

**Побудова системи підтримки прийняття рішень на основі теорії байєсівських мереж для моделювання поведінки складних систем**

**Построение системы поддержки принятия решений на основе теории сетей Байеса для моделирования поведения сложных систем**

**Construction decision support system by using Bayesian network's theory for modeling behavior of complex systems.**

**1. Номер державної реєстрації теми: 0109U000300**

**2. Науковий керівник: д.т.н., проф. Бідюк П.І., Бидюк П.И., Bidyuk P.I.**

**3. Суть розробки, основні результати**

**(укр.)**

Розроблено низку математично обґрунтованих методів для розв'язання слабоструктурованих задач моделювання, прогнозування та класифікації на основі використання мереж Байеса із прихованими вершинами. Створена нова п'ятикрокова методика знаходження параметрів мережі Байеса з прихованими вершинами на основі алгоритму максимізації математичного очікування. Для визначення мір зв'язку між вершинами мереж Байеса запропоновано використати такі коефіцієнти: Пірсона; Чупрова; Крамера; лямбда Гудмана та значення взаємної інформації. Для розв'язання задачі моделювання поведінки складних систем запропонована оригінальна методика побудови та застосування гібридних мереж Байеса. Вона ґрунтується на використанні структурного EM або градієнтного методів для побудови топології мережі та формування імовірнісного висновку на основі як точних так і наближених методів. Для формування висновку у гібридній мережі Байеса запропоновано новий підхід до дискретизації неперервних змінних із використанням методів кластерного аналізу.

На основі запропонованих методів і алгоритмів розроблена оригінальна архітектура системи підтримки прийняття рішень для розпізнавання образів, а також програмно реалізована система підтримки прийняття рішень для інтелектуального аналізу даних на основі байєсівських мереж. При цьому забезпечена можливість оперативної модифікації створених обчислювальних процедур за рахунок відкритої модульної архітектури комп'ютерної системи підтримки прийняття рішень при моделюванні та прогнозуванні станів. На основі розробленого програмного забезпечення побудовано низку практично значимих моделей у вигляді гібридних мереж та мереж з прихованими вершинами.

Побудову математичних імітаційних моделей виконано на основі статистичних (експериментальних) даних та експертних оцінок, які поповнюються та оновлюються у процесі використання моделі. Для перевірки коректності теорії використано аналітичні процедури та обчислювальні експерименти з використанням значних об'ємів статистичних даних та експертних оцінок.

**(рос.)**

Разработан набор математически обоснованных методов для решения слабоструктурованных задач моделирования, прогнозирования и классификации на основе использования сетей Байеса со скрытыми вершинами. Создана новая пятишаговая методика нахождения параметров сети Байеса со скрытыми вершинами на основе алгоритма максимизации математического ожидания. Для определения меры связи между вершинами сети Байеса предложено использование коэффициентов: Пирсона; Чупрова; Крамера; лямбда Гудмана и значение взаимной информации. Для решения задачи моделирования поведения

сложных систем предложена оригинальная методика построения и применение гибридных сетей Байеса. Методика основывается на использовании структурного EM или градиентного методов для построения топологии сети и формирование вероятностного вывода на основе как точных, так и приближенных методов. Для формирования вывода в гибридной сети Байеса предложен новый подход к дискретизации непрерывных переменных с использованием методов кластерного анализа.

На основе предложенных методов и алгоритмов разработанная оригинальная архитектура системы поддержки принятия решений для распознавания образов, а также в виде компьютерной программы реализована система поддержки принятия решений для интеллектуального анализа данных на основе байесовских сетей. При этом обеспеченная возможность оперативной модификации созданных вычислительных процедур за счет открытой модульной архитектуры, компьютерной системы поддержки принятия решений, при моделировании и прогнозировании состояний. На основе разработанного программного обеспечения построен набор практически значимых моделей в виде гибридных сетей и сетей со скрытыми вершинами.

Построение математических имитационных моделей выполнено на основе статистических (экспериментальных) данных и экспертных оценок, которые пополняются и обновляют в процессе использования модели. Для проверки корректности теории использованы аналитические процедуры и вычислительные эксперименты с использованием значительных объемов статистических данных и экспертных оценок.

