УДК 007:681.3.00

О.С. Балабанов

ПРАВИЛА ПІДБОРУ СЕПАРАТОРІВ У БАЄСІВСЬКИХ МЕРЕЖАХ

Запропоновано і обґрунтовано набір тверджень та правил, які допомагають знаходити мінімальні за розмірами сепаратори в моделях ймовірнісних залежностей, структурованих як ациклічні орієнтовані графи (тобто в баєсівських, гаусових та гібридних мережах). Правила призначені для методів виведення структури моделі з даних за допомогою знаходження сепараторів (тобто виявлення умовної незалежності змінних). Ми сформулювали необхідні вимоги до членів мінімальних сепараторів, виходячи з критерію d-сепарації та властивостей ациклічних орграфів. Ці вимоги та правила дозволяють відсіяти деякі кандидати у сепаратор і спростити задачу пошуку сепараторів, а, також, і задачу ідентифікації структури моделі залежностей.

Вступ

Предметом розгляду є ймовірнісні моделі залежностей, структуровані як ациклічні орієнтовані графи (АОГ-моделі) [1–8]. Такі моделі звуться баєсівськими мережами, коли змінні – номінальні (дискретні), гаусовими мережами, коли змінні – неперервні, а залежності – лінійні з нормальними розподіленнями, та, відповідно, гібридними мережами, коли є змінні різних типів. АОГ-моделі стають популярним апаратом і застосовуються для моделювання стохастичних залежностей та каузальних відношень.

Графові моделі залежностей забезпечують більш системну репрезентацію, ніж традиційні види моделей – логічні формули, регресійні рівняння, правила класифікації, діагностики чи розпізнавання образів і т.д. АОГ-моделі глибше відображають механізми, задіяні у предметній галузі, і в принципі здатні відбивати каузальні відношення. Достоїнства АОГ-моделей – наочність, компактність, здатність відображати причинно-наслідкові зв'язки, обчислювальна ефективність ймовірнісного виведення від свідчень. Ці властивості забезпечують ефективне розв'язання задач медичної, технічної і комп'ютерної діагностики [9], розпізнавання мови [10], прогнозування наслідків рішень і дій людини або агента і т.д. АОГ-моделі становлять апарат ймовірнісних експертних систем.

Привабливим способом застосування стало виведення (ідентифікація) АОГ-моделі системи з статистичних даних спостережень за системою. Виведення АОГ-моделей з даних спостережень сприяє „інсайту” щодо структури і зв’язків процесів у досліджуваній предметній області, дозволяє відтворити структурно адекватну систему впливів. Ця аналітична технологія широко застосовується у дослідженнях в медико-біологічній галузі, зокрема, для аналізу механізмів експресії генів [11]. Також така технологія дозволяє вивести з даних модель для прогнозування погоди [12] та ін. Наприклад, виходячи з даних Світового банку, проведено аналіз причин і наслідків бідності в 80 країнах, що розвиваються [13]. Результатом аналізу (за допомогою алгоритма ‘PC’) стала модель, фрагмент якої показано на рис. 1. Модель показує фактори народжуваності в бідних країнах. GDP – це величина прибутку сімейного господарства на душу населення; Gini index – коефіцієнт Джині (індекс концентрації доходів).

Відомо, що проблема виведення структур баєсівських мереж з даних є NP-важка [14]. Тож відомі методи стикаються з великими труднощами вже коли кількість змінних досліджуваної системи сягає кількох десятків. Ідентифікація структури АОГ-моделі методами “constraint-based” підходу (на які тут орієнтуємося) ґрунтується на пошуку сепараторів. Оскільки при пошуку сепараторів виникають складні переборні задачі, бажано знаходити мінімальні сепаратори. Це важливо також і в задачах виведення

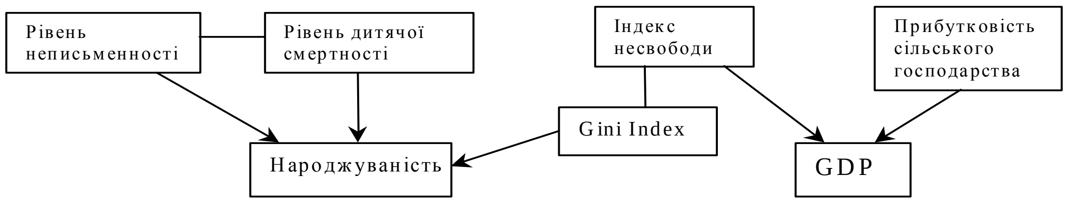

Рис.1. Виведена модель факторів народжуваності в бідних країнах

наслідків від свідчень на АОГ-моделі, бо інформація розповсюджується у структурі моделі через сепаратори. Ми показуємо нові можливості для зменшення складності розв'язання названих задач, пропонуємо засоби для звуження області пошуку мінімальних сепараторів і даємо відповідне граф-теоретичне обґрунтування. Матеріал викладається переважно на строгому формальному рівні (з графічними ілюстраціями). Наскільки відомо автору, результати є цілком оригінальними. Всі результати роботи чинні для всього класу АОГ-моделей, включаючи будь-які гібридні моделі.

1. Теоретичні основи та проблема

Нагадаємо потрібні елементарні поняття. Позначатимемо дугу символом $\rightarrow$. Ребро – це дуга, орієнтація якої може бути невідома або ігнорується. Вершини графа звуться суміжними, якщо вони поєднані ребром, що позначається x–y. *Шлях* суміжності (або просто шлях) у графі – це послідовність дотичних ребер (будь-якої орієнтації) без повторення вершин. Тобто кожна чергова вершина (за винятком останньої) поєднана ребром з наступною вершиною цього шляху. Як виняток, крайні вершини шляху можуть збігатися, і тоді цей шлях називають циклом. *Орпуть* або *оршлях* (тобто строго орієнтований шлях) – це шлях, на якому всі ребра є орієнтовані в напрямку одного й того самого кінця шляху (орієнтовані узгоджено). *Орцикл* – це оршлях, на якому початкова вершина збігається з останньою. Ациклічний орієнтований граф (АОГ) – це орграф, в якому немає орциклів. Надалі терміни граф та орграф слід розуміти як АОГ. Якщо існує дуга $x\rightarrow y$, то вершина x зветься батьком вершини y. Якщо існує оршлях $x\rightarrow\cdots\rightarrow y$, то вершина y зветься нащадком вершини x.

