

Лабораторна робота №6

«ОСНОВИ РОБОТИ З БАЙЄСІВСЬКИМИ МЕРЕЖАМИ ДОВІРИ В СИСТЕМІ HUGIN»

Мета роботи: вивчення побудови експертних систем на основі байєсівських мереж довіри (БМД), методами їх реалізації і прийомами роботи з ними в системі HUGIN.

Короткі теоретичні відомості

1. Основи побудови найпростішої байєсівської мережі довіри

Розглянемо приклад побудови найпростішої експертної системи, що дозволяє оцінювати стан плодкових дерев. Досвідчений садівник, яка професійно володіє знаннями з фізіології рослин одного разу виявив, що його улюблене яблучне дерево втратило листя і він хоче з'ясувати, чому це сталося. Будучи фахівцем у своїй галузі (експертом), він прагнути змодельовати ситуацію, шляхом побудови експертної системи, яка в нагоді б і іншим менш досвідченим садівникам. Він знає, що:

ЯКЩО	"Дерево засихає при нестачі вологи"
АБО	"Дерево хворіє"
ТО	"Листя опадає".

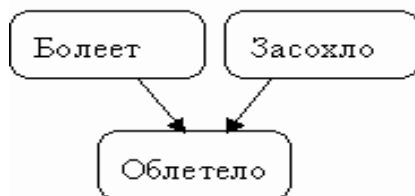


Рисунок 6.1 - Графічне представлення БМД

Дана ситуація може бути змодельована байєсівською мережею довіри [3], що містить три вершини подій: "Хворіє", "засохли" і "Облетіло". При цьому кожна подія може приймати лише одне з двох можливих станів:

- "Хворіє" - стану 'хворіє' або 'ні';
- "Засохло" - стану 'засохло' або 'ні';

– "Облетіло" - стану 'облетіло' або 'ні'.

Байєсівська мережа довіри, наведена на рис.6.1, моделює той факт, що є причинно-наслідкова залежність від вершини "Хворіє" до вершини "Облетіло" і від вершини "засохли" до вершини "Облетіло". Це відображено стрілками на БМД. Коли є причинно-наслідкова залежність від вершини А до іншої вершини В, то ми очікуємо, що коли А знаходиться в деякому певному стані, це впливає на стан В. Слід бути уважним, коли моделюється залежність в БМД. Іноді зовсім не очевидно, який напрямок повинен мати зв'язок. Наприклад, в розглянутому прикладі, ми говоримо, що є залежність від "Хворіє" до "Облетіло", так як коли дерево хворіє, це може викликати опадання його листя. Опадання листя є наслідком хвороби, а не хвороба - наслідком опадання листя.

На рис.6.1 дано графічне представлення БМД. Однак, це тільки якісне уявлення. Перед тим, як назвати це повністю БМД необхідно визначити кількісне уявлення, тобто множину таблиць апріорних і умовних ймовірностей (табл. 6.1 – 6.3). Зауважимо, що всі три таблиці показують імовірність перебування деякої вершини в певному стані, обумовленим станом її батьківських вершин. Але так як вершини "Хворіє" і "засохли" не мають батьківських вершин, то їх ймовірності є маргінальними, тобто не залежить (не залежать від) ні від чого. На такому прикладі ми розглянули, що і як описується дуже простою БМД.

Таблиця 6.1 - Апріорні та умовні ймовірності для стану «Хворіє»

P (Хворіє)	
Хворіє = 'хворіє'	Хворіє = «ні»
0.1	0.9

Таблиця 6.2 - Апріорні та умовні ймовірності для стану «Засохло»

P (засохло)	
Засохло = 'засохло'	Засохло = "ні"
0.1	0.9

Таблиця 6.3 - Априорні та умовні ймовірності для стану «Облетіло»

P (Облетіло Хворіє, Засохло)				
	Засохло = 'засохло'		Засохло = "ні"	
	Хворіє = 'хворіє'	Хворіє = «ні»	Хворіє = 'хворіє'	Хворіє = «ні»
Облетіло= "так"	0.95	0.85	0.90	0.02
Облетіло = "ні"	0.05	0.15	0.10	0.98

Сучасні програмні засоби (такі як MSBN, HUGIN та інші) забезпечують інструментарій для побудови таких мереж, а також можливість використання БМД для введення нових свідочств і отримання рішення (висновку) за рахунок перерахунку ймовірностей у всіх вершинах, які відповідають цим свідченням.