**(англ.)**

Developed some of methods for solving ill-structured problems for modeling, prediction and classification. All methods use Bayesian networks. Proposed a new five step method for finding the parameters of Bayesian networks with hidden nodes. Method bases on an expectation maximization algorithm. Suggested Pearson's, Chuprov's, Cramer's, Goodman's and mutual information coefficients for finding interconnections between Bayesian network's nodes. For solving the problem of modeling the behavior of complex systems proposed original method for construction and application hybrid Bayesian networks. Methodology uses structural EM or gradient methods for constructing the topology of network. For probabilistic inference uses exact and approximate algorithms. For build probabilistic inference in hybrid Bayesian networks proposed a new approach for discretization of continuous variables by using the methods of cluster analysis.

Developed the original architecture of decision support systems for pattern recognition by using of proposed methods and algorithms. Realized computer program for data mining by using Bayesian networks. Developed program gives another users the possibility of rapid modification of computational procedures through an open modular architecture of decision support system for modeling and forecasting. Constructed some of practically important models in the form of hybrid networks and networks with hidden nodes by using developed software.

Construction of mathematical simulation model is based on statistical (experimental) data and expert estimates, which is maintained and updated in the process of using the model. For validation the correctness of the proposed theory used analytical procedures and computational experiments with big number of statistical data and expert estimates.

#### **4. Наявність охоронних документів на об'єкти права інтелектуальної власності:**

- Терентьев О. М., Трофименко Д. В., Давиденко В. І. авторське свідоцтво № 34443 України на комп'ютерну програму "Bayesian Network Master BNetMaster". – № 34657; заявл. 09.06.2010; опуб. 09.08.2010.
- Гуз Н. С., Терентьев О.М., Коршевнюк Л.О., Бідюк П.І. "Пристрій для розпізнавання символів на основі мереж Байеса". подано документи для оформлення патенту.

## **5. Порівняння зі світовими аналогами.**

Розроблена комп'ютерна програма “Bayesian Network Master BNetMaster” та запропоновані методики побудови і використання гібридних мереж з прихованими вершинами відповідають світовому рівню. Завдяки розробленій п'яти крокової методиці знаходження параметрів мережі з прихованими вершинами на основі алгоритму максимізації математичного очікування та методики побудови гібридних мереж Байєса із використанням градієнтного методу.

## **6. Економічна привабливість для просування на ринок**

Розроблені методи побудови топології мереж Байєса та ймовірнісного висновку в них відносяться до надзвичайно популярного напрямку сучасної математики “інтелектуального аналізу даних”, призначеного для автоматичного пошуку причинно-наслідкових взаємозв'язків та моделюванні різноманітних процесів на основі використання великих статистичних баз даних. Вартість аналогічних бізнес-версій програмних засобів складає від 140 до 5200 €, в залежності від типу ліцензії та комплектації. Наприклад, французька BayesiaLab, канадська Netica та датська Hugin-Expert.

## **7. Потенційні користувачі.**

Розроблений пакет прикладних програм BNetMaster разом з отриманими фундаментальними та прикладними результатами НДР можуть бути використані для розв'язання задач моделювання, розпізнавання, прогнозування та діагностування статичних і динамічних процесів у науково-дослідних та проектних інститутах Національної академії наук України, в галузевих інститутах, в Інституті економіки, Інституті проблем моделювання в енергетиці, Інституті космічних досліджень, на виробничих машинобудівних підприємствах, у переробній промисловості, медичних закладах, банківських та фінансових установах.

Одним з перспективних прикладних напрямів використання результатів виконання НДР є створення комерційних інформаційних систем підтримки прийняття рішень при моделюванні, прогнозуванні та діагностуванні технічних систем та технологічних процесів. Наприклад, в подальшому можливе використання отриманих результатів в межах створення центру компетенції SAS при ННК ПСА з метою розширення функціональних можливостей цієї системи.

## **8. Стан готовності розробки.**

Розроблена працююча версія комп'ютерної програми BNetMaster для побудови топології та ймовірнісного висновку в мережах Байєса з прихованими вершинами на основі статистичних даних. Бажано розширення функціональних можливостей програми за рахунок реалізації більш складного мультидисциплінарного застосування разом з нейронними мережами.