АОГ-модель визначена як (G, ϑ), де G – ациклічний орієнтований граф (де кожній змінній відповідає вершина графа), а ϑ – сукупність локально заданих параметрів. З огляду на взаємно-однозначну відповідність, терміни „змінна” (моделі) та „вершина” (графа) вживаються як взаємозамінні.

Визначення 1. *Колізором* (колайдером, collider) в орграфі зветься фрагмент із двох суміжних дуг вигляду $x\rightarrow y\leftarrow z$. Якщо колізор $x\rightarrow y\leftarrow z$ є частиною шляху π в орграфі, то y називають *колізорною* (*колайдерною*) *вершиною на шляху* π. (Зазначимо, що вершина y одночасно може бути неколізорною на якомусь іншому шляху). *Безколізорний* (безколайдерний) шлях в орграфі – це шлях, який не містить жодного колізора.

Заради лаконічності будемо називати безколізорний шлях *ланцюгом*. (При цьому слід пам'ятати, що це не збігається з поняттям Марковського ланцюга, бо може існувати декілька безколізорних шляхів між двома заданими вершинами.)

Відношення між структурою моделі і фактами умовної незалежності, репрезентованими в АОГ-моделі, формалізовано через критерій *d-сепарації* [1, 2].

Визначення 2 (d-сепарація). Шлях π в АОГ-моделі зветься *d-закритим* (*d-блокованим*) за допомогою (кондиціонування, блокування) множини вершин $\boldsymbol{S}$, якщо і тільки якщо

- існує вершина x, $x\in\pi$, $x\in\boldsymbol{S}$, причому на шляху π лежить дуга $x\rightarrow$ чи $\leftarrow x$, або
- на шляху π лежить хоча б один колізор $\rightarrow y\leftarrow$, з тим, що $y\notin\boldsymbol{S}$ та не існує такої $z\in\boldsymbol{S}$, що є оршлях $y\rightarrow\cdots\rightarrow z$.

Множина вершин $\boldsymbol{S}$ *d-сепарує* вершини x та y ($x,y\notin \boldsymbol{S}$) якщо і тільки якщо всі шляхи між x та y є *d-закритими* за допомогою множини вершин $\boldsymbol{S}$. Будемо позначати таку d-сепарацію предикатом $\mathrm{Ds}(x\bot\boldsymbol{S}\bot y)$. У разі, коли $\boldsymbol{S} = \varnothing$, будемо записувати такий предикат скорочено як $\mathrm{Ds}(x\bot\bot y)$.

Якщо хоча б один шлях між x та y не є d-закритим (тобто є d-відкритим), то говорять, що вершини x та y *d-з'єднані*. Факт d-з'єднання будемо виражати у формі $\neg\mathrm{Ds}(x\bot\boldsymbol{S}\bot y)$.

Визначення 3. Якщо предикат $\mathrm{Ds}(x\bot\boldsymbol{S}\bot y)$ діє в АОГ G, то $\boldsymbol{S}$ зветься *сепаратором* в G для пари вершин (x, y).

Критерій d-сепарації дозволяє проводити аналіз моделі чисто графовим апаратом. Зокрема, безпосередньо з критерію d-сепарації випливає

Правило 1. Якщо в АОГ є лише одна вершина, d-з'єднана з x (залежна від x), тобто множина $\{z|\neg\mathrm{Ds}(z\bot\bot x)\}=\{y\}$, то x–y.

Явно за цих умов немає альтернативного варіанта забезпечити відкритий шлях між вершинами x та y, як тільки з'єднати їх ребром.

Поняття умовної незалежності відоме з теорії ймовірностей і статистики [15]. Відношення умовної незалежності змінних x та y при фіксації значень (кондиціонуванні) набору змінних $\boldsymbol{S}$ будемо виражати формулою $\mathrm{Pr}(x\bot\boldsymbol{S}\bot y)$, де $x,y\notin\boldsymbol{S}$. Для дискретних змінних незалежність $\mathrm{Pr}(x\bot\boldsymbol{S}\bot y)$ означає

$$p(y|\boldsymbol{S},x) = p(y|\boldsymbol{S}). \qquad (1)$$

Для частотних оцінок ймовірностей відповідна рівність виконується в асимптотичному сенсі. У гаусових мережах умовна незалежність проявляється як нульова (з точністю до статистичного гамору) величина коефіцієнта частної кореляції. Безумовна незалежність є просто спеціальним випадком умовної незалежності з порожньою умовою, тобто $\mathrm{Pr}(x\bot\varnothing\bot y)$, або коротко $\mathrm{Pr}(x\bot\bot y)$. Факт безумовної залежності будемо виражати як $\neg\mathrm{Pr}(x\bot\bot y)$.

Мета аналітика – ідентифікувати структуру моделі "індуктивно", тобто вивести її з даних спостережень чи вимірювань (спираючись на кілька методологічних постулатів) [3– 5, 7, 14].

Загальновідомо [1–5]: якщо в АОГ немає ребра x–y, то існує сепаратор для пари вершин (x, y). Зокрема, таким сепаратором є об'єднання батьків вершин x та y. Отже, маємо принцип у формі правила $\{\neg\exists\boldsymbol{S}: \mathrm{Pr}(x\bot\boldsymbol{S}\bot y)\} \Rightarrow (x–y)$.

Відомі методи і алгоритми ідентифікації структури АОГ-моделі можна поділити на два основних підходи – "constraint-based" (або "сепараційний") та "оптимізаційний" (апроксимаційний) [3–5, 7, 14]. Маємо на меті підсилення сепараційних методів. Робота таких методів полягає у пошуку сепараторів для кожної пари змінних. Найвідоміший метод цього підходу – алгоритм 'PC' [3, 4].

Відома теорема АОГ-семантики [16] стверджує, що з факту d-сепарації в АОГ випливає істинність відповідного твердження умовної незалежності, тобто

$$\mathrm{Ds}(x\bot\boldsymbol{S}\bot y) \Rightarrow \mathrm{Pr}(x\bot\boldsymbol{S}\bot y). \qquad (2)$$

Теорема АОГ-семантики забезпечує процедуру зчитування тверджень умовної незалежності з графа АОГ-моделі. Але процес ідентифікації моделі за статистичними даними має рухатись у зворотному напрямку, відштовхуючись від результатів тестування незалежності.

