

***В. Вартузов, С. Бабічев, В. Литвиненко, А. Фефелов**

*Державний науково-дослідний експертно-криміналістичний центр МВС України,
Херсонський національний технічний університет

МОДЕЛЬ АНАЛІТИЧНОЇ СИСТЕМИ ПРОФІЛЮВАННЯ НАРКОТИЧНИХ РЕЧОВИН НА ОСНОВІ МЕРЕЖІ БАЙЕСА

© Вартузов В., Бабічев С., Литвиненко В., Фефелов А., 2012

Наведено аналітичну модель системи профілювання наркотичних речовин, що дозволяє оцінити ймовірність шляху транзиту наркотиків та місце їх виготовлення. Побудована мережа Байєса, що враховує характер взаємозв'язку між вузлами системи, в якій як цільовий був прийнятий вузол, що визначає назву лабораторії виробництва наркотичної речовини. Показано реалізацію запропонованої моделі на прикладі трьох лабораторій, шести дилерів, які поставляють товар до місця призначення і чотирьох компонент наркотичної речовини, які додають дилери в процесі транспортування товару.

Ключові слова: байєсовська мережа, профілювання наркотичних речовин.

In the article presents the analytical model of the system profiling of the drugs allowing estimating the probability of drug transit route of the drugs and placing their manufacture. Bayesian network is constructed, taking into account the nature of the relationship between the nodes of the system, which has as its target was adopted node, defines the name of the laboratory drug production. Shows the implementation of the model proposed by the example of the three laboratories, six dealers that supply goods to the destination and the four components of a narcotic substance, added dealers during its transportation.

Key words: Bayesian network profiling of drugs.

Вступ

Зловживання забороненими наркотиками – серйозна проблема всесвітнього характеру, яку можна вирішити лише у межах міжнародного співробітництва. Один з найефективніших заходів боротьби з наркотиками – запобігання їх виробництву. У разі синтетичних наркотиків контроль над прекурсорами є важливим компонентом стратегії запобігання їх виробництву. Дослідження в області створення аналітичної системи профілювання наркотичних речовин можуть забезпечити правоохоронні органи корисною інформацією. Може бути виявлено подібність між пробами за хімічним складом, а матеріал з різних вилучених партій можна класифікувати за групами родинних проб. Найважливішим наслідком для правоохоронних органів є той факт, що цим можна виявити специфічні зв'язки, наприклад, між постачальниками наркотиків і особами, що вживають наркотики, визначити тенденції поширення наркотиків і мережі їх збуту, а також ідентифікувати джерела, зокрема географічне походження проб.

Постановка проблеми

У [1] показано, що незаконні наркотики та психотропні речовини, як правило, являють собою складну суміш, і лише в окремих випадках містять чисту наркотичну речовину. Поряд з наркотичною або психотропною речовиною проби можуть містити одну або більше різних типів основних компонентів. Кожну компоненту залежно від характеру аналізу речовини можна охарактеризувати за допомогою різних якісних або кількісних характеристик. У цей момент часу одні характеристики є відомими, а інші – невідомими. Наприклад, хроматограма наркотичної речовини або процентний вміст його компонент, визначений шляхом хімічного аналізу, є підмножиною відомих характеристик, а

випускна лабораторія і шлях поставки – підмножиною невідомих. Завдання полягає в тому, щоб, маючи набір відомих характеристик досліджуваної хімічної речовини визначити його невідомі характеристики. У роботах [2–5] автори для розв'язання таких типів задач запропонували використовувати мережу Байєса. Основна проблема при цьому полягає у виборі структури мережі з урахуванням характеру взаємозв'язку між вузлами і апріорне визначення умовних ймовірностей вузлів предків мережі. Однак у криміналістиці і зокрема для виявлення шляхів постачання наркотичних речовин мережі Байєса сьогодні не одержали належного поширення.

Мета роботи – розроблення моделі системи профілювання наркотичних речовин на основі мережі Байєса.