## **9. Існуючі результати впровадження.**

Результати роботи впроваджені у вигляді системи підтримки прийняття рішень для оцінювання кредитних ризиків в ПромінвестБанку, що дозволить більш точно оцінювати ймовірність дефолту фізичних осіб.

За матеріалами роботи захищена кандидатська дисертація за темою: “Моделі і методи побудови та аналізу байєсівських мереж для інтелектуального аналізу даних”.

Результати роботи впроваджені у навчальний процес Національного технічного університету «Київський політехнічний інститут», Херсонського національного технічного університету та Херсонського морського інституту. В ННК ПСА результати роботи використовуються в навчальних дисциплінах: “Моделювання економіки перехідного

періоду”, “Аналіз часових рядів”, “Прикладна статистика” та “Проектування комп’ютерних інформаційних систем”. Створюється новий курс: “Мережі Байєса і дерева рішень”.

Заплановано розроблених методів та алгоритмів сумісно з компанією SAS Institute Ukraine для розробки нового статистичного модуля SAS Bayesian Net.

#### 10. Назва підрозділу, телефон, e-mail.

ННК «ПСА» НТУУ «КПІ»

Тел.: 241-86-59

E-mail: [pbidyuke@gmail.com](mailto:pbidyuke@gmail.com)

#### 11. Перелік публікацій за матеріалами досліджень за період виконання розробки

1. Терентьев О.М. Модели і методи побудови та аналізу байєсівських мереж для інтелектуального аналізу даних // Дисертаційна робота. – Київ, 2009. – 258 с.
2. Terentyev A. N., Bidiuk P. I., Mironova A. V., Medin N. Y. Comparison of Data Mining Methods while Credit Rating of Natural Persons // Journal of automation and information sciences. – NY.: Begell House Inc, 2009. – 41. – P. 71-80.
3. Терентьев О.М., Бидюк П. І., Коршевнік Л. О. Алгоритм вероятностного вывода в байесовских сетях // Міжнародний науково-технічний журнал “Системні дослідження та інформаційні технології”. – 2009. – № 2. – С. 107-111.
4. Терентьев А.Н., Бидюк П.И., Миронова А.В., Медин Н.Ю. Сравнение методов интеллектуального анализа данных при оценивании кредитоспособности физических лиц // Международный научно-технический журнал “Проблемы управления и информатики”. – К.: ИКИ НАНУ-НКАУ, 2009. – № 5. – С. 141-149.
5. Бидюк П.И., Давиденко В.И., Трофименко Д.В., Терентьев А.Н. Сравнительный анализ методов оценки взаимосвязи между вершинами при построении байесовских сетей // Международный научно-технический журнал “Проблемы управления и информатики”. – К.: ИКИ НАНУ-НКАУ, 2010. – № 6. – С. 64-73.
6. Бидюк П.І., Кузнєцова Н. Порівняльний аналіз характеристик моделей оцінювання ризиків кредитування // Наукові вісті НТУУ «КПІ». – 2010. – № 1, – С. 42-53.
7. Баклан Я.І., Бидюк П.І. Застосування динамічних байєсівських мереж до розв’язання задач розпізнавання (аутифікації) користувачів // Проблеми інформаційних технологій (Херсон) . – 2009, – 2, – С. 20-32.
8. Кузнєцова Н.В., Бидюк П.І. Порівняльний аналіз моделей оцінювання ризиків кредитування // Вісник Херсонського Національного технічного університету. – 2010. – № 2 (38). – С. 52-62.
9. Коршевнік Л.О., Бидюк П.І. Система для оцінювання і прогнозування стану підприємства на основі мереж Байєса // Вісник Чорноморського державного університету ім. Петра Могили. – 2010, – С. 60 -74.
10. Бидюк П.І., Терентьев О.М., Коновалюк М.М. Байєсівські мережі в технологіях інтелектуального аналізу даних // Вісник Чорноморського державного університету ім. Петра Могили. – 2010. – С. 6-16.
11. Кузнєцова Н. В. Гібридні мережі Байєса: основні особливості і точні методи формування висновку // Тр. Одес. политехн. ун-та. – Одесса, 2009. – Вып.1(31). – С. 114-121.
12. Бидюк П. І., Кузнєцова Н. В. Наближені методи формування ймовірнісного висновку та їх модифікація для гібридних мереж Байєса // Наукові праці. Серія: Комп’ютерні технології (Миколаївський державний гуманітарний університет ім. Петра Могили комплексу НаУКМА). – Миколаїв, 2009. – Вип. 104, т. 117. – С. 17-30.
13. Колбасюк Д.А., Терентьев О.М. Прогнозування продаж торговельного підприємства із використанням методів інтелектуального аналізу даних / Менеджмент малого і середнього