Як методологічну базу для методів ідентифікації АОГ-моделей теоретики прийняли припущення *необманливості* розподілу ймовірностей для АОГ-моделі [3, 4], що можна виразити як

$$\mathrm{Pr}(x\bot\boldsymbol{S}\bot y) \Rightarrow \mathrm{Ds}(x\bot\boldsymbol{S}\bot y). \qquad (3)$$

Зіставляючи (2) та (3), отримуємо

$$\mathrm{Ds}(x\bot\boldsymbol{S}\bot y) \Leftrightarrow \mathrm{Pr}(x\bot\boldsymbol{S}\bot y). \qquad (4)$$

Але беззастережно покладатися на припущення необманливості в повному обсязі не можна (див., зокрема, [7]). Відомі методи й алгоритми сепараційного підходу використовують більш обережні правила висновування. Загальноприйнятим є таке правило ідентифікації відсутності ребра

$$(\exists\boldsymbol{S} : \mathrm{Pr}(x\bot\boldsymbol{S}\bot y)) \Rightarrow \neg(x–y). \qquad (5)$$

Кардинальність умови $\boldsymbol{S}$ визначає складність і надійність тестування умовної незалежності $\Pr(x\perp \boldsymbol{S}\perp y)$. Чим більшою є множина кандидатів у члени сепаратора, тим складнішою є переборна задача пошуку сепаратора. Із зростанням кардинальності $\boldsymbol{S}$ збільшується перелік потрібних статистик і обсяг їх обчислення. Це особливо ускладнює виведення дискретних моделей. Для оптимізації пошуку сепаратора алгоритм ‘PC’ [3–5] обмежує множину кандидатів у члени сепаратора для (x, y) множиною вершин, які вважаються суміжними до x або до y на даному етапі ідентифікації (хоча часто не всі вони є насправді суміжні). Для реальних моделей така тактика не розв'язує всіх проблем. Мабуть найгіршим є те, що алгоритм продовжує шукати сепаратор у багатьох випадках, коли його не існує, і, як наслідок, ризикує припуститися помилки при тестуванні умовної незалежності з великою кардинальністю умови. Тому є потреба знайти нові можливості вдосконалення алгоритмів пошуку. Бажано знаходити сепаратори найменшої кардинальності.

Для пошуку мінімальних сепараторів ми не можемо скористатися відомими (з транспортних задач) методами пошуку розрізу мережі та мінімального потоку з огляду на те, що:

- залежності блокуються маніпуляцією вершин (змінних), а не ребер;
- властивості d-сепарації відмінні від “транспортних аналогів”.

Не можемо скористатися і іншими методами теорії графів, бо на момент пошуку сепаратора структура графа є незавершена або цілком невідома. Проблема полягає не лише в тім, щоб знайти мінімальні сепаратори, але й досягти цього економно, тобто виконати якнайменше тестів (мінімізувати кількість невдалих тестів) та допоміжних обчислень.

2. Властивості членів сепаратора. Формування правил

Доцільно ввести наступні поняття [17, 18].

Визначення 4. Локально-мінімальним сепаратором в АОГ для пари вершин (x, y) зветься такий сепаратор $\boldsymbol{S}$, що після вилучення з $\boldsymbol{S}$ будь-якого його члена (елемента) він перестає бути сепаратором для (x, y). Формально це записується як $\mathrm{Ds}(x\perp \boldsymbol{S}\perp y)$, $\forall z\in \boldsymbol{S}$: $\neg\mathrm{Ds}(x\perp \boldsymbol{S}\backslash\{z\}\perp y)$.

Визначення 5. Назвемо сепаратор $\boldsymbol{S}^*$ для пари вершин (x, y) в АОГ мінімальним сепаратором, якщо для (x, y) не існує сепаратора меншої кардинальності, тобто для всіх інших сепараторів $\boldsymbol{S}$ для пари (x, y) маємо $|\boldsymbol{S}| \geq |\boldsymbol{S}^*|$.

Зауважимо, що може бути декілька мінімальних сепараторів для пари вершин (x, y) в тому самому АОГ. Позначатимемо мінімальний та локально-мінімальний сепаратор для пари вершин (x, y) відповідно як $\mathrm{S_{min}}(x,y)$ та $\mathrm{S_{lom}}(x,y)$.

Поняття локально-мінімального сепаратора важливо також для алгоритмів розмірковування на моделях (в експертних системах) [1, 5]. При виведенні від свідчення x до цільової змінної y інформація має проходити через кожну $z\in \mathrm{S_{lom}}(x,y)$ кожного $\mathrm{S_{lom}}(x,y)$.

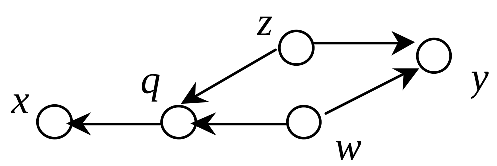

Рис. 2. Приклад моделі, де мінімальний та локально-мінімальний сепаратори не перетинаються

Зауваження 1. Зрозуміло, що кожний мінімальний сепаратор є водночас і локально-мінімальним. Тож всі необхідні вимоги до членів локально-мінімального сепаратора є також необхідними вимогами до членів мінімального сепаратора. Водночас зворотне невірне, тобто не кожний локально-мінімальний сепаратор є мінімальним. Локально-мінімальний сепаратор може не мати жодного спільного члена з мінімальним сепаратором. Приклад показано на рис.2, де $\mathrm{S_{lom}}(x,y) = \{z, w\}$, $\mathrm{S_{min}}(x,y)=\{q\}$.

Твердження 1. Якщо в АОГ діє $\neg\mathrm{Ds}(x\perp\perp y)$ і множина $\boldsymbol{S}$ є локально-мінімальним сепаратором для пари вершин

(x, y), то у складі $\boldsymbol{S}$ існує щонайменше одна якась вершина, що лежить на якомусь ланцюгу між вершинами x та y.

Це випливає з критерію d-сепарації. Тривіально зрозуміло, що між x та y існують ланцюг(и). Якби серед них було ребро x–y, то ніякого $S_{lom}(x,y)$ не існувало б. А оскільки такий сепаратор існує, то всі ланцюги між x та y мають довжину два чи більше ребер. Всі ті ланцюги є закриті за допомогою $\boldsymbol{S}$. Звідси, на кожному з тих ланцюгів лежить якась вершина $z \in \boldsymbol{S}$. □

Слід додати, що це твердження не можна посилити. Тобто можливо, що всі члени $S_{lom}(x,y)$, за винятком одної вершини z, не лежать ні на якому ланцюгу між вершинами x та y.

Твердження 2 (“подвійне 1-відсікання”; ‘double 1-cutting’). Якщо в АОГ існує якась вершина z, що забезпечує $Ds(w \perp z \perp x)$ та $Ds(w \perp z \perp y)$, то вершина w не входить до складу жодного локально-мінімального сепаратора для пари вершин (x, y).