При реалізації наведеної БМД в середовищі HUGIN слід мати на увазі, що дана система не у всіх режимах підтримує роботу з кіріличними шрифтами. Тому, в подальшому, будемо користуватися позначеннями вершин англійською мовою. При цьому вихідна БМД має вигляд, наведений на рис.6.2



Рисунок 6.2 - Графічне представлення БМД в середовищі HUGIN

Порядок виконання роботи

1. Завантаження системи HUGIN

Після запуску HUGIN Runtime, відкриється вікно системи HUGIN, яке містить панель меню, панель інструментів (tool bar pane), а також панель редагування вершин (node edit pane) і панель графічного відображення мережі (network pane). Новий порожній мережу довіри отримує ім'я "unnamed1" і автоматично відкривається у вікні мережі (network window).

Після запуску системи HUGIN (рис. 6.3) вона автоматично переводиться в режим редагування (edit mode), що дозволяє негайно приступити до побудови нової БМД і визначення стану її вершин.

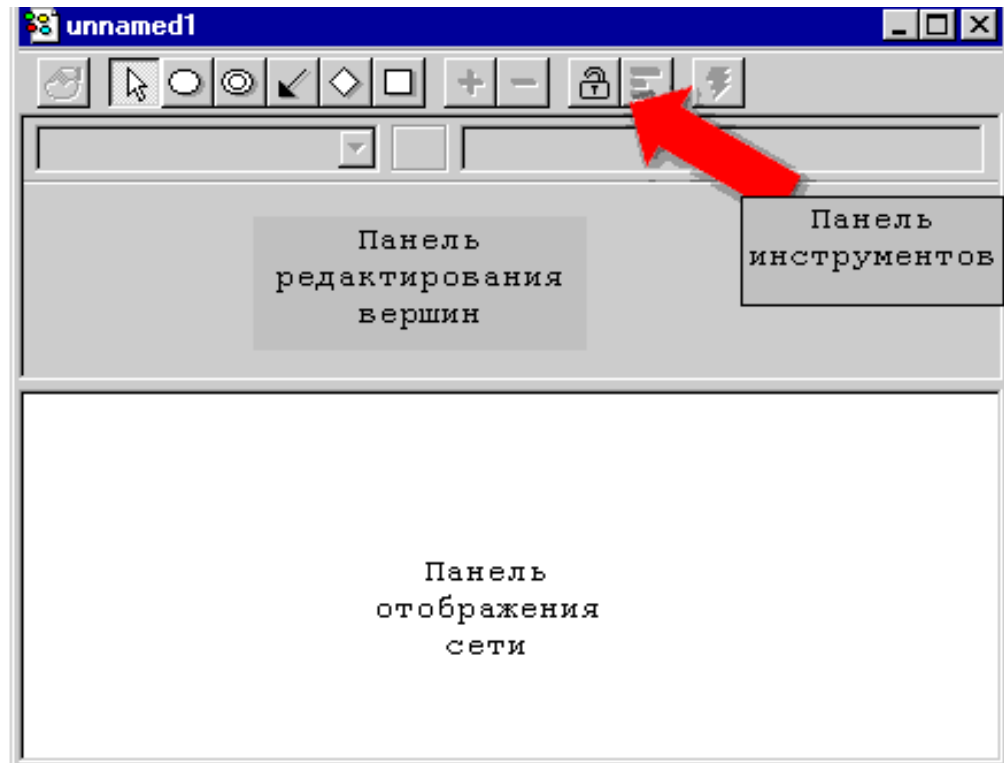


Рисунок 6.3 - Вид користувальницького інтерфейсу системи HUGIN

Іншим важливим режимом роботи системи є режим виконання (run mode), який дозволяє використовувати БМД для отримання необхідних результатів.