бізнесу: ефективність, конкурентоспроможність, стійкість: сб. науч. трудов по материалам VIII-й междунар. науч.-прак. конф., г. Севастополь, 21-25 вересня 2009. – Севастополь: СевНТУ, 2009. – С. 42-43.

14. Колбасюк Д.А., Бідюк П.І., Терентьев О.М., Шумейко О.М. Прогнозування темпів приросту ВІЛ/СНІД хворих в Україні на 2016 рік з використанням методів регресійного аналізу / Интеллектуальные системы принятия решений и проблемы вычислительного интеллекта: сб. науч. трудов по материалам междунар. конф., 17-21 мая 2010 г., Евпатория. – Херсон: ХНТУ, 2010. – Т. 2 – С. 80-82.

15. Бідюк П.І., Терентьев О.М., Миронова О.В. Класифікація кредитоспроможності фізичних осіб за допомогою теорії дерев рішень та методу Монте-Карло / Интеллектуальные системы принятия решений и проблемы вычислительного интеллекта: сб. науч. трудов по материалам междунар. конф., 17-21 мая 2010 г., Евпатория. – Херсон: ХНТУ, 2010. – Т. 2 – С. 326-329.

16. Бідюк П.І., Литвиненко В.І., Фефелов А.О. Інформаційна гібридна система технічного діагностування / Интеллектуальные системы принятия решений и проблемы вычислительного интеллекта: сб. науч. трудов по материалам междунар. конф., 17-21 мая 2010 г., Евпатория. – Херсон: ХНТУ, 2010. – Т. 1 – С. 475.

17. Литвиненко В.И., Бідюк П.І., Фефелов А.А. Построение вероятностного вывода в сетях Байеса на основе LS-метода / Интеллектуальные системы принятия решений и проблемы вычислительного интеллекта: сб. науч. трудов по материалам междунар. конф., 17-21 мая 2010 г., Евпатория. – Херсон: ХНТУ, 2010. – Т. 1 – С. 260.

18. Терентьев О.М Побудова кредитних скорінгових моделей із використанням аналітичної системи SAS Enterprise Miner / Системний аналіз та інформаційні технології: матеріали 12-ї Міжнародної науково-технічної конференції SAIT 2010, Київ, 25–29 травня 2010 р. / ННК «ПСА» НТУУ «КПІ». – К.: ННК «ПСА» НТУУ «КПІ», 2010. – С. 523.

19. Бідюк П.І., Збруцька І.О. Аналіз стану підприємства за допомогою байєсівського методу / 3-я Всеукраїнська Научно-практическа конференція молодих учених, Севастополь, 26-30 апреля, 2010. – С. 49.

20. Кузнецова Н. В. Аналіз кредитоспроможності позичальника з використанням гібридних мереж Байеса / Н. В. Кузнецова // Матеріали XI міжнар. наук.-техн. конф. «Системний аналіз та інформаційні технології», SAIT 2009. – К. : ННК «ПСА» НТУУ «КПІ», 2009. – С. 129.

21. Bidyuk P. I., Gasanov A.S., Omelchenko O.V. Specific features of information distribution in Bayesian networks / XV-th International Conference – Problems of Decision Making Under Uncertainties, Lviv, May 17-21, 2010. – P. 43.

22. Бідюк П. І., Кузнецова Н. В. Основні етапи побудови і приклади застосування мереж Байеса // Системні дослідження та інформаційні технології. – 2007. – № 4. – С. 26–39.