Отже, маємо *правило* 2 (“подвійного 1-відсікання”):

$$[\exists z\text{: } Ds(w \perp z \perp \{x,y\})] \Rightarrow w \notin S_{lom}(x,y).$$

Зауважимо, що правило 2 неявно вбудовано в алгоритм ‘PC’ (як спеціальний випадок).

З огляду на тактику ‘PC’, логічно виникає питання про розширення твердження 2. Перше питання: чи можна виключити вершину w з множини кандидатів у склад сепаратора (зокрема, мінімального) для (x, y) на підставі лише “однобічного 1-відсікання”: $\exists z$: $Ds(w \perp z \perp x)$ або $Ds(w \perp z \perp y)$. Відповідь буде ні. Прикладом є модель, показана на рис. 3, де $Ds(w \perp z \perp x)$, проте $S_{min}(x,y) = \{t, w\}$. Більш того, тут вершина w (як і вершина t) є обов’язковим (незамінним) членом будь-якого сепаратора для (x, y). Ще загальніше: вершина w буде незамінним членом будь-якого сепаратора для (x, y), навіть якщо вбудувати показаний фрагмент графа в контекст оточуючого графа моделі (де можуть з'явитися додаткові шляхи між x та y). Отже, посилити правило 2 зазначеним чином не вдається.

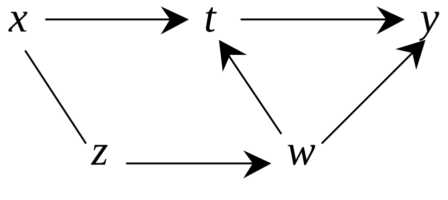

Рис. 3. Модель, де “однобічно 1-відсікається” член мінімального сепаратора

Посилити це правило також не можна і в іншому аспекті, а саме, послабити (розширити) умову так, щоб вершина w відсікалася від вершин x та y не одною й тою вершиною z, а різними. Такий варіант відсіювання кандидатів до складу сепаратора потребує застережень. Аналог твердження 2 буде не вірний, що демонструється наступними прикладами. Тривіальним прикладом є модель $x \leftarrow q \leftarrow w \rightarrow z \rightarrow y$. Інший приклад такого випадку показано на рис. 4.

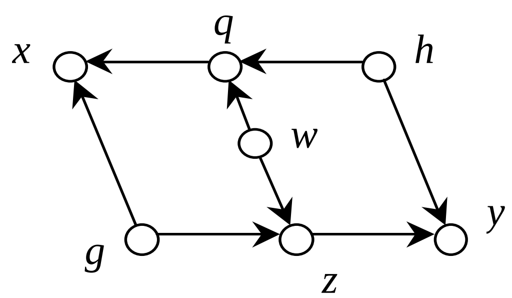

Рис. 4. Модель, що ілюструє випадок двох 1-відсікань (‘pair of 1-cutting’)

Для цієї моделі виконуються сепарації $Ds(w \perp q \perp x)$ та $Ds(w \perp z \perp y)$, тобто розширена версія умови твердження 2. Однак вершина w входить до складу локально-мінімального сепаратора, а саме $S_{lom}(x,y) = \{q, z, w\}$. Тому цей сепаратор не є мінімальним. Існують два мінімальних сепаратора для (x, y): $\{g, q\}$ та $\{h, z\}$, які не містять вершини w. Як бачимо, розширення правила “подвійного 1-відсікання” до правила “двох 1-відсікань” може мати практичний сенс. Але існують інші моделі, де виключення з пошуку сепаратора вершини w за таких умов може призвести до втрати мінімального сепаратора. (Зауважимо, що коли модель має приховані змінні, таке правило може призвести до неможливості знайти існуючий сепаратор. Але модель із прихованими змінними вже виходить за межі класу АОГ).

Підсумувати ці міркування можна таким твердженням про "заміщення" члена сепаратора.

Якщо в АОГ вершина *w* входить до складу якогось локально-мінімального сепаратора для пари вершин (*x*, *y*), та існують якісь дві вершини *q*, *z*, такі, що Ds($w\perp q\perp x$) та Ds($w\perp z\perp y$), то існує якийсь сепаратор для (*x*, *y*), що не містить вершини *w*.

Довести це просто. Дійсно, тривіальним сепаратором для (*x*, *y*) є об'єднання множин батьків вершин *x* та *y*. Зрозуміло, що вершина *w* не входить до цього об'єднання (як несуміжна до *x* та *y*). □

Більш того, це доведення збереже чинність і в разі, якщо відкинути першу частину умови і узагальнити другу частину умови так, що відсікання вершини *w* від *x* та *y* відбувається за допомогою довільних множин вершин. Отже ми обґрунтували

Твердження 3 (правило "двох відсікань"; 'pair of cutting'). Якщо сепаратор для (*x*, *y*) існує й існують множини вершин ***S***, ***R***, такі, що Ds($w\perp \boldsymbol{S}\perp x$) та Ds($w\perp \boldsymbol{R}\perp y$), то існує якийсь сепаратор для (*x*, *y*), що не містить вершини *w*.

Виключення з пошуку сепаратора тих несуміжних вершин, які відсікаються від *x* чи *y* за допомогою якихось третіх вершин (1-сепараторами) є більш виправдане, ніж коли відсікання досягається складними сепараторами. Подаємо випадок, коли виключення вершини зі списку потенційних членів сепаратора на підставі того, як вона відсікається від *x* та *y* за допомогою 2-сепараторів, веде до втрати мінімального сепаратора. Це ілюструє модель на рис. 5, де маємо Ds($w\perp\{g,h\}\perp x$) та Ds($w\perp\{q,z\}\perp y$). Проте $S_{min}(x,y)=\{w, r, t\}$, й інших мінімальних сепараторів немає. В таких ситуаціях алгоритм 'PC' не знайде мінімального сепаратора для (*x*, *y*), але має знайти {*g*, *h*, *r*, *t*} або {*q*, *z*, *r*, *t*}. Та незважаючи на те, тактика алгоритма 'PC' в багатьох випадках виправдовується міркуванням скорочення множини кандидатів у сепаратор (а це спрощує перебір).

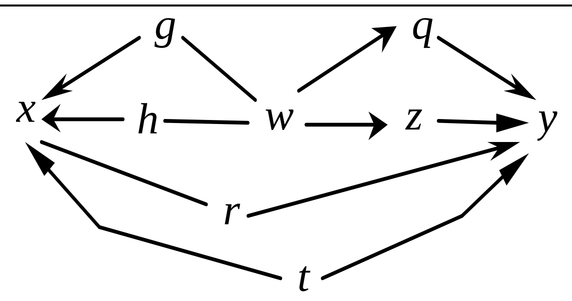

Рис.5. Модель, де мінімальний сепаратор для (*x*, *y*) включає вершину, несуміжну до *x* та *y*.