2. Додавання нових вершин до БМД

Побудова БМД починається з визначення окремих вершин, що входять в проєктовану БМД. При цьому для створення вершин з дискретними станами, визначення їх властивостей і завдання причинно - наслідкових зв'язків між вершинами використовуються, наведені на рис. 6.4, інструменти інтерфейсу системи, які активуються шляхом натискання лівої кнопки мишки на їх піктограмах.




-  - вершини з дескретними станами [discrete chance tool]
-  - властивості вершин [node properties tool]
-  - додавання зв'язків [casual arrow tool]

Рисунок 6.4 - Піктограми інструментів призначеного для користувача інтерфейсу системи HUGIN

Для побудови даного прикладу БМД, перше, що необхідно зробити - це створити вершину Sick ("Хворіє"). Для цього необхідно:

- Встановити режим додавання вершин з дискретними станами, вибравши відповідну піктограму в панелі інструментів (рис. 6.4).
- Клацнути мишкою в будь-якому місці панелі відображення мережі (рис. 6.3), де передбачається розміщення додається вершини.

Після клацання в панелі відображення мережі повинна відобразитися у вигляді овалу вершина, назва якої за замовчуванням буде C1. Відповідно до нашої моделлю, необхідно цій вершині привласнити назву Sick. Для цього:

- Виділіть курсором миші потрібну вершину.
- Встановіть режим визначення властивостей вершини [Node Properties], клацнувши на відповідній піктограмі панелі інструментів (рис. 6.4).
- Змініть вміст полів Name і Label на Sick.
- Натисніть "ОК".

Name- це внутрішнє ім'я вершини, поки Label є назвою вершини. Якщо поле Label не задано (як це було до того моменту, поки ми не змінили поле Label), то його значення відповідає внутрішньому імені. Внутрішнє ім'я може містити символи 'a' - 'z' і 'A' - 'Z', цифри "0" - "9", а також знак підкреслення '_'.

Вершини Dry ("засохли") і Loses ("Облетіло") додаються в БМД таким же чином. Можна додавати вершини не натискаючи кожного разу на піктограму вершин з дискретними станами, а одного разу вибравши режим додавання і утримуючи клавішу SHIFT клацнути мишею у вікні відображення мережі стільки раз, скільки вершин з необхідно додати в БМД.

3. Встановлення причинно-наслідкові зв'язки між вершинами БМД

На поточний момент проєктована БМД має вигляд, схожий на той, що наведений на рис. 6.5. Наступний етап проєктування БМД полягає у встановленні причинно - наслідкових зв'язків між подіями. Ці зв'язки в моделі БМД відображаються у вигляді стрілок, що з'єднують між собою вершини БМД. Для додавання стрілок від вершини Sick до вершини Loses і від Dry до Loses, необхідно зробити наступне:

- Натисніть іконку додавання зв'язків (рис. 6.4).
- Протягніть мишею стрілку від Sick до Loses, натиснувши ліву кнопку і утримуючи клавішу SHIFT.
- Протягніть мишею стрілку від Dry до Loses натиснувши ліву клавішу.

Тепер можливо отримати повне якісне уявлення, подібне зображеному на рис. 6.2. Наступний крок - це завдання станів і таблиць умовних ймовірностей для кожної вершини.