23. Кузнецова Н. В. Системний підхід до аналізу кредитних ризиків з використанням мереж Байеса / Н. В. Кузнецова, П. І. Бідюк // Наукові вісті НТУУ «КПІ». – 2008. – №3. – С. 11–24.

**12. Фото / схема, слайди презентації розробки в електронному вигляді (рекламного характеру)**



Національний технічний університет України  
"Київський політехнічний інститут"

## “Побудова системи підтримки прийняття рішень на основі теорії байєсівських мереж для моделювання поведінки складних систем” 2234-п

Номер державної реєстрації  
теми: 0109U000300

Науковий керівник:  
д.т.н., проф. каф. “ММСА”  
Бідюк П.І.,



Київ - 2010

1

### Поняття системного аналізу

- Системний аналіз - науковий метод пізнання, що представляє собою послідовність дій з установаження структурних зв'язків між змінними або елементами досліджуваної системи.
- Історичний розвиток системних ідей

2500-2000 р. до н.е	400-300 р. до н.е.	1500 р.р.	1600 р.р.	1700-1800 р.р.	1920-ті р.р.	1940-і р.р.	1950-ті р.р.	2000-...
Поява слова "система" в Др. Греції	Демокріт, Арістотель	Коперник	Галілей, Ньютон	Ламберт, Кант Фіхте, Гегель	Богданов	Л. фон Бергаланфі	Вінер	Згуровський, Панкратова
	Ціле більше суми його частин	Геліоцентрична картина світу	Ціле-частина, причина-наслідок...	Системність наукових знань	Ідеї кібернетики	Концепція організму як системи	Кібернетика	Системний підхід, системна методологія
								

## Типи проблем системного аналізу

- **Добре структуровані** (*well-structured*) – істотні залежності з'ясовані дуже добре
 

Математичне програмування

Дослідження операцій
  
- **Неструктуровані** (*unstructured*) – містять лише опис найважливіших ресурсів, ознак і характеристик, кількісні залежності між якими абсолютно невідомі
 

Теорія евристичних рішень
  
- **Слабоструктуровані** (*ill-structured*) – містять як якісні, так і кількісні елементи
 

Методологія системного аналізу

СППР

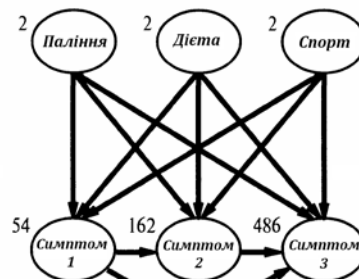
## Для чого необхідні приховані вершини в мережах Байєса?

Діагностична мережа для серцевого захворювання з можливістю існування прихованої змінної; кожна змінна має по 3 можливі стани.



а)

Мережа з прихованою вершиною "Хвороби серця" – кількість параметрів 78



б)

Мережа з видаленою вершиною "Хвороби серця" – кількість параметрів 708

## Застосування прихованих вершин в мережах Байєса

### Переваги:

- спрощення структури мережі
- зменшення кількості параметрів, необхідних для визначення БМ, а це в свою чергу дозволяє різко зменшити об'єм даних, необхідних для визначення в процесі навчання цих параметрів

### Недоліки:

- ускладнення задачі навчання
- у разі невідомої структури мережі невирішеною проблемою залишається визначення кількості прихованих вершин, місце їх розташування та кількість станів кожної з них

## Відома структура, приховані вершини

### EM-алгоритм

Схема:

1:  $i = 0$

   вибір  $\theta^0$

2: repeat

3:      $i = i + 1$

4:      $\theta^i = \arg \max_{\theta} Q(\theta : \theta^{i-1})$

5: until  $Q(\theta^i : \theta^{i-1}) \leq Q(\theta^{i-1} : \theta^{i-1})$

■ на кроці **E** оцінюється значення  $N^*$  на основі параметрів  $\Theta^i$

■ на кроці **M** вибирається найкраще значення параметрів  $\Theta^{i+1}$ , максимізуючи  $Q$

$Q(\theta : \theta^*) = \sum_i \sum_{X_i = x_k} \sum_{P_i = pa_j} N_{ijk}^* \log P_{\theta_{jk}}(D)$  - математичне очікування логарифмічної функції правдоподібності