Твердження 4 (базова теорема про членів локально-мінімального сепаратора). Якщо в АОГ вершина *z* входить до складу якогось локально-мінімального сепаратора для пари вершин (*x*, *y*), то:

а) існує якийсь оршлях ρ від вершини *z* до вершини *x* або *y*;

б) якщо не існує жодного ланцюга між *x* та *z*, який не проходить через *y*, то мають існувати не менше як два якісь ланцюги λ_1 та λ_2 між *z* та *y*, які не проходять через *x* та закінчуються дугами $\rightarrow y$; (відтинки ланцюгів λ_1 та λ_2, що прилягають до вершини *y*, можуть взаємно накладатися).

Це уточнення твердження з [17].

Доведення. Звертаємось до (а). Позначимо символом ***S*** локально-мінімальний сепаратор для (*x*, *y*), членом якого є *z*. Вершина *z* закриває якийсь шлях τ між *x* та *y*, причому *z* є неколізорною вершиною на шляху τ. (Це негайно випливає з критерію d-сепарації та визначення локально-мінімального сепаратора). Доведення будується за ітеративною схемою. Якщо вершина *z* лежить на деякому ланцюгу між вершинами *x* та *y*, то негайно отримуємо бажане. (Якщо ланцюг розрізати на частини, то не менше ніж одна частина буде оршляхом). Інакше, нехай вершина *z* не лежить ні на якому ланцюгу між вершинами *x* та *y*. Розглянемо шлях τ, який закривається за допомогою *z*. Будемо просуватися по τ, розпочавши з дуги $z\rightarrow$. Зрозуміло, що пройшовши певний оршлях, ми або дістанемось вершини *x* чи *y* (і тоді доведення закінчено), або дійдемо до колізора. Нехай цей найближчий до вершини *z* колізор на цій частині шляху τ буде $\rightarrow q_1\leftarrow$. Оскільки *z* є членом локально-мінімального сепаратора ***S*** і закриває шлях

τ, то колізор $\rightarrow q_1 \leftarrow$ є d-відкритим при кондиціонуванні множини вершин $\boldsymbol{S}\backslash\{z\}$. Тож або маємо $q_1 \in \boldsymbol{S}$, або $\boldsymbol{S}$ включає якогось нащадка вершини q_1, скажімо t_1 (це найближчий такий нащадок). Відповідно маємо оршлях $z \rightarrow \cdots \rightarrow q_1$ або $z \rightarrow \cdots \rightarrow q_1 \cdots \rightarrow t_1$. Тепер розглядаємо відповідно q_1 або t_1 як члена локально-мінімального сепаратора аналогічно тому, як ми розглядали вершину z. Повторюючи викладені міркування ітеративно, будемо продовжувати оршлях від z до чергового члена q_2 або t_2 множини вершин $\boldsymbol{S}$ і так далі. Оскільки множина $\boldsymbol{S}$ є скінченою, а повторне попадання в ту саму вершину неможливо (строго орієнтований цикл неможливий за визначенням), то ми неодмінно дійдемо по оршляху від z до вершини x або y. Отримали бажане (а). □

Переходимо до пункту (б). Вершина z закриває якийсь шлях між x та y; якщо ми попрямуємо по цьому шляху від z в бік вершини x, то дійдемо до якогось відкритого колізора $\rightarrow q \leftarrow$. Зрозуміло, вершина q або якийсь її нащадок входять до складу $\boldsymbol{S}$. Тож, згідно пункту (а) твердження, існує оршлях $q \rightarrow \cdots \rightarrow y$. (Якщо уявити оршлях $q \rightarrow \cdots \rightarrow x$, то отримаємо ланцюг $z \cdots \rightarrow q \rightarrow \cdots \rightarrow x$, що суперечить умові (б)). Отже, маємо перший потрібний ланцюг λ_1: $z - \cdots \rightarrow q \rightarrow \cdots \rightarrow y$. Нагадаємо, що вершина z закриває якийсь шлях між x та y, отож існує шлях λ_{zy} від z в сторону вершини y, який розпочинається з іншої дуги, ніж та, яка веде в сторону вершини x. Якщо ми почнемо рухатись по цьому шляху λ_{zy} від z в сторону вершини y, то або дійдемо до y, не зустрівши колізора (що завершує доведення), або зустрінемо колізор $\rightarrow w \leftarrow$. Оскільки цей колізор – відкритий за кондиціонування $\boldsymbol{S}$, то, аналогічно вищесказаному, має бути оршлях $w \rightarrow \cdots \rightarrow y$. Таким чином, отримуємо другий потрібний ланцюг λ_2: $z - \cdots \rightarrow w \rightarrow \cdots \rightarrow y$. □

Зауваження 2 (до твердження 4). Якщо в АОГ маємо $z, w \in \mathrm{S_{lom}}(x,y)$, то не обов'язково існують оршляхи з вершин z та w до одної й тої самої вершини з пари (x, y). Наприклад, для АОГ на рис. 4 маємо $\mathrm{S_{lom}}(x,y) = \{q, z, w\}$, і існує оршлях $q \rightarrow x$, але не існує оршляху від z до x. Але існує $z \rightarrow y$.

Для підвищення ефективності пошуку сепараторів треба інтенсивно скорочувати кількість кандидатів до складу сепаратора. Інтуїтивно видається, що мають бути ще якісь ознаки, за якими можна розпізнати вершини, які не є членами потрібного локально-мінімального сепаратора.

Зокрема, можна висунути таку гіпотезу: якщо одна вершина із пари (x, y) сепарує вершину z від іншої вершини пари (x, y), то вершина z (правдоподібно) не є членом ніякого $\mathrm{S_{lom}}(x,y)$. За цією гіпотезою стоїть інтуїтивна здогадка, яку можна висловити так: перша вершина "відсторонює" третю вершину від другої вершини. Тому третя вершина не лежить "поміж" першою та другою, і отже, правдоподібно не є членом локально-мінімального сепаратора для першої та другої вершини. Це можна сформулювати як таке правило "відсторонення" кандидатів у сепаратор:

$$\mathrm{Ds}(y \perp x \perp z) \Rightarrow z \notin \mathrm{S_{lom}}(x,y).$$

Отже, формулюємо наступне твердження [18].

