, яка потребує спеціальної обробки та прийняття відповідних рішень на основі результатів такої обробки. Це в першу чергу пов'язано з тим що успішний розвиток будь-якої організаційної структури, установи, підприємства чи системи в першу чергу залежить від її здібності адекватно та оперативно реагувати на зміни зовнішнього оточення, а також здібності прогнозувати результати тих чи інших змін. Корисна модель може бути застосована в технічних, технологічних системах та при

моделюванні процесів економічних, екологічних, соціальних і біологічних систем [1].

Відомий пристрій для прогнозування випадкових подій [патент РФ RU №2290689C1, G06F17/18, 2006, бюл. №36] [2], який передбачає підвищення достовірності визначення стану виробничої чи телекомунікаційної системи в умовах неоднозначності (нечіткості) параметрів сигналів, що моделюються. До складу пристрою входять блок керування, блок моделі системи, блок імітаторів станів ділянок системи, блок формування сигналів відмов, блок реєстрації, блок перетворення сигналів відмов. Недоліками пристрою є достатньо вузькі функціональні можливості, оскільки стан системи визначається лише в просторі станів безвідмовної роботи і станів відмов, відсутні можливості врахування різноманітних вхідних даних.

(13) U

(11) 28751

(19) UA

Відомий статистичний аналізатор [заявка на винахід РФ RU №2004134293/09, G06F17/18, 2006, бюл. №13] [3], який базується на використанні блоку визначення належності вибірки до одного з чотирьох типів та застосуванні блоку обчислення математичного очікування та блоку визначення Гама-функції. Недоліками даного пристрою постають також достатньо вузькі функціональні можливості через застосування зафіксованої обмеженої типізації вибірок даних, відсутність можливості врахування різнотипних вхідних даних та недостатня якість визначення вихідного сигналу без здійснення імовірнісного виводу.

Відомі пристрій вибору оптимального дерева рішень для обробки даних [міжнародна заявка за РСТ №WO03/075182A1, G06F17/30, 12.09.2003] [4] та аналізатор дерева рішень [заявка на патент США US №2004/0199484A1, G06N5/02, G06F9/44, G06N7/02, G06N7/06, G06F17/00, 07 Д 1.2004] [5]. Вони базуються на застосуванні Байєсового підходу, проте вони не передбачають здійснення імовірнісного виводу вихідного сигналу на основі різнотипних слабоструктурованих даних, та потребують відносно багато часу на одержання вихідного сигналу.

Найбільш близьким до заявленої корисної моделі аналогом є система з адаптивною мережею Байєса (МБ) моделювання обробки даних [патент США US №6973459B1, G06F17/30, 06.12.2005] [6], яка містить функціональні блоки керування, вводу даних, виводу результатів, зберігання навчальних даних, побудови моделі мережі Байєса, пам'яті побудованої мережі, вибору і заміни моделі та блок визначення і прогнозування вихідного значення. Основним недоліком системи є орієнтація на однотипні структуровані дані, що суттєво звужує галузі її застосування.

Задача корисної моделі, що вирішується, полягає у розробці пристрою автоматичної обробки (перетворення) статистичних даних будь-якого типу для одержання вихідного сигналу, що визначає поточний та прогнозний стан системи чи процесу, який здатний враховувати спільно як кількісні, так і якісні вхідні дані, нечіткі та неповні дані, зменшує час одержання результату, підвищує якість і точність виводу.