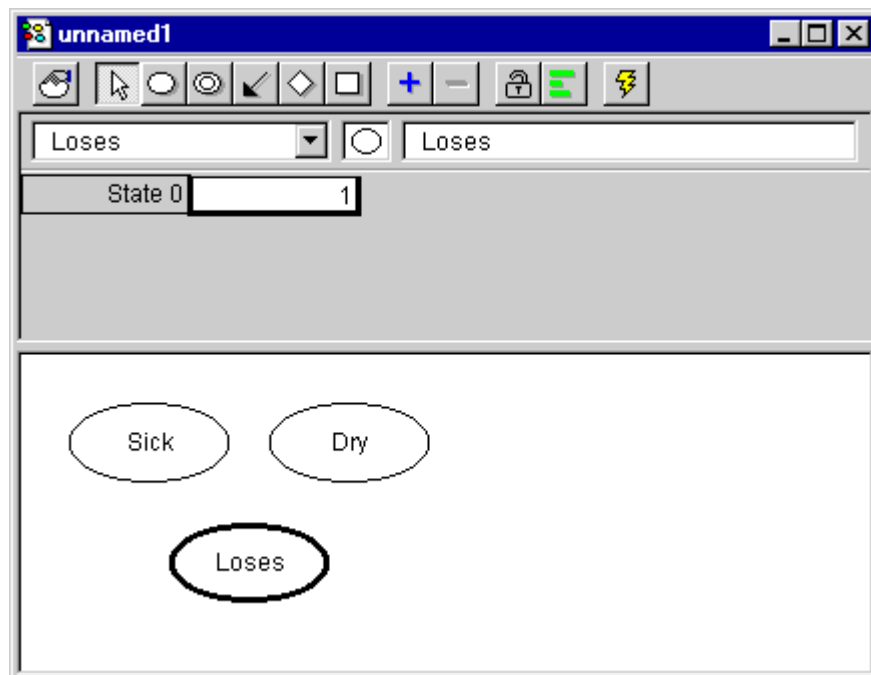


Рисунок 6.5 - Вид БМД без причинно-наслідкових зв'язків в системі HUGIN

4. Визначення всіх можливих стану кожної з вершин БМД

Раніше, кожна з вершин визначилась так, що кожна з них може перебувати в одному з двох станів: вершина Sick в станах 'sick' ('хворіє') і 'not' (немає), вершина Dry в станах 'dry' (' засохло ') і 'not' (немає), а вершина Loses в станах - ' yes '(та) і ' no '(немає). Як задаються стану кожної з вершин БМД в системі HUGIN розглянемо на прикладі визначення станів вершини Sick. Для цього необхідно:

- зробити активної вершину Sick, вибравши її в списку на панелі інструментів або двічі клацнувши на ній мишею.
- натиснути піктограму додавання станів на панелі інструментів (рис.6.6).
- якщо клацнути мишею, перевести курсор в ТУВ на поле, що містить текст State 0 і ввести в нього текст 'sick', задаючи один зі станів вершини. Потім State 1 замінити на 'not', задаючи другий стан вершини.



Рисунок 6.6 - Піктограми додавання і видалення станів в системі HUGIN

Самостійно виконайте аналогічні операції для вершини Dry, визначаючи її стану 'dry' і 'not', і вершини Loses, визначаючи її стану 'yes' і 'no'. Зверніть увагу, що при визначенні вершини Loses, таблиця її умовних ймовірностей буде відрізнитися від ТУВ вершин Sick і Dry. Це пов'язано з тим, що вершина Loses має батьківські вершини, яких Sick і Dry не мають.

5. Задання значення таблиць умовних ймовірностей кожної з вершин БМД

Наступний крок проектування БМД - це коректне завдання значень ТУВ кожної з вершин. За замовчуванням система HUGIN для всіх вершин встановлює рівномірний розподіл. Для прикладу значення ТУВ були задані табл. 6.1 – 6.3 і в нових позначеннях матимуть вигляд табл. 6.4 – 6.6.

Таблиця 6.4 - Таблиця умовних ймовірностей для вершини «Хворіє»

P (Sick)	
Sick = 'sick'	Sick = 'not'
0.1	0.9

Таблиця 6.5 - Таблиця умовних ймовірностей для вершини «засохли».

P (Dry)	
Dry = 'dry'	Dry = 'not'
0.1	0.9

Таблиця 6.6 - Таблиця умовних ймовірностей для вершини «Облетіло».

P (Loses Sick, Dry)				
	Dry = 'dry'		Dry = 'not'	
	Sick = 'sick'	Sick = 'not'	Sick = 'sick'	Sick = 'not'
Loses = 'yes'	0.95	0.85	0.90	0.02
Loses = 'no'	0.05	0.15	0.10	0.98

Процес заповнення таблиць умовних ймовірностей розглянемо на прикладі заповнення ГУВ для вершини Sick:

- виберіть вершину Sick.
- клацніть на поле містить Sick = 'sick'.
- введіть значення 0.1, відповідне табл. 6.4.
- клацніть на поле містить Sick = 'not'.
- введіть значення 0.9, відповідне табл. 6.4.

Таким же чином задайте значення для Dry і Loses. Коли будите вводити значення для Loses, будьте уважні, щоб значення відповідали тим, що позначені в табл.6.6.