Викладення основного матеріалу

Байєсовська мережа [4] являє собою пару $\langle G, V \rangle$, у якій перша компонента G – це спрямований нециклічний граф, що відповідає випадковим змінним і записується як набір умов незалежності: кожна змінна незалежна від її батьків в G . Друга компонента пари V – це множина параметрів, що визначають мережу. Кожна змінна $X^{(i)} \in G$ представляється у вигляді вершини. Повна спільна ймовірність БМ обчислюється за формулою:

$$P_B(X^{(1)}, \dots, X^{(N)}) = \prod_{i=1}^N P_B(X^{(i)} | Pa(X^{(i)})).$$

З математичного погляду БМ – це модель подання наявних і відсутніх імовірнісних залежностей. При цьому зв'язок $A \rightarrow B$ є причинним, коли подія A – причина виникнення B , тобто коли існує механізм, відповідно до якого значення, прийняте A , впливає на значення, прийняте B .

Нехай Ω – множина подій випадкових експериментів. Цей вибірковий простір містить всі можливі значення випадкової змінної. Припустимо, що дві змінні E і H деяким чином пов'язані між собою. Умовна ймовірність події E визначається за виразом:

$$p(E | H_k) = \frac{p(E \cap H_k)}{p(H_k)}, \quad (1)$$

де \cap – операція перетину множин. Ця залежність дає змогу визначити ймовірність події E , якщо існує конкретна подія H_k . Якщо E_1, E_2, \dots, E_n – взаємовиключні події такі, що $\cup_{i=1}^n E_i = \Omega$, то кажуть, що події E_i формують повну (вичерпну) множину. Дві змінні не зв'язані (не перетинаються), якщо вони не мають однакових значень. Якщо дві змінні є вичерпними і незв'язаними, то можна записати, що

$$E = \cup_i (E \cap H_i), \quad (E \cap H_i) \cap (E \cap H_j) = \emptyset, \quad i \neq j. \quad (2)$$

Теорія побудови байєсівських мереж ґрунтується на припущенні, що події є вичерпними і не перетинаються. Якщо ця умова не виконується, то результати застосування мережі будуть неконсистентними (тобто неточними). У випадку, коли події вичерпні і не перетинаються, то ймовірність події E можна обчислити за допомогою умовних ймовірностей:

$$p(E) = \sum_{i=1}^n p(E \cap H_i) = \sum_{i=1}^n p(E | H_i) \cdot p(H_i). \quad (3)$$

Використовуючи рівняння (3), суму перетинів події E з H можна подати так:

$$p(E \cap H_k) = p(E | H_k) \cdot p(H_k) = p(H_k | E) \cdot p(E). \quad (4)$$

З рівності $p(E | H_k) \cdot p(H_k) = p(H_k | E) \cdot p(E)$ знайдемо, що

$$p(H_k | E) = \frac{p(E | H_k) \cdot p(H_k)}{p(E)},$$

а із врахуванням (3) отримаємо вираз:

$$p(H_k | E) = \frac{p(E | H_k) \cdot p(H_k)}{\sum_{i=1}^n p(E | H_i) \cdot p(H_i)}, \quad (5)$$

який являє собою формулу Байеса. На основі цієї формули будується байесовська мережа.

В (5) H_k означає будь-яку гіпотезу з n можливих. Ймовірності $p(E | H_k)$ задають експерти апріорно або розраховують за навчальними даними [6, 7]. Тобто їх можна розглядати як відповідь на запитання: “Якою буде ймовірність деякого виміру, якщо відомо, яка гіпотеза була реалізована?”. Ймовірності $p(E | H_k)$ є дуже корисними, тому що, як правило, легше знайти ймовірність послідовності подій типу причина–наслідок, ніж навпаки. Значення $p(H_k)$ називають апріорними ймовірностями; вони визначають початкові ймовірності для всіх гіпотез. Сила байесівського методу полягає в тому, що апріорні ймовірності можна уточнювати (оновлювати) відповідно до реалій проходження досліджуваного процесу. Це дає змогу уточнювати ймовірності подій при надходженні додаткової інформації. Знаменник виразу (5) можна розглядати як нормуючий член, який встановлює значення ймовірності між 0 та 1.

Профілювання – це процедура ідентифікації й аналізу домішок у наркотичних речовинах, вилучених з незаконного обігу і на основі цієї інформації встановлення, якими методами вони були синтезовані, а також визначення спільних шляхів походження.

Профілювання передбачає систематичне збирання і розділення за групами вилучених з незаконного обігу наркотичних речовин на основі фізичної і хімічної інформації, отриманої внаслідок їх дослідження, яке містить аналіз і визначення домішок, щоб порівняти різні наркотичні зразки. Отримані результати забезпечують важливе джерело інформації для подальших аналізів.