Твердження 5 ("відсторонення" кандидатів у сепаратор; 'placing aside'). Якщо в АОГ-моделі вірне $\mathrm{Ds}(z \perp x \perp y) \& \neg \mathrm{Ds}(z \perp\perp y)$, то вершина z не є членом ніякого локально-мінімального сепаратора для пари вершин (x, y).

Доведення. Нехай в АОГ вірне $\neg \mathrm{Ds}(x \perp\perp y)$, $\mathrm{Ds}(z \perp x \perp y)$, $\neg \mathrm{Ds}(z \perp\perp y)$, і припустимо $z \in \mathrm{S_{lom}}(x,y)$. З огляду на $\mathrm{Ds}(z \perp x \perp y)$ усі шляхи між z та y, які не проходять через x, є колізорні. Тож не існує ніякого оршляху $z \rightarrow \cdots \rightarrow y$. Звідси, згідно твердження 4 (а) та факту $z \in \mathrm{S_{lom}}(x,y)$, існує оршлях $z \rightarrow \cdots \rightarrow x$. Тому, всі ланцюги між z та x мають кінцеву дугу $\rightarrow x$ (заборона орциклів). Водночас $\neg \mathrm{Ds}(z \perp\perp y)$ та $\mathrm{Ds}(z \perp x \perp y)$ означає, що існують ланцюги між z та y, і то всі вони проходять через x. Розглянемо відтинки цих шляхів від вершини x до вершини y. Кожен такий відтинок має першу дугу $x \rightarrow \cdots$, бо інакше $\neg \mathrm{Ds}(z \perp\perp y)$ є неможливе. (Вище з'ясовано, що всі ланцюги між z та x

мають дугу $\rightarrow x$). Отож, існує оршлях (оршляхи) $x\rightarrow\cdots\rightarrow y$.

Візьмемо локально-мінімальний сепаратор для (x, y), до якого належить z, та позначимо його $\boldsymbol{S}_z$. Тоді множина $\boldsymbol{S}_z\backslash\{z\}$ не є локально-мінімальним сепаратором для (x, y). Тобто за блокування вершин $\boldsymbol{S}_z\backslash\{z\}$ існує відкритий шлях τ між вершинами x та y, який проходить через z (згідно d-сепарації). Вище з'ясовано, що всі ланцюги між z та y проходять через x. Отож, існує колізорний шлях τ_2 між z та y, який є частиною шляху τ (тобто не проходить через x) і який відкривається при кондиціонуванні вершин $\boldsymbol{S}_z\backslash\{z\}$.

Таким чином, аби був відкритий шлях τ між вершинами x та y через z за кондиціонування вершин $\boldsymbol{S}_z\backslash\{z\}$, необхідно, щоб кондиціонування вершин $\boldsymbol{S}_z\backslash\{z\}$ відкривало всі колізори q_i на відтинку τ_2 цього шляху між z та y.

Отож, (згідно d-сепарації) для кожного такого колізора q_i маємо $q_i\in\boldsymbol{S}_z$ або має існувати оршлях $q_i\rightarrow\cdots\rightarrow w_i$, де $w_i\in\boldsymbol{S}_z$. Візьмемо колізор q_1, найближчий до вершини z. Оскільки цей колізор q_1 є відкритий при кондиціонуванні вершин $\boldsymbol{S}_z\backslash\{z\}$, то існують $w_1\in\boldsymbol{S}_z$ та $q_1\rightarrow\cdots\rightarrow w_1$. (Зокрема, можливо $q_1\equiv w_1$). Згідно твердження 4 (а), існує $w_1\rightarrow\cdots\rightarrow x$ або $w_1\rightarrow\cdots\rightarrow y$. Проте випадок $w_1\rightarrow\cdots\rightarrow y$ є неможливим з огляду на те, що тоді конкатенація шляхів $z-\cdots-q_1\rightarrow\cdots\rightarrow w_1$ та $w_1\rightarrow\cdots\rightarrow y$ дає ланцюг між вершинами z та y, який не проходить через x (а неможливість цього показана вище). Отже, має бути оршлях $w_1\rightarrow\cdots\rightarrow x$. Звідси випливає, що кондиціонування вершини x відкриває колізор $\rightarrow q_1\leftarrow$ на шляху τ_2 між z та y. Водночас за кондиціонування вершин $\boldsymbol{S}_z\backslash\{z\}$ є відкритим шлях τ та його відтинок між вершинами q_1 та y. Отже, на цьому відтинку зустрінемо наступний колізор $\rightarrow q_2\leftarrow$, який є відкритий за кондиціонування вершин $\boldsymbol{S}_z\backslash\{z\}$, тобто існує оршлях $q_2\rightarrow\cdots\rightarrow w_2$, де $w_2\in\boldsymbol{S}_z$. Повторюючи викладені міркування рекурсивно, ми прийдемо в останній на шляху τ_2 колізор $\rightarrow q_k\leftarrow$, такий, що існує оршлях $q_k\rightarrow\cdots\rightarrow w_k$, де $w_k\in\boldsymbol{S}_z$ й існує оршлях $w_k\rightarrow\cdots\rightarrow x$. Таким чином, відстеживши названі шляхи, отримуємо шлях $z\rightarrow\cdots\rightarrow x\leftarrow\cdots\leftarrow w_k\leftarrow\cdots\leftarrow q_k\leftarrow-\cdots-y$, на якому є лише один колізор $\rightarrow x\leftarrow$. Отже, має бути $\neg\mathrm{Ds}(z\perp x\perp y)$, що суперечить умові твердження. □

Доведення пояснюється на рис. 6.

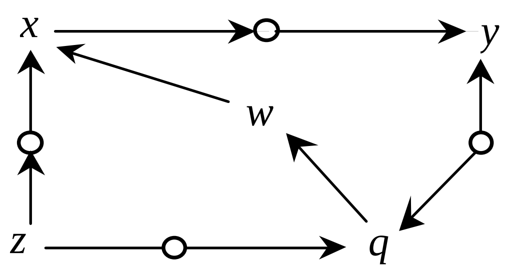

Рис.6. Ілюстрація доведення твердження 5 (‘placing aside’)

Отже, твердження 5 обґрунтовує коректність *правила* 3 (“відсторонення”):

$$[\mathrm{Ds}(z\perp x\perp y)\&\neg\mathrm{Ds}(z\perp\perp y)]\Rightarrow z\notin \mathrm{S}_{\mathrm{lom}}(x,y).$$

Зрозуміло, що змінні x та y в антецеденті можна міняти місцями.