Після завдання таблиць умовних ймовірностей для всіх вершин байєсівської мережі довіри, вікно відображення мережі буде виглядати так, як показано на рис. 6.7.

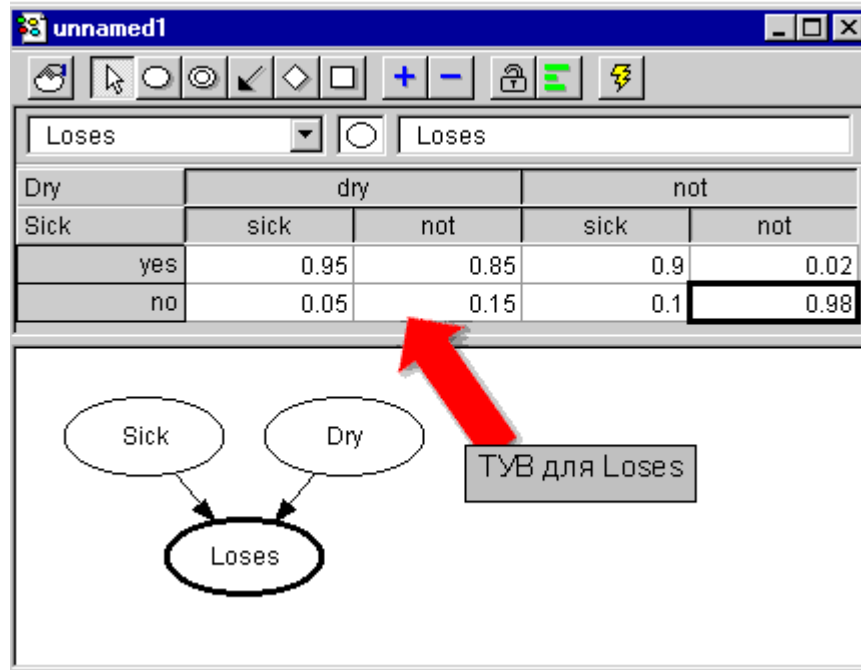


Рисунок 6.7 - Вікно відображення мережі з заданими таблицями умовних ймовірностей

На цьому конструювання проектованої БМД закінчується.

Тепер необхідно зберегти результати роботи. Це можна зробити наступним чином:

- оберіть "Save As" в меню "File".
- введіть ім'я файлу (наприклад apple.hkb або lab1.hkb).
- натисніть "Save".

3. Компілювання спроектованої БМД і робота з нею

Після того як БМД спроектована і визначена в системі HUGIN, її необхідно скопіювати і подивитися як вона працює. Для цього необхідно перевести систему в режим обчислень. З цією метою клацніть мишкою по піктограмі режиму

обчислень [run mode tool button] в панелі інструментів (рис. 6.8). Якщо Ви чітко слідували опису лабораторної роботи, то помилок компіляції не буде. Компіляція такої малої БМД, як наша, закінчиться швидко і система перейде в режим обчислень ["run" mode].



Рисунок 6.8 - Піктограма режиму обчислень

В цьому режимі вікно мережі поділено на дві вертикальні секції (рис. 6.9). Ліва секція являє собою панель списку вершин [node list pane], а права - панель відображення мережі [network pane].