Профіль вмісту хімічних домішок в пробі – це профіль вмісту домішок, які є результатом спільного екстрагування і виготовлення (профіль вмісту домішок), а також розріджувачів.

Наркотики та психотропні речовини, як правило, є складною сумішшю, лише в окремих випадках містять чисту наркотичну речовину, незалежно від того, чи мають вони рослинне походження. Поряд з наркотичною чи психотропною речовиною проби можуть містити один або більше трьох різних типів основних компонентів [1]:

- **природні компоненти**, присутні в сировині (наприклад, лист коки, опій) і використовувані у виробництві певних наркотиків “рослинного походження”, таких як кокаїн або героїн; в процесі виробництва наркотику їх екстрагують спільно з наркотичною речовиною і не видаляють повністю із кінцевого продукту;
- **побічні продукти**, що утворюються під час виробництва наркотику і пов’язані із способом його виготовлення;
- **розчинники та розбавлювачі**, медичні препарати тощо, які можуть бути додані на будь-якому наступному за виготовленням етапі в ланцюзі збуту наркотиків.

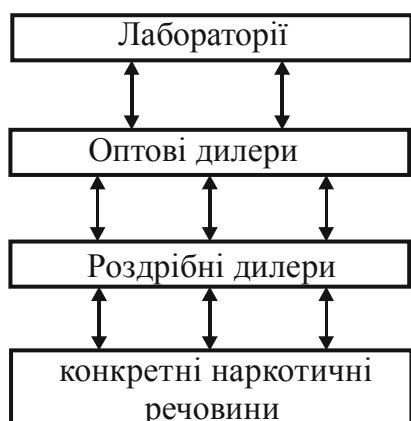


Рис. 1. Загальна схема моделі системи профілювання наркотичних речовин

Проби синтетичних наркотиків, таких як амфетаміни, можуть містити різні синтетичні домішки, зокрема залишки основних хімічних речовин, що використовувалися в процесі виготовлення наркотиків, а також побічні продукти, що є результатом побічних реакцій. Наявність і відносна концентрація таких домішок залежать від якості вихідних матеріалів, шляхів їхнього синтезу, умов проведення хімічних реакцій, ступеня очищення кінцевого продукту (продуктів) і найбільше – від кваліфікації хіміка, що працює в підпільних умовах. Крім того, на різних рівнях ланцюга розповсюдження до наркотиків можуть додаватися різні домішки, наприклад, медичні препарати, з метою збільшення обсягу кінцевого продукту та, відповідно, доходу.

Загальну схему моделі системи профілювання наркотичних речовин наведено на рис. 1.

За цією моделлю з кожною наркотичною речовиною можуть бути асоційовані різноманітні поняття, які його характеризують. При цьому пропонується модель не обмежує дослідника жорсткою односпрямованістю дій, а може гнучко використовуватися як для виявлення причинних характеристик (наприклад, за відомим хімічним складом і місцем вилучення визначити лабораторію та можливі шляхи поставки), так і передбачати наслідки (наприклад, якщо було виявлено, що деяка лабораторія випустила партію речовини, то визначити, появи якого типу речовин і на якій території слід очікувати найближчим часом). Описані вище властивості моделі на рис. 1 проілюстровано стрілками, що показують рух інформації між виділеними блоками. Слід зазначити, що внаслідок специфіки роботи байєсівських мереж всі висновки за цією моделлю щодо шуканої інформації мають імовірнісний характер і наведені у вигляді рангового (за значеннями імовірності вірності того або іншого висновку) списку. Остаточне рішення приймає дослідник.

Для розв'язання поставленої задачі пропонується використовувати класифікувальну структуру байєсівської мережі, приклад якої зображено на рис. 2. Як видно з рисунка, така мережа враховує не тільки незалежні, але й пов'язані між собою атрибути. Це зручно, наприклад, у тому випадку, коли необхідно отримати інформацію про канал постачання. Структуру байєсівської мережі, орієнтованої на розв'язання задачі профілювання наркотичних речовин, наведено на рис. 3. Дилери 1,2 і 3 являють собою клас оптових дилерів, а 4,5 і 6 – роздрібні дилери.