Поставлена задача вирішується завдяки тому, що пристрій для обробки слабоструктурованих даних на основі мереж Байєса містить блок керування, послідовно з'єднані блок вводу даних, блок побудови мережі Байєса, блок пам'яті зберігання побудованої мережі Байєса, блок виводу результатів, блок агрегування даних, вхід якого з'єднаний з виходом блоку вводу даних, та вихід якого з'єднаний з входом блоку побудови мережі Байєса. Крім того, пристрій містить послідовно з'єднані шинами даних блок побудови таблиці умовних імовірностей, вхід якого з'єднаний з виходом блоку агрегування даних, блок пам'яті зберігання таблиці умовних імовірностей та блок пам'яті формування імовірнісного виводу, вхід якого з'єднаний з виходом блоку пам'яті зберігання побудованої мережі Байєса, та вихід якого з'єднаний з входом блоку виводу результатів. Додатково пристрій містить блок перевірки

множини інстанційованих значень, вхід якого з'єднаний з виходом блоку вводу даних, та послідовно з'єднані шинами даних блок побудови матриці емпіричних значень сумісного розподілу імовірностей всієї мережі Байєса, вхід якого з'єднаний з виходом блоку агрегування даних, блок пам'яті зберігання матриці емпіричних значень сумісного розподілу імовірностей всієї мережі Байєса та блок пам'яті формування імовірнісного виводу, вхід якого з'єднаний з виходом блоку пам'яті зберігання побудованої мережі Байєса і з виходом блоку вводу даних, та вихід якого з'єднаний з входом блоку виводу результатів. Структура операційних блоків пристрою та їх функціональна реалізація забезпечують побудову мережі Байєса вузлів станів системи за навчальними статистичними даними та визначення вихідного сигналу пристрою у виконання імовірнісного виводу у побудованій мережі Байєса. При цьому пристрій містить блок агрегування даних і передбачає два ланцюги блоків формування імовірнісного виводу для визначення результуючого сигналу: блоки виводу через таблицю умовних імовірностей (ТУІ) та через матриці емпіричних значень спільного розподілу імовірностей всієї мережі Байєса.

Технічним результатом корисної моделі є забезпечення обробки слабоструктурованих даних, поєднання даних різного типу, спільного врахування кількісних і якісних вхідних даних, зменшення часу на одержання результату і підвищення якості імовірнісного виводу для визначення вихідного результуючого сигналу.

Перелік фігур креслення.

Фіг. Структура пристрою для обробки слабоструктурованих даних на основі мереж Байєса.

На Фіг. зображені функціонально-операційні блоки пристрою та зв'язки між ними. Тонкими стрілками показані сигнали керування, а широкими стрілками - шини даних, що з'єднують відповідні блоки.

Суть корисної моделі полягає у створенні пристрою обробки слабоструктурованих даних, який складається з таких операційних блоків: блока керування 1, блока вводу даних 2, блока агрегування даних 3, блока побудови МБ 4, блока пам'яті зберігання побудованої МБ 5, блока перевірки множини інстанційованих значень 6, блока побудови матриці емпіричних значень спільного розподілу імовірностей всієї мережі Байєса 7, блока пам'яті зберігання матриці емпіричних значень спільного розподілу імовірностей всієї мережі Байєса 8, блока пам'яті формування імовірнісного виводу А 9, блока побудови ТУІ 10, блока пам'яті зберігання ТУІ 11, блока пам'яті формування імовірнісного виводу Б 12, блока виводу результатів 13. Блок керування 1 з'єднаний з блоком вводу даних 2, блоком агрегування даних 3, блоком побудови МБ 4, блоком перевірки множини інстанційованих значень 6, блоком побудови матриці емпіричних значень спільного розподілу імовірностей всієї мережі Байєса 7, блоком пам'яті формування імовірнісного виводу А 9, блоком побудови ТУІ 10, блоком пам'яті формування імовірнісного виводу Б

12 двосторонніми шинами керування, а з блоком виводу результатів 13 односторонньою шиною керування (від блоку 1 у бік блоку 13) для передачі сигналів керування. Сигнали керування на Фіг. зображені тонкими стрілками. Крім цього, певні блоки пристрою поєднані один з одним односторонніми шинами даних. Так вихід блоку 2 з'єднаний з входом блоку 3, блоку 6 і блоку 9; вихід блоку 3 - з входом блоку 4, блоку 7 і блоку 10; вихід блоку 4 - з входом блоку 5; вихід блоку 5 - з входом блоку 9 і входом блоку 12; вихід блоку 7 - з входом блоку 8; вихід блоку 8 - з входом блоку 9; вихід блоку 9 - з входом блоку 13; вихід блоку 10 - з входом блоку 11; вихід блоку 11 - з входом блоку 12; вихід блоку 12 - з входом блоку 13.