Це правило здатне скоротити кількість кандидатів до складу сепаратора і, отже, створює передумови підвищення надійності та зменшення переборної і обчислювальної складності методів ідентифікації структур АОГ-моделей. Більше того, в деяких ситуаціях воно дозволяє вичерпно ідентифікувати ребро, тобто взагалі припинити пошук сепаратора для даної пари вершин, з'ясувавши, що сепаратора не існує. Дійсно, якщо всі потенційні кандидати до сепаратора “відсторонюються”, то можна зробити висновок, що сепаратора для цієї пари не існує.

Отже, з тверджень 1 та 5 випливає *правило* 4 (“швидкої” ідентифікації ребра), яке формально записується як

$$\neg\mathrm{Ds}(x\perp\perp y)\&\{\forall z\in \boldsymbol{U}\backslash\{x,y\}:[\mathrm{Ds}(z\perp\perp x)\vee \vee\mathrm{Ds}(z\perp\perp y)\vee\mathrm{Ds}(z\perp x\perp y)\vee\mathrm{Ds}(z\perp y\perp x)]\}\Rightarrow \Rightarrow x-y,$$

де $\boldsymbol{U}$ – множина всіх вершин графа; символ “$\vee$” – означає диз’юнкцію.

Твердження 6. Нехай $\neg\mathrm{Ds}(x\perp\perp y)$ та маємо

$\boldsymbol{V} = \{v| \neg \mathrm{Ds}(v \perp\perp x) \vee \neg \mathrm{Ds}(v \perp\perp y)\}$,

$\boldsymbol{Q} = \{q| \mathrm{Ds}(q \perp x \perp y) \& \neg \mathrm{Ds}(q \perp\perp y)\}$,

$\boldsymbol{R} = \{r| \mathrm{Ds}(r \perp y \perp x) \& \neg \mathrm{Ds}(r \perp\perp x)\}$,

та множина $\boldsymbol{V} \backslash (\boldsymbol{Q} \cup \boldsymbol{R} \cup \{x,y\})$ складається з єдиного елемента z. Тоді, якщо $\neg \mathrm{Ds}(x \perp z \perp y)$, то існує ребро $x–y$.

Випливає з тверджень 1, 4 та 5.

Це твердження дає правило 5 ("ідентифікації ребра за наявності єдиного не відстороненого"). (Не записуємо його окремо з огляду на громіздкість.)

Твердження 7. Нехай в АОГ вірне $\neg \mathrm{Ds}(x \perp\perp y)$ та маємо відсторонення $\mathrm{Ds}(\boldsymbol{Q} \perp x \perp y)$ та $\mathrm{Ds}(x \perp y \perp \boldsymbol{R})$. Нехай буде $\boldsymbol{M} = \{z| \neg \mathrm{Ds}(z \perp\perp x) \& \neg \mathrm{Ds}(z \perp\perp y)\}$.Тоді, якщо $\boldsymbol{M} \backslash [(\boldsymbol{Q} \cup \boldsymbol{R}) \cap \boldsymbol{M} \cup \{x,y\}] = \varnothing$, то існує $x–y$.

Випливає з тверджень 1 та 5.

Це твердження дає правило 6 ("ідентифікації ребра за відсутності посередника").

Твердження 8. Нехай $\neg \mathrm{Ds}(x \perp\perp y)$ й існує сепаратор для пари вершин (x, y), і нехай $\boldsymbol{Q}$ – це множина всіх вершин q, для яких вірне $\mathrm{Ds}(q \perp x \perp y) \vee \mathrm{Ds}(q \perp y \perp x) \vee \mathrm{Ds}(q \perp\perp x) \vee \vee \mathrm{Ds}(q \perp\perp y)$. Тоді, якщо множина $\boldsymbol{U} \backslash (\boldsymbol{Q} \cup \{x,y\})$ (де $\boldsymbol{U}$ – множина всіх вершин графа) складається з єдиного елемента z, то z є обов'язковим членом всіх сепараторів для пари вершин (x, y).

Доведення. За зазначених умов має бути ланцюг $x–z–y$. Якби замість нього був будь-який довший ланцюг $x–q–\cdots–z–\cdots–y$, то вершина q теж була б членом множини $\boldsymbol{U} \backslash (\boldsymbol{Q} \cup \{x,y\})$. Якби було $x \rightarrow z \leftarrow y$, то з факту існування сепаратора для (x, y) випливає (див. твердження 1) існування якогось ланцюга з двох чи більше ребер, який не проходить через вершину z (заборона орциклів). Отож, той ланцюг проходив би через іншу вершину t, тобто має існувати $x–\cdots–t–\cdots–y$. Тоді вершина t теж була б членом множини $\boldsymbol{U} \backslash (\boldsymbol{Q} \cup \{x,y\})$. Отож, з заперечення існування ланцюга $x–z–y$ випливає протиріччя умові.

3. Застосування правил у виведенні структур АОГ-моделей

Деякі з наведених тверджень можуть здатися тривіальними. Ймовірно, може спасти на думку, що з'ясувати значення деяких (графових) предикатів в лівих частинах запропонованих правил буде складніше, ніж прямо знайти потрібний сепаратор. Проте сумніви мають розсіятись, щойно ми переходимо від аналіза графа до аналіза даних. Запропоновані правила призначені головним чином для ідентифікації моделі (коли структура графа ще невідома). При цьому графові предикати замінюються їх емпіричними "двійниками" згідно (4), та "обчислюються" за допомогою виконання тестів на відбірці даних. Наприклад, емпіричний факт $\neg \mathrm{Pr}(x \perp\perp y)$ беззастережно свідчить про існування ланцюга або ребра $x–y$. Але при малих відбірках даних багато свідчень у формі емпіричних фактів умовної незалежності є вразливими до відбіркових ухилень (і то у різній мірі для різних форматів тестів), тому мають бути застосовані з обачністю. Зокрема, емпіричний факт $\mathrm{Pr}(x \perp\perp y)$ можна трактувати як робастне (для реалістичних розмірів відбірки даних) свідчення відсутності ребра $x–y$, але не як свідчення відсутності ланцюга [7].