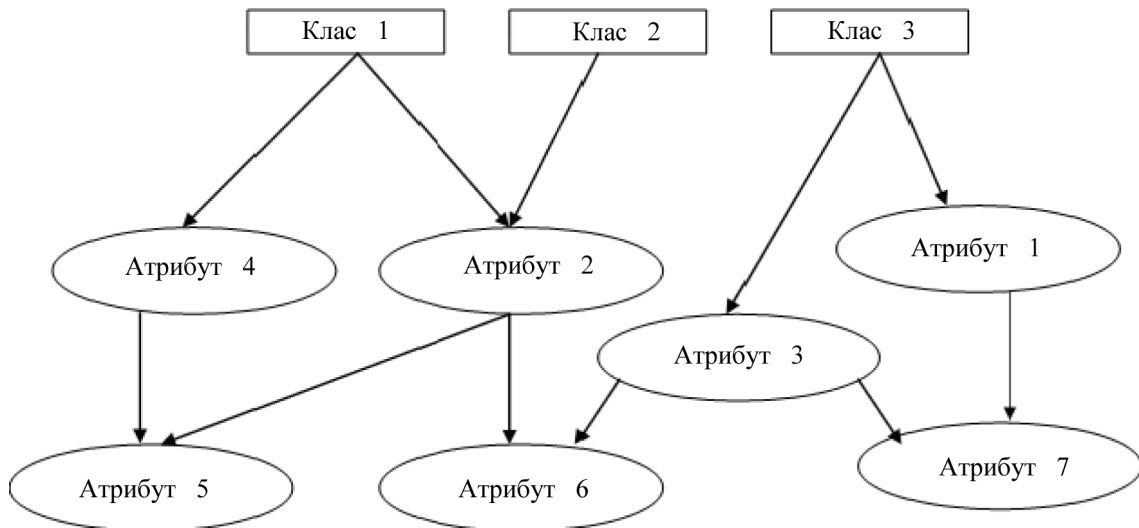


Рис. 2. Приклад структури байєсівської мережі для розв'язання задач класифікації

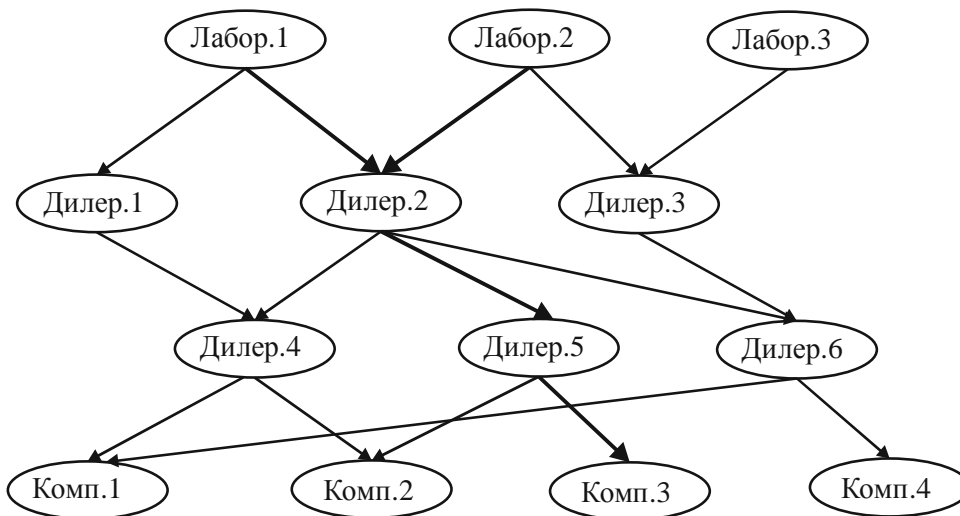


Рис. 3. Структура байєсівської мережі для розв'язання задачі профілювання наркотичних речовин

Так, наприклад, аналіз наркотичної речовини показав, що воно поряд з іншими компонентами містить компоненту 3. Як видно з рис. 3, цю компоненту було додано роздрібним дилером 5, який отримав товар від оптового дилера 2. Поставки цього дилера йдуть, своєю чергою, від лабораторій 1 і 2. Уточнити конкретну лабораторію, яка постачала товар, можна, аналізуючи інші компоненти наркотичної речовини. Отже, встановивши на підставі існуючої бази даних [1] характер взаємозв'язку між вузлами мережі, можна з високим ступенем ймовірності передбачити шлях поставки товару в пункт його збуту.