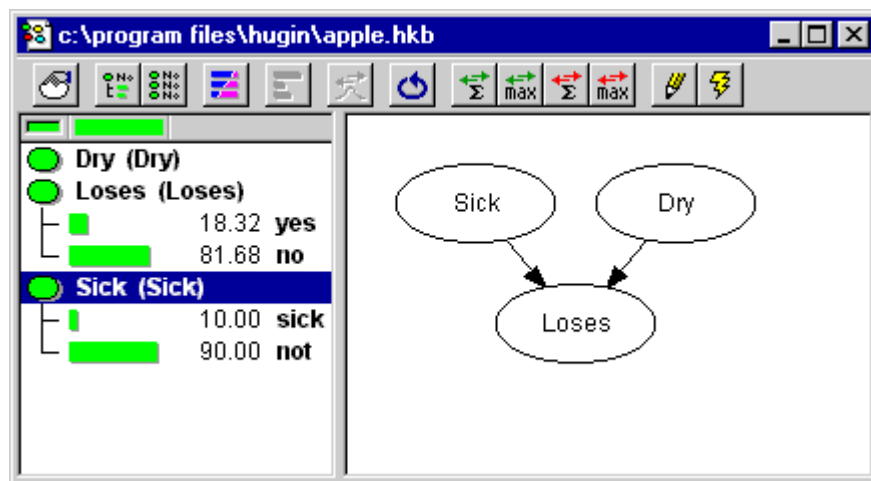


Рисунок 6.9 - Вікно мережі в режимі обчислень

У лівій частині вікна відображаються вершини БМД, всі можливі стани кожної з вершин, а також ймовірності перебування вершини в кожному з станів. в секції списку вершин [node list pane]. Подивитися ці ймовірності можна двічі клацнувши на імені вершини в списку вершин. Тепер розкриємо вершини Loses і Sick. Для цього двічі клацніть на імені вершини Loses, а потім двічі клацніть на Sick. Ви також можете переглянути ймовірності перебування всіх вершин у всіх можливих станах, натиснувши піктограму розкриття списку вершин [expand node list tool], другу зліва на панелі інструментів.

4. Поширення ймовірностей в БМД при надходженні нових свідчень

Тепер, уявімо собі, що Ви хочете використовувати БМД для знаходження ймовірності хвороби яблучного дерева, маючи інформацію про втрати ним листя. Іншими словами в базу знань нашої експертної системи додається новий факт про втрату деревом листя і нас цікавить результат виведення ЕС про хвороби яблучного дерева. Реалізації такого типу запиту до експертної системі на основі БМД виконується наступним чином:

- розкрийте список всіх вершин, натиснувши [expand node list tool].
- введіть факт втрати листя двічі клацнувши на стані 'yes' для вершини Loses.
- натисніть піктограму поширення (рис. 6.10) на панелі інструментів для поширення даного факту на всю БМД.



Рисунок 6.10 - Піктограма поширення

- отримайте ймовірність перебування події Sick ("Хворіє") в стані 'sick' ('хворіє'). Результати поширення ймовірностей по байєсівській мережі довіри наведені на рис. 6.11.

Ймовірність того, що дерево хворіє при обліку причини - наслідкових зв'язків, зазначених в БМД і наявності факту про обпаданні листя складе 0.49.

Якщо отримане значення не збігається з представленим вище, можливо невірне введення даних при заповненні таблиць умовних ймовірностей. Необхідно перевірити таблиці для всіх вершин.

5. Вікна контролю

В п.п.4 панель списку вершин використовувалася для введення фактів і знаходження ймовірностей. Теж саме можливо зробити використовуючи вікно контролю [monitor windows]. Вікна контролю показують ту ж інформацію що і панель списку вершин, але з'являється можливість спостереження цих вікон поруч з відповідними вузлами БМД в панелі відображення мережі. Можливо відк-

риту таке вікно для будь-якої вершини, але кращий варіант - це відкривати лише ті вікна, які становлять особливий інтерес. Інакше, ці вікна будуть займати занадто багато місця на екрані дисплея.