Робастним є свідчення у формі $\mathrm{Pr}(x \perp y \perp z) \& \neg \mathrm{Pr}(x \perp\perp z)$, тож робастною є емпірична версія правила 3 (відсторонення). Справді, це свідчення означає, що після закриття шляху, який йде через вершину y, буде: або між змінними x та z вже немає відкритих шляхів, або такі шляхи існують, проте всі вони слабкі. Але якщо такий шлях $x–\cdots–z$ виявився слабким, то подовжений (протягнутий далі) шлях $x–\cdots–z–\cdots–y$ буде ще слабкішим. І тому виглядає неправдоподібним, що змінна z – необхідна для емпіричної сепарації $\mathrm{Pr}(x \perp \boldsymbol{S} \perp y)$. Зрозуміло, що це не строге доведення, а лише евристичне міркування. (Треба враховувати різні умови відповідних незалежностей). Та все ж попередній аналіз підтверджує практичну надійність емпіричної версії правила відсторонення.

У процесі ідентифікації моделі визначаються фрагменти структури графа та окремі графові факти. Деякі факти можуть бути відомі апріорі. Тому корисно застосовувати комбінації графових та емпіричних свідчень. Але при цьому треба дотримуватись певних застережень. Більшість відомих алгоритмів виведення структури з даних спочатку ідентифікує всі ребра, а вже потім орієнтує їх. Наявність шляху між *x* та *y* в неорієнтованому графі не є підставою для твердження про існування ланцюга між *x* та *y*. Для вилучення коректних графових фактів з незавершеного графа треба враховувати стратегію алгоритма виведення структури моделі. Якщо алгоритм стартує з повного графа і послідовно видаляє зайві ребра (як це робить алгоритм ‘PC’), то можна зчитувати з поточного графа свідоцтва про модель у формі Ds($x \perp\!\!\!\perp y$), але забороняється зчитувати факти у формі ¬Ds($x \perp\!\!\!\perp y$). Якщо ж алгоритм стартує з пустого графа і послідовно вставляє необхідні ребра, то все буде навпаки: зчитування фактів ¬Ds($x \perp\!\!\!\perp y$) буде коректним для моделі, а фактів Ds($x \perp\!\!\!\perp y$) – некоректним.

Висновки

Запропоновано і обґрунтовано твердження та правила, які дозволяють адаптивно оптимізувати пошук сепараторів у процесі виведення структур АОГ-моделей (зокрема, баєсівських мереж та гібридних моделей з будь-якою формою параметризації) з статистичних даних. Ці правила:

- забезпечують виявлення мінімальних сепараторів;
- сприятимуть більш ранньому скороченню списка кандидатів у сепаратор і зменшенню кількості необхідних тестів;
- спрощують і прискорюють ідентифікацію деяких ребер (або їх відсутності);
- підтримують використання апріорних знань про структуру.

Отже, ми показали нові можливості і техніку підвищення ефективності і надійності методів та алгоритмів індуктивного відтворення структур ймовірнісних моделей залежностей.

1. *Pearl J.* Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. – San Mateo: Morgan Kaufmann, 1988.
2. *Pearl J.* Causality: Models, Reasoning, and Inference. – Cambridge Univ. Press, 2000. – 526 p.
3. *Spirtes P., Glymour C., Scheines R.* Causation, prediction and search. 2-nd Ed., New York: MIT Press, 2001. – 496 p.
4. *Scheines R., Spirtes P., Glymour C., Meek C., Richardson T.* The TETRAD Project: Constraint Based Aids to Causal Model Specification // Multivariate Behavioral Research, 1998. – 33 (1) – P. 65 – 118.
5. *Neapolitan R.E.* Learning Bayesian Networks. – Prentice Hall, 2003. – 674 p.
6. *Андон Ф.И., Балабанов А.С.* Структурные статистические модели: инструмент познания и моделирования // Системні дослідження та інформаційні технології. – 2007. – № 1. – С. 79 – 98.
7. *Балабанов А.С.* К выводу структур моделей вероятностных зависимостей из статистических данных // Кибернетика и системный анализ. – 2005. – № 5. – С. 19 – 31.
8. *Балабанов А.С.* Выделение знаний из баз данных: передовые компьютерные технологии интеллектуального анализа данных // Математические машины и системы. – 2001. – № 1/2. – С. 40 – 54.
9. *Burnell L., Horvitz E.* Structure and chance: melding logic and probability for software debugging // Communications of the ACM. – 1995 – **38**, Issue 3. – P. 31 – 41.
10. *Nefian A., Liang L., Pi X., Liu X., Murphy K.* Dynamic Bayesian Networks for Audio-Visual Speech Recognition // EURASIP, J. of Applied Signal Processing. – 2002. – **11** – P. 1 – 15.
11. *Using* Bayesian Network to Analyze Expression Data / N. Friedman, M. Linial, I. Nachman, and D. Pe'er // J. Computational Biology. – 2000. – **7** – P. 601 – 620.
12. *Kennett R.J., Korb K.B., Nicholson A.E.* Seabreeze Prediction Using Bayesian Networks // Lecture Notes in Computer Science, Vol. 2035/2001. –Berlin/Heidelberg: Springer: 2001.
13. *Bessler D.A.* On World Poverty: Its Causes and Effects. – Food and Agricultural Organization (FAO) of the United Nations. – Research Bulletin. – Rome. – 2003. – 50 p.

14. *Chickering D.M., Meek C., Heckerman D.* Large-Sample Learning of Bayesian Networks is NP-Hard. / In: Proceedings of 19th Conf. on Uncertainty in Artificial Intelligence. Acapulco, Mexico. – Morgan Kaufmann. – 2003. – P. 124 – 133.
15. *Dawid A.P*. Conditional independence in statistical theory (with discussion) // J. of Roy. Statist. Soc. – 1979. – **41-B**. – P. 1 – 31.
16. *Verma T., Pearl J.* Causal Networks: Semantics and Expressiveness. / in: R. Shachter, T.S. Levitt, L.N. Kanal (Eds.), Uncertainty in Artificial Intelligence. – 4 – Elsevier Science Publishers, North-Holland, Amsterdam, 1990. – P. 69 – 76.
17. *Балабанов А.С.* Восстановление структур систем вероятностных зависимостей из данных. Аппарат генотипов переменных // Проблемы управления и информатики. – 2003. – № 2. – С. 91 – 99.
18. *Балабанов О.С.* Дослідження шляхів підвищення обчислювальної ефективності методів ідентифікації моделей залежностей. – Звіт з НДР, Шифр 3/01К.02-02. – К., Інститут програмних систем НАН України, 2005. – 36 с.

Отримано 19.09.2007

***Про автора*:**

Балабанов Олександр Степанович,
старший науковий співробітник,
кандидат технічних наук.

***Місце роботи*:**

Інститут програмних систем
НАН України, 03187, Київ-187,
проспект Академіка Глушкова, 40.
факс: (38044) 526 6263 ;
Тел.: (044) 526 6249.
E-mail: bas@isofts.kiev.ua