Розв'язували задачу побудови мережі Байєса, використовуючи програмне середовище GeNIe 4.0. Як цільовий вузол приймали вузол, який визначає лабораторію виробництва наркотиків. Модель байєсівської мережі для розв'язання задачі профілювання наркотичних речовин наведено на рис. 4 і 5.

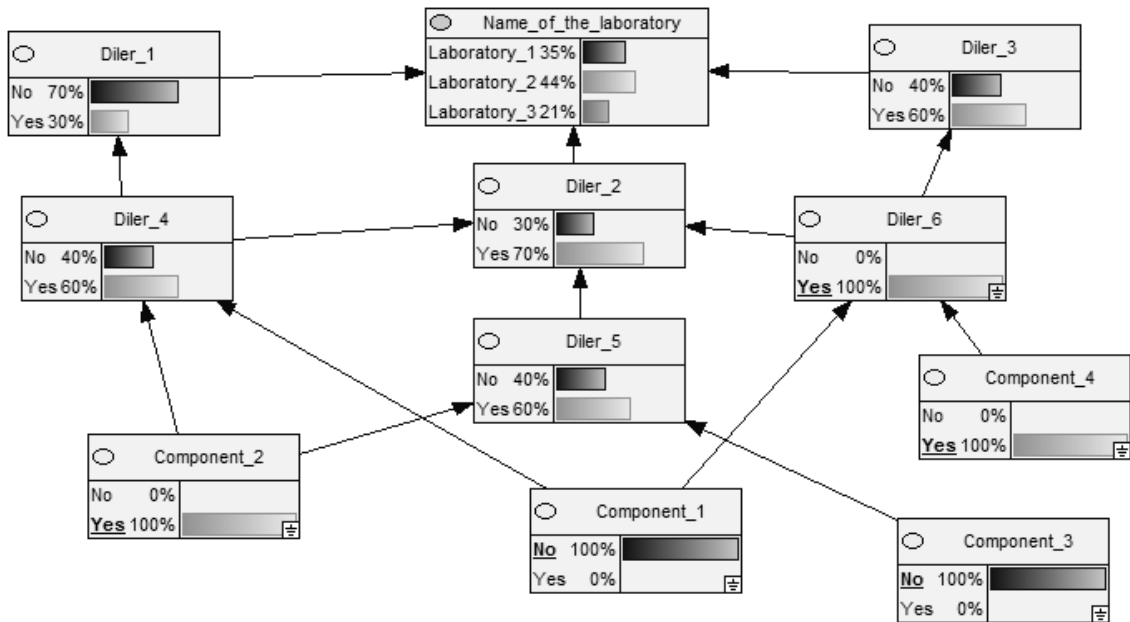


Рис. 4. Модель байєсівської мережі для розв'язання задачі профілювання наркотичних речовин за наявності компонент 2 і 4