Блок керування 1 призначений для генерації сигналів керування, що забезпечують роботу всього пристрою, та обробку сигналів керування, які поступають від зв'язаних з ним блоків.

Блок вводу даних 2 призначений для вводу даних в пристрій, які складаються з а: множини навчальних даних та б: множини інстанційованих значень (значень вузлу МБ з одиничною ймовірністю).

Блок агрегування даних 3 призначений для виконання операції агрегування даних відносно множини навчальних даних введених в блок вводу даних 2. Операція агрегування здійснюється наступним чином, якщо якийсь вузол з множини навчальних даних приймає числові значення, то ці числові значення замінюються інтервальними оцінками.

Блок побудови МБ 4 призначений для виконання операції побудови МБ за навчальними даними, переданими з блока агрегування даних 3. Для побудови МБ застосовується евристичний метод побудови МБ за навчальними даними [7], який складається з двох етапів. На першому етапі здійснюється розрахунок значення взаємної інформації між всіма вузлами мережі. На другому етапі виконується цілеспрямований пошук, який застосовує в якості функції оцінки описання мінімальною довжиною.

Блок пам'яті зберігання побудованої МБ 5 призначений для зберігання побудованої МБ.

Блок перевірки множини інстанційованих значень 6 призначений для перевірки множини інстанційованих значень. Якщо множина інстанційованих значень не порожня, то пристрій продовжує працювати з групою операційних блоків 7-9, в протилежному випадку з групою операційних блоків 10-12.

Блок побудови матриці емпіричних значень спільного розподілу ймовірностей всієї мережі Байєса 7 призначений для побудови матриці емпіричних значень спільного розподілу ймовірностей всієї мережі Байєса за навчальними даними.

Блок пам'яті зберігання матриці емпіричних значень спільного розподілу ймовірностей всієї мережі Байєса 8 призначений для зберігання матриці емпіричних значень спільного розподілу ймовірностей всієї мережі Байєса за навчальними даними.

Блок пам'яті формування ймовірнісного виводу А 9 призначений для розрахунку ймовірнісного виводу. Операція виконується у такий спосіб:

маючи інформацію про структуру моделі МБ та інстанційовані значення виконується послідовний перебір всіх вузлів МБ. Якщо вузол не є інстанційованим, то виконується розрахунок усіх можливих станів цього вузла. Після чого проводиться послідовний перебір усіх строк матриці емпіричних значень спільного розподілу ймовірностей всієї МБ, і якщо значення вузла строки співпадає із значеннями інстанційованих вузлів і значенням аналізованого вузла, то відповідне значення додається до значення ймовірності відповідного стану аналізованого вузла. Після цього виконується нормування значень ймовірностей станів аналізованого вузла.

Блок побудови ТУІ 10 призначений для побудови ТУІ для кожного вузла МБ за множиною навчальних даних.

Блок пам'яті зберігання ТУІ 11 призначений для зберігання ТУІ всіх вузлів МБ.

Блок пам'яті формування ймовірнісного виводу Б 12 призначений для виконання ймовірнісного виводу. Для здійснення ймовірнісного виводу застосовується класичний метод прямого розповсюдження ймовірностей по МБ [8]. Спочатку рахуються ймовірності значень кореневих вузлів, тобто вузлів у яких відсутні батьківські вузли. Після цього визначаються ймовірності усіх інших вузлів за формулою Байєса та ТУІ.

Блок виводу результатів 13 призначений для здійснення виводу остаточного результату - розрахованих прогнозних значень вузлів МБ.