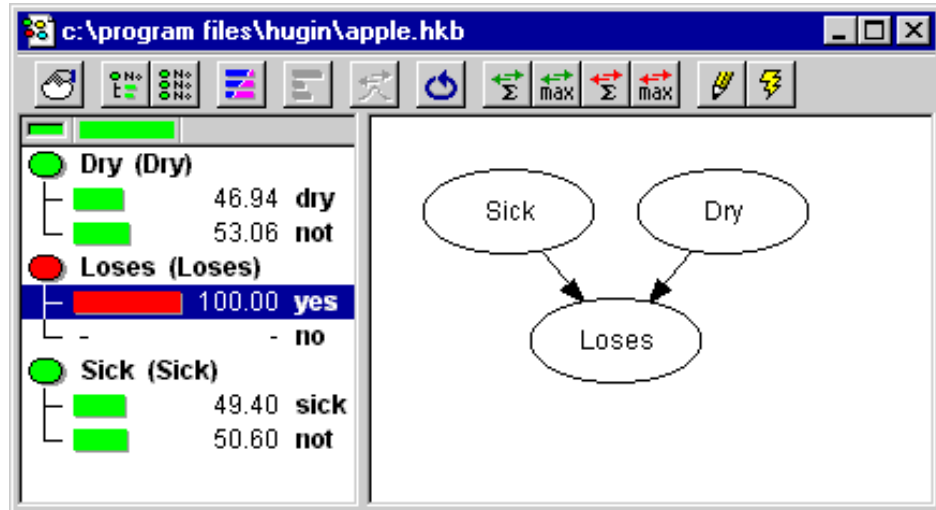


Рисунок 6.11 - Результати поширення ймовірностей, які відображаються у вікні мережі

Тепер відкриємо такі вікна для вершин Sick і Loses і повторимо розрахунок. Спершу ініціалізуємо БМД, тобто наведемо її до вихідного стану. Для цього:

- натисніть піктограму початкової ініціалізації в вигляді кільця зі стрілкою [initialize tool button], розташовану зліва від піктограми поширення.

Після того, як мережу ініціалізовано, можливо відкрити вікна контролю для вершин Sick і Loses (рис. 6.12). Для цього:

- оберіть Sick і Loses (утримуйте SHIFT для вибору більш ніж однієї вершини).
- оберіть "Show Monitor Windows" (Показувати вікна контролю) з пункту меню "View" (Вид).

6. Визначення найбільш ймовірної комбінації станів

На основі аналізу надійшов факту про обпаданні листя, можна з ймовірністю 0.47 зробити припущення, що дерево засихає і з ймовірністю 0.49, що дерево хворіє. При цьому як для події Sick, так і для події Dry більш ймовірними є стан 'not'. Це може породити впевненість, що найкраща комбінація станів, коли Sick і Dry знаходяться в стані 'not'. Однак це невірний висновок.

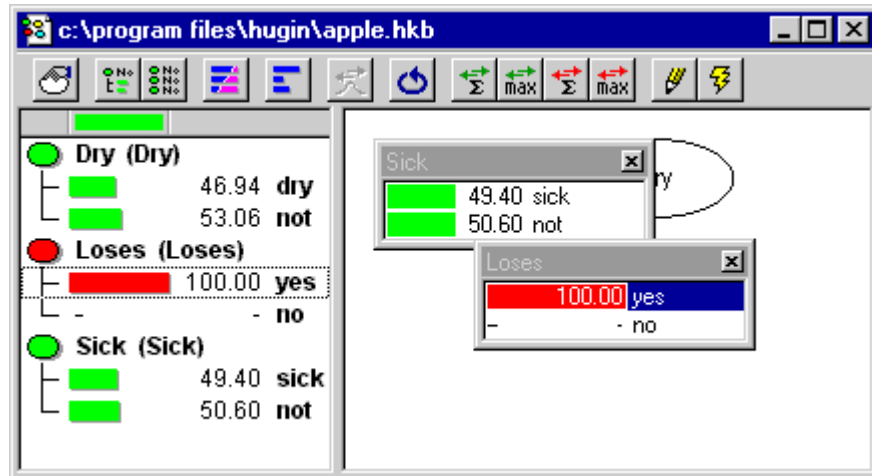


Рисунок 6.12 - Розташування вікон контролю вершин у вікні мережі

Для того, щоб знайти найбільш ймовірну комбінацію станів всіх вершин, необхідно замість поширення сум використовувати поширення максимумів. Піктограма поширення максимумів [Max propagation tool] знаходиться на панелі інструментів праворуч від піктограми поширення сум.

Якщо натиснути на цю піктограму можливо отримати нове поширення в вікнах відображення мережі і вершин. При цьому, кожне з станів вершин, що має значення 100.00, буде належить до найбільш ймовірною комбінації станів. Для розглянутого прикладу буде отримана одна унікальну комбінацію, при якій найбільш ймовірним є, що подія Sick (Хворіє) знаходиться в стані 'sick', а подія Dry (засохла) - в стані 'not'.