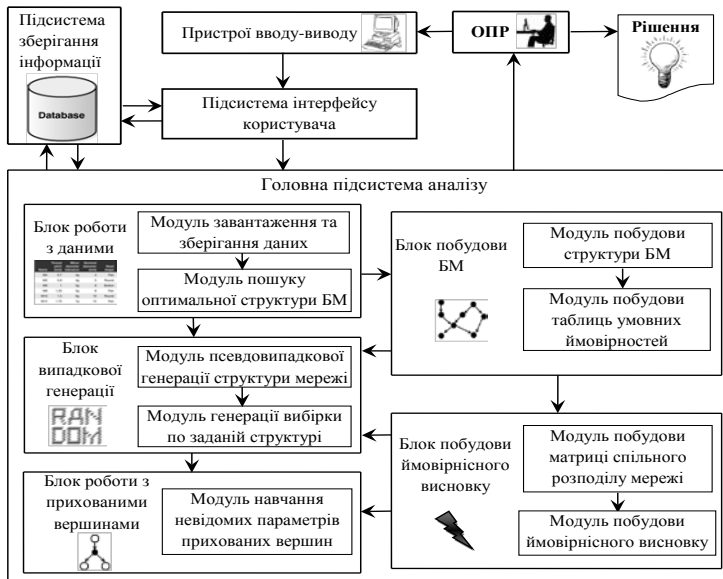
$N_{ijk}^* = N \cdot P(X_i = x_k, P_i = pa_j | \theta^*)$  - отримується за допомогою ймовірнісного висновку в мережі Байєса

$Pa(X_i)$  - множина предків вершини  $X_i$

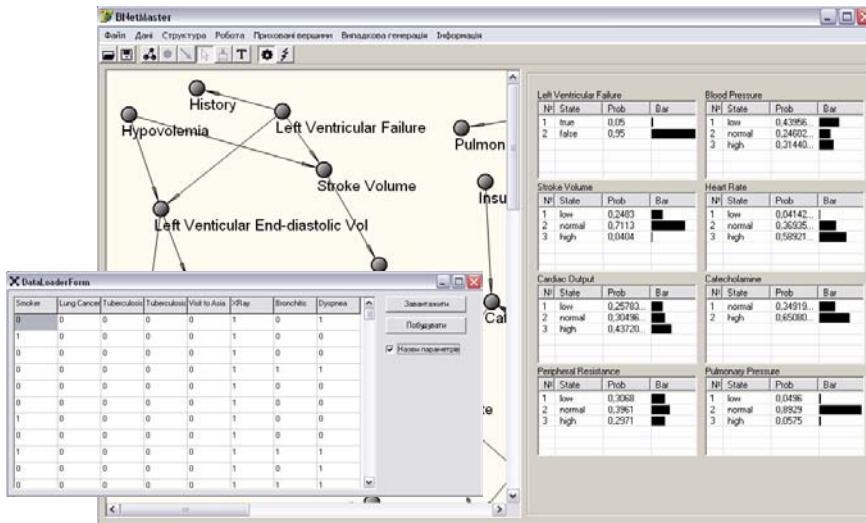
$N$  - розмір вибірки множини навчальних даних  $D$