Функціонування запропонованого пристрою складається з такої послідовності операцій.

Крок 1. Блок керування 1 подає сигнал 1 для приведення блока вводу даних 2 в робочий стан.

Крок 2. В блок вводу даних 2, через шину даних 3, здійснюється введення даних в пристрій. Дані для вводу складаються з множини навчальних даних та множини інстанційованих значень.

Крок 3. Після того, як введення даних завершено, з блока вводу даних 2 подається сигнал 2, про успішне завершення операції введення даних в пристрій, в блок керування 1.

Крок 4. Блок керування 1 подає сигнал 4 для приведення блока агрегування даних 3 в робочий стан.

Крок 5. В блоці агрегування даних 3 здійснюється операція агрегування числових даних, для чого з блок вводу даних 2 через шину даних 6 передається множина навчальних даних. Операція агрегування здійснюється у такий спосіб: якщо якийсь вузол з множини навчальних даних приймає числові значення, то ці числові значення замінюються інтервальними оцінками.

Крок 6. З блока агрегування даних 3 в блок керування 1 подається сигнал 5 про успішне завершення, операції агрегування.

Крок 7. Блок керування 1 подає сигнал 7 для приведення блока побудови МБ 4 в робочий стан.

Крок 8. Блок побудови МБ 4 здійснює побудову МБ за навчальними даними, які передаються по шині даних 9 з блока агрегування даних 3. Для побудови МБ застосовується евристичний метод побудови МБ за навчальними даними [7], який складається з двох етапів. На

першому етапі здійснюється розрахунок значення взаємної інформації між всіма вузлами мережі. На другому етапі виконується цілеспрямований пошук, який застосовує в якості функції оцінки описання мінімальною довжиною.

Крок 9. В блоці пам'яті 5 здійснюється зберігання побудованої МБ, для чого з блока побудови МБ 4, по шині даних 10, передається відповідна інформація.

Крок 10. Блок побудови МБ 4, подає сигнал 8 в блок керування 1 про успішне виконання операції побудови МБ та збереження МБ в блоці пам'яті 5.

Крок 11. Блок керування 1 подає сигнал 11 приведення блока перевірки множини інстанційованих значень 6 в робочий стан.

Крок 12. В блоці 6 здійснюється перевірка множини інстанційованих значень. Для чого по шині даних 13 передається множина інстанційованих значень. Якщо множина інстанційованих значень не порожня пристрій продовжує працювати з кроку 13, в противному випадку з кроку 23.

Крок 13. З блока 6 в блок керування 1 подається сигнал 12, який містить інформацію про те, що множина інстанційованих значень не порожня.

Крок 14. Блок керування 1 подає сигнал 14 приведення блока 7 в робочий стан.

Крок 15. В блоці 7 здійснюється побудова матриці емпіричних значень спільного розподілу ймовірностей всієї мережі Байєса за навчальними даними. Ця інформація передається по шині даних 16 з блока агрегування даних 3. В блоці 7 виконується розрахунок ймовірностей здійснення кожної можливої спільної події всіх вузлів мережі.

Крок 16. В блоці пам'яті 8 здійснюється зберігання матриці емпіричних значень спільного розподілу ймовірностей всієї мережі Байєса, для чого з блока 7 по шині даних 17 передається відповідна інформація.

Крок 17. Блок 7 подає сигнал 15 в блок керування 1 про успішне виконання операцій побудови матриці емпіричних значень спільного розподілу ймовірностей всієї мережі Байєса та збереження цієї матриці в блоці пам'яті 8.

Крок 18. Блок керування 1 подає сигнал 18 приведення блока 9 в робочий стан.