З отриманого результату випливає, що незважаючи на те, що стан Sick = 'sick' менш ймовірно ніж стан Sick = 'not', разом з тим саме цей стан Sick = 'sick' входить в найбільш ймовірну комбінацію станів вершин БМД при надходженні факту про обпаданні листя., в той час як стан Sick = 'not' в неї не входить. Це

означає, що необхідно бути дуже обережними в своїх висновках про результати поширення ймовірностей.

Після отримання висновку про найбільш імовірною комбінації станів вершин БМД, можливо виконати обчислення ймовірність настання цієї (або будь-який інший) комбінацій станів за умови настання введених в БМД свідощів.

7. Розрахунок ймовірності комбінацій станів

Тут ми розповімо про техніку розрахунку ймовірності найбільш імовірною комбінації станів, отриманих при наявності факту, що яблучне дерево втрачає своє листя. Ця ймовірність записується у вигляді:

$$P(\text{Sick} = \text{'sick'}, \text{Dry} = \text{'not'} \mid \text{Loses} = \text{'yes'})$$

і для її визначення можна скористатися відомою формулою з теорії ймовірності [7]:

$$P(A, B) = P(A \mid B) P(B),$$

переписавши її у вигляді:

$$P(A \mid B) = P(A, B) / P(B),$$

Кожен раз при виконанні розрахунку поширення сум для БМД, в нижньому лівому кутку HUGIN буде відображатися ймовірність (значення $P(\text{All})$) одночасного настання декількох подій, яку можна записати в такий спосіб:

$$P(A_1, A_2, \dots, A_n)$$

Якщо тепер замінимо A на $\text{Sick} = \text{'yes'}$, $\text{Dry} = \text{'not'}$ і B на $\text{Loses} = \text{'yes'}$, то вищенаведена формула для визначення умовної ймовірності може бути переписана у вигляді:

$$P(\text{Sick} = \text{'yes'}, \text{Dry} = \text{'not'} \mid \text{Loses} = \text{'yes'}) = P(\text{Sick} = \text{'sick'}, \text{Dry} = \text{'not'}, \text{Loses} = \text{'yes'}) / P(\text{Loses} = \text{'yes'})$$

Якщо вибрати стан 'yes' для вершини Loses і зробити розрахунок поширення сум, то можливо здобути ймовірність перебування вершини Loses в стані 'yes'. Її значення буде 0.1832. (Тобто $P(\text{Loses} = \text{'yes'}) = 0.1832$). Визначивши чому дорівнює $P(\text{Loses} = \text{'yes'})$, залишилося визначити чому дорівнює $P(\text{Sick} = \text{'sick'}, \text{Dry} = \text{'not'}, \text{Loses} = \text{'yes'})$. Це робиться в такий спосіб:

- введіть Sick = 'sick', Dry = 'not', і Loses = 'yes' в БМД.
- виконайте розрахунок поширення сум (натисніть sum propagation tool).
- подивіться $P(\text{Sick} = \text{'sick'}, \text{Dry} = \text{'not'}, \text{Loses} = \text{'yes'})$ як значення $P(\text{All})$ в нижньому лівому кутку.

Це значення має дорівнювати 0.081. Тепер можливо розрахувати необхідну ймовірність, що буде дорівнює:

$$P(\text{Sick} = \text{'yes'}, \text{Dry} = \text{'not'} \mid \text{Loses} = \text{'yes'}) = 0.081 / 0.1832 = 0.442$$

Т.ч., на основі надійшов факту про обпаданні листя експертна система робить висновок про те, що найбільш ймовірною є ситуація, пов'язана з тим, що дерево захворіло і не засохло. При цьому ймовірність такої ситуації дорівнює 0.442.

Контрольні питання

1. Основи побудови найпростішої байєсівської мережі довіри
2. Основні прийоми роботи в HUGIN при побудові байєсівських мереж довіри
3. Додавання вершин в HUGIN
4. Встановлення зв'язків в HUGIN
5. Визначення станів вершин байєсівських мереж довіри в HUGIN
6. Завдання таблиць умовних ймовірностей в HUGIN
7. Компілювання байєсівських мереж довіри і робота з ними

8. Поширення ймовірностей в байєсівських мережах довіри
9. Вікно контролю в HUGIN
10. Найбільш ймовірні комбінації станів
11. Розрахунок ймовірності комбінацій станів