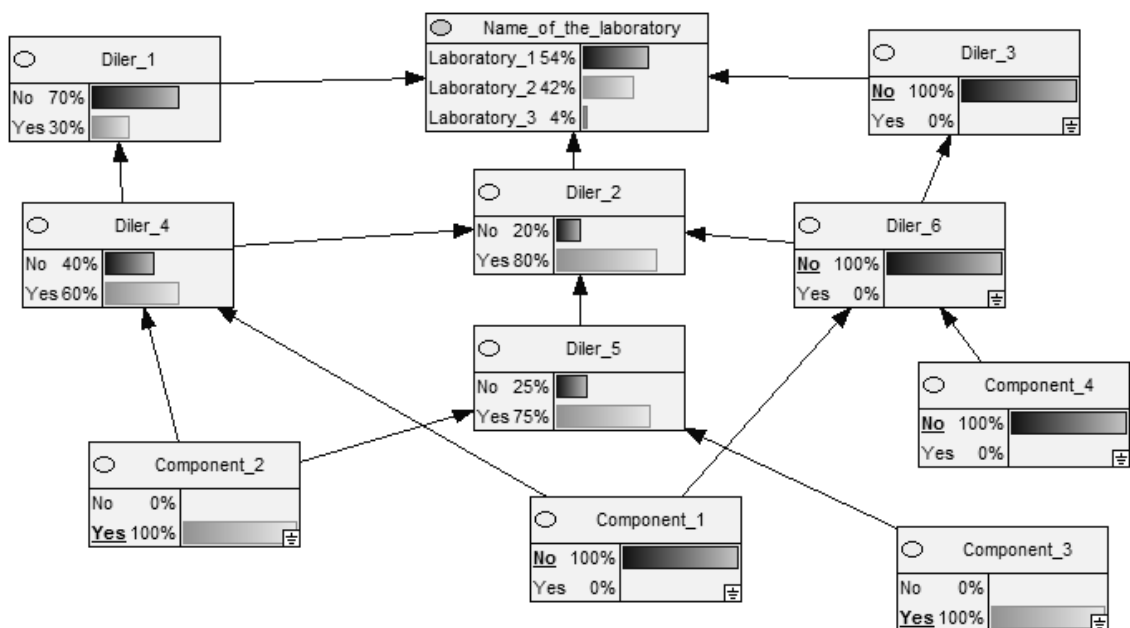


Рис. 5. Модель байєсівської мережі для розв'язання задачі профілювання наркотичних речовин за наявності компонент 2 і 3

Як видно з рис. 4, за наявності в наркотичній речовині компонент 2 і 4, з ймовірністю 44 % можна стверджувати, що виробництво здійснювалося в лабораторії 2, а поставку проводили через роздрібного дилера 6, який, своєю чергою, отримав товар з ймовірністю 70 % від оптового дилера 2 і з ймовірністю 60 % від оптового дилера 3. Модель, яку наведено на рис. 5, дає змогу зробити висновок, що за наявності в речовині компонент 2 і 3 з високою ймовірністю можна стверджувати, що виробництво здійснювалося в лабораторії 1 і шлях поставки проходив через дилерів 2 і 5.

Висновок

У роботі розроблено модель байєсівської мережі для розв'язання задачі профілювання наркотичних речовин. Запропонована модель дає змогу, аналізуючи наркотичну речовину, оцінити ймовірність місця виробництва цієї речовини і визначити можливі шляхи його транспортування до місця збуту. В процесі налагодження моделі враховували два стани компоненти наркотичної речовини: присутній або відсутній. При цьому в силу рівноцінності цих станів кожному з них було приписано ймовірність 50 %. Однак за наявності бази даних з точним складом компонентів цієї чи іншої речовини і відомою його передісторією можливе точніше налагодження моделі з урахуванням обсягу компоненти. Подальшим напрямом досліджень авторів є реалізація цієї моделі за реальними даними відповідно до [1].

1. Организация Объединённых Наций. Описание свойств наркотиков/составление профилей содержания примесей. Основные понятия и концепции. Руководство для национальных правоохранительных органов и лабораторий экспертизы наркотиков. ST/NAR/32/Rev.1 / Организация Объединённых Наций. – Нью-Йорк, 2004. – С. 1–18. 2. Бідюк П.І., Терентьев О.М., Коршевнюк Л.О. Интеллектуальный анализ слабоструктурированных данных за помощью байесовых сетей / Звіт по результатам виконання робіт за грантом НТУУ „КПІ” № 3/5-ГР, 2006, 2007р. – 85 с. 3. Згуровський М.З., Бідюк П.І., Терентьев О.М. Системна методика побудови байєсових мереж // “Наукові вісті” НТУУ “КПІ”. – 2007. – №4. – С. 47–61. 4. Терентьев А.Н., Бідюк П.И., Коршевнюк Л.А. Байесовская сеть – инструмент интеллектуального анализа данных // Проблемы управления и информатики. – К.: ИКИ НАНУ-НКАУ, 2007. – № 4. – С. 83–92. 4. Terentyev A.N., Gasanova L.T. Bayesian networks in credit scoring / The second international conference on control and optimization with industrial applications (COIA-1008), Baku, June 2-8, 2008. – Baku: Institute of applied mathematics BSU, 2008. – P. 171. 5. Терентьев А.Н., Бідюк П.И., Коршевнюк Л.А. Алгоритм вероятностного вывода в байесовских сетях / Материалы IX международной научно-технической конференции “Системный анализ и информационные технологии” (САИТ-2007), г. Киев, 15–19 мая 2007. – К.: Екмо, 2007. – С. 76. 6. Бідюк П.И., Терентьев А.Н. Метод вероятностного вывода в байесовских сетях по обучающим данным // Кибернетика и системный анализ. – 2007. – № 3. – С. 93–99.