## Архітектура СППР “BNetMaster”



## СППР “BNetMaster”

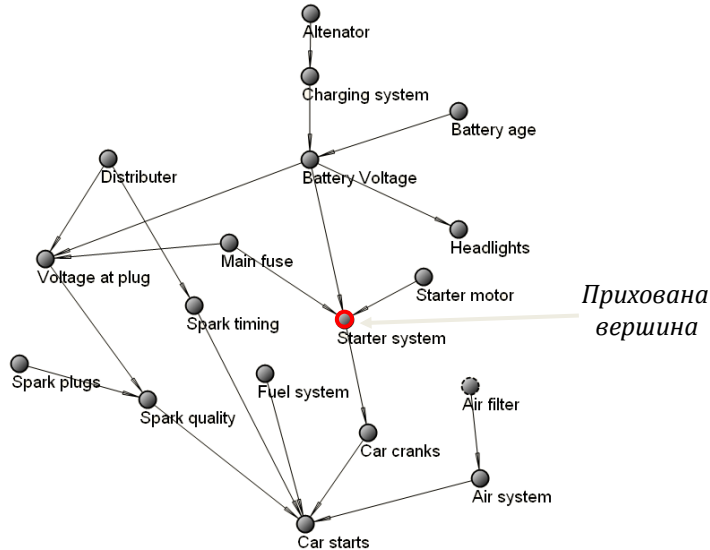


## Мережа CarStarts: Опис вхідних даних

- Використовується для діагностики старту роботи автомобіля на основі інформації про свічки запалення, роботу фар, головного запобіжника тощо
- Мережа містить 18 вершин та 19 дуг
- Список вершин:
 

– Генератор	<input type="checkbox"/> Головний запобіжник
– Система зарядки	<input type="checkbox"/> Розподільювач
– Термін дії акумулятора	<input type="checkbox"/> Напруга на свічці
– Напруга акумулятора	<input type="checkbox"/> Свічка запалення
– Фари	<input type="checkbox"/> Встановлення моменту запалення
– Повітряний фільтр	<input type="checkbox"/> Якість запалення
– Повітряна система охолодження	<input type="checkbox"/> Паливна система
– Мотор стартера	<input type="checkbox"/> Початок роботи автомобіля
– Система старту	<input type="checkbox"/> Заведення авто

## Мережа CarStarts: Топологія



## Мережа CarStarts: Отримані результати

- Значення ймовірностей станів прихованої вершини

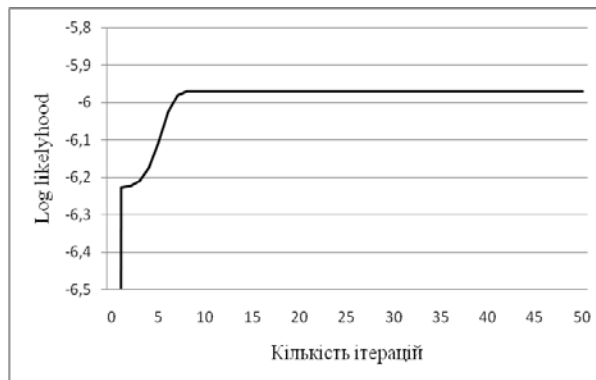
Стан	Початкові	Отримані в результаті експерименту
Okay	0,60	0,55
Faulty	0,40	0,45

- Таблиці умовних ймовірностей

Початкові					Експеримент				
MF	SM	BV	Okay	Fault	MF	SM	BV	Okay	Fault
okay	Okay	strong	0,98	0,02	okay	Okay	strong	0,88	0,12
okay	Okay	weak	0,90	0,10	okay	Okay	weak	0,83	0,17
okay	Okay	dead	0,10	0,90	okay	Okay	dead	0,1	0,9
okay	Faulty	strong	0,02	0,98	okay	Faulty	strong	0,00	1,00
okay	Faulty	weak	0,01	0,99	okay	Faulty	weak	0,00	1,00
okay	Faulty	dead	0,01	0,99	okay	Faulty	dead	0,00	1,00
...	...	...	...	...	...	...	...	...	...
blown	Okay	weak	0	1,00	blown	Okay	weak	0	1,00
blown	Okay	dead	0	1,00	blown	Okay	dead	0	1,00

## Мережа CarStarts: Отримані результати

Графік логарифмічної функції правдоподібності



- Збільшення логарифму функції правдоподібності з -16,90 до -5,97

## Огляд ринку програмного забезпечення для роботи з мережами Байєса

- Розроблена СППР VNetMaster не поступається за функціональними можливостями більшості програмних продуктів, що присутні на ринку
- VNetMaster забезпечує основні функції СППР для роботи з мережами Байєса – параметричне та структурне навчання БМ, формування ймовірнісного висновку
- VNetMaster має відкритий вихідний код, що дозволяє розширювати функціональні можливості продукту
- Графічний інтерфейс користувача та безкоштовне використання СППР VNetMaster є значною конкурентною перевагою

**Дякую за увагу!**