Крок 19. В блоці 9 здійснюється розрахунок ймовірнісного виводу. Для цього з блока вводу даних 2 в блок 9 по шині даних 20 передається множина інстанційованих значень, а з блока пам'яті зберігання побудованої МБ 5 по шині даних 21 передається інформація про структуру моделі МБ. Надалі маючи інформацію про структуру МБ та інстанційовані значення виконується послідовний перебір всіх вузлів МБ. Якщо вузол не є інстанційованим, то потрібно виконати розрахунок усіх можливих станів цього вузла. Для чого з блока пам'яті 8 в блок 9 по шині даних 22 передаються значення матриці емпіричних значень спільного розподілу ймовірностей всієї мережі Байєса. Після чого проводиться послідовний перебір усіх строк матриці емпіричних значень спільного розподілу ймовірностей всієї МБ. Якщо значення вузла строки співпадає зі значеннями інстанційованих вузлів і значенням

аналізованого вузла, то відповідне значення додається до значення ймовірності відповідного стану аналізованого вузла. Після цього виконується нормування значень ймовірностей станів аналізованого вузла.

Крок 20. Блок 9 подає сигнал 19 в блок керування 1 про успішне виконання операції формування ймовірнісного виводу.

Крок 21. Блок керування 1 подає сигнал 23 приведення блока виводу результатів 13 в робочий стан.

Крок 22. Блок виводу результатів 13 здійснює вивід остаточного результату - розрахованих прогнозних значень вузлів по шині даних 25 для чого з блока пам'яті формування ймовірнісного виводу 9 по шині даних 24 передається відповідна інформація.

Крок 23. З блока 6 в блок керування 1 подається сигнал 12, який містить інформацію про те, що множина інстанційованих значень порожня.

Крок 24. Блок керування 1 подає сигнал 26 приведення блока побудови ТУІ 10 в робочий стан.

Крок 25. Блок 10 виконує побудову ТУІ для кожного вузла МБ, для чого з блока агрегування даних 3 по шині даних 28 передаються навчальні дані.

Крок 26. В блоці пам'яті 11 здійснюється зберігання ТУІ всіх вузлів МБ, для чого з блока 10 по шині даних 29 передається відповідна інформація.

Крок 27. Блок 10 подає сигнал 27 в блок керування 1 про успішне виконання операцій побудови ТУІ для всіх вузлів МБ та їх збереження в блоці пам'яті 11.

Крок 28. Блок керування 1 подає сигнал 30 приведення блока 12 в робочий стан.

Крок 29. В блоці 12 виконується ймовірнісний вивід. Для цього з блока 11 по шині даних 32 передаються ТУІ, а з блока пам'яті зберігання побудованої МБ 5 по шині 33 передається інформація про структуру моделі МБ. Для здійснення ймовірнісного виводу застосовується класичний метод прямого розповсюдження ймовірностей по МБ [8]. Спочатку рахуються ймовірності значень кореневих вузлів, тобто вузлів у яких відсутні батьківські вузли. Після цього рахуються ймовірності усіх інших вузлів за формулою Байєса та ТУІ.

Крок 30. Блок 12 подає сигнал 31 в блок керування 1 про успішне виконання операції формування ймовірнісного виводу.

Крок 31. Блок керування 1 подає сигнал 23 приведення блока виводу результатів 13 в робочий стан.

Крок 32. Блок виводу результатів 13 здійснює вивід остаточного результату - розрахованих прогнозних значень вузлів, по шині даних 25, для чого з блока пам'яті формування ймовірнісного виводу 12, по шині даних 34, передається відповідна інформація.

Обробка запропонованим пристроєм кількісних і якісних даних, перетворення слабоструктурованих та нечітких даних, зменшення часу на одержання результуючого вихідного сигналу відбувається через

використання блоку агрегування даних 3 та блоків 4,5 і блоків 7-12, що забезпечують визначення вихідного сигналу пристрою на основі принципу ймовірнісного виводу у мережі Байєса вузлів станів системи, яка будується за навчальними статистичними даними.

Наявність в пристрої двох ланцюгів операційних блоків формування ймовірнісного виводу для визначення результуючого сигналу (блоків виводу через таблицю умовних ймовірностей (блоки 10-12) та блоків виводу через матриці емпіричних значень спільного розподілу ймовірностей всієї мережі Байєса (блоки 7-9) забезпечує підвищення якості і точності ймовірнісного виводу для визначення вихідного результуючого сигналу. Така структурна схема пристрою для одержання більш точного і якісного ймовірнісного виводу передбачає використання необхідних блоків в залежності від вхідних даних, які поступають на вхід пристрою - у блок вводу даних 2.

Таким чином, заявлена корисна модель - пристрій для обробки слабоструктурованих даних на основі мереж Байєса, - вирішує поставлену задачу і забезпечує досягнення технічного результату.

Практична реалізованість корисної моделі впливає з використання операційних блоків, які забезпечують виконання елементарних функціональних процедур з введення, зберігання, перетворення, передачі даних та видачі вихідного результуючого сигналу.

Список використаної літератури

1. Бидюк П.И., Терентьев А.Н., Коршевиюк Л.А. Применение байесовских сетей для интеллектуального анализа данных// Интеллектуальные системы принятия решений і прикладні аспекти інформаційних технологій: Матеріали НІЖ. Т.2 - Херсон: ПП Вишемирський В.С., 2007. - с.15-17.

2. Устройство для прогнозирования случайных событий. Пат. RU 2290689C1 РФ, МПК G06F17/18; Оpubл. 27.12.2006, Бюл. изобретений и полезных моделей №36.

3. Статистический анализатор. Заявка на изобретение RU №2004134293/09 РФ, МПК G06F17/18; Оpubл. 27.12.2006, Бюл. изобретений и полезных моделей №13.

4. Apparatus and method for selecting an optimal decision tree for data mining. Міжнародна заявка за РСТ. - № WO03/075182A1 (PCT/KR03/00407), МПК G06F17/30; Оpubл. 12.09.2003.

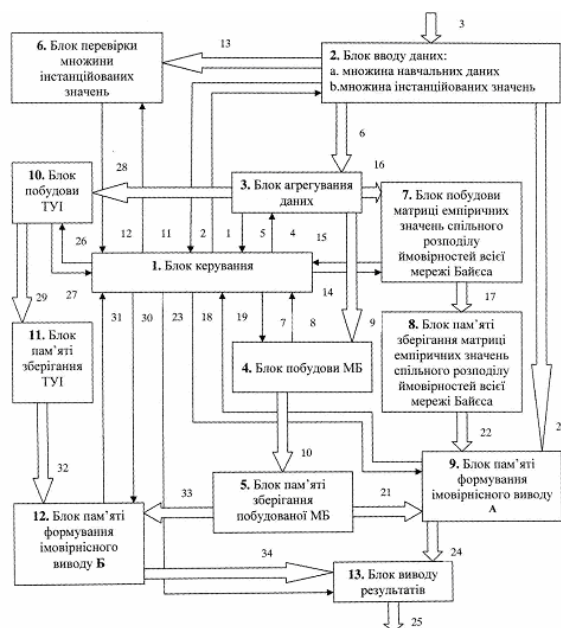
5. Decision Tree Analysis. Заявка на патент US №2004/0199484A1 США, МПК G06N5/02, G06F9/44, G06N7/02, G06N7/06, G06F17/00; Оpubл. 07.11.2004.

6. Adaptive Bayes network data mining modeling. Пат. US №6973459B1 США, МПК G06F17/30; Оpubл. 06.12.2005.

7. Терентьев А.Н., Бидюк П.И. Эвристический метод построения Байесовских сетей// Математические машины и системы. - 2006. - 3.- с.12-23.

8. Бидюк П.И., Терентьев А.Н., Гасанов А.С. Построение и методы обучения Байесовских

сетей// Кибернетика и системный анализ. - 2005. - 4. - с.133-147.



Фиг.