

Міністерство освіти і науки України  
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «КИЄВО-МОГИЛЯНСЬКА  
АКАДЕМІЯ»

Факультет інформатики

Кафедра інформатики

**Магістерська робота**

Освітній ступінь: магістр

**На тему: «РОЗРОБКА СОЦІАЛЬНО-ОРІЄНТОВАНОГО  
ЗАСТОСУНКУ З ВИКОРИСТАННЯМ БАЄСІВСЬКОЇ БАЗИ  
ЗНАНЬ»**

Виконала: студентка 2 року навчання

Спеціальності

122 Комп'ютерні науки

Салій Анна Сергіївна

Керівник:

доцент,

кандидат фіз.-мат. наук

Ющенко Ю.О.

Рецензент М. М. Глибовець

Магістерська робота захищена

з оцінкою \_\_\_\_\_

Секретар ЕК С.А. Мелешенко

« \_\_\_\_ » \_\_\_\_\_ 2022

Київ 2022

## Календарний план виконання курсової роботи

**Тема: Розробка соціально-орієнтованого застосунку з використанням  
басівської бази знань**

**Календарний план виконання роботи:**

| №<br>п/п | Назва етапу                                           | Термін<br>виконання<br>етапу |
|----------|-------------------------------------------------------|------------------------------|
| 1.       | Отримання завдання.                                   | 20.10.2022                   |
| 2.       | Огляд технічної літератури за темою роботи.           | 05.12.2022                   |
| 3.       | Вивчення методів аналізу даних.                       | 19.12.2022                   |
| 4.       | Написання умов до текстової частини роботи.           | 19.01.2022                   |
| 5.       | Написання умов до практичної частини проекту.         | 21.01.2022                   |
| 6.       | Написання першої частини.                             | 22.02.2022                   |
| 7.       | Перегляд змісту роботи з керівником.                  | 23.02.2022                   |
| 8.       | Написання другої частини.                             | 24.04.2022                   |
| 9.       | Перегляд змісту роботи з керівником.                  | 25.04.2022                   |
| 10.      | Написання висновків.                                  | 28.04.2022                   |
| 11.      | Перегляд змісту роботи з керівником.                  | 30.05.2022                   |
| 12.      | Створення слайдів для доповіді та написання доповіді. | 02.06.2022                   |
| 13.      | Аналіз отриманих результатів з керівником.            | 04.06.2022                   |
| 14.      | Корегування роботи .                                  | 09.06.2022                   |
| 15.      | Остаточне оформлення роботи та слайдів.               | 11.06.2022                   |

Студент \_\_\_\_\_

Керівник \_\_\_\_\_

“ \_\_\_\_\_ ”

## Зміст

|                                                                |    |
|----------------------------------------------------------------|----|
| Анотація                                                       | 4  |
| Вступ                                                          | 5  |
| Розділ 1. Теоретична частина                                   | 7  |
| 1.1 Історія виникнення баєсівської мережі                      | 7  |
| 1.2 Поняття баєсівської мережі                                 | 7  |
| 1.3 Аналіз методів побудови баєсівської мережі                 | 15 |
| 1.3.1 Ручна побудова                                           | 15 |
| 1.3.2 Автоматична побудова                                     | 17 |
| 1.4 Структура баєсівської мережі                               | 17 |
| Висновки до розділу 1                                          | 18 |
| Розділ 2. Дослідження предметної області                       | 19 |
| 2.1 Проблема розподілу трафіку                                 | 19 |
| 2.2 Задача матричного оцінювання                               | 20 |
| 2.3 Комбіновані методи                                         | 22 |
| Висновки до розділу 2                                          | 23 |
| Розділ 3. Запропонована модель прогнозування трафіку           | 24 |
| 3.1 Побудова простої моделі в практичних випадках              | 24 |
| 3.2 Повний варіант моделі                                      | 25 |
| 3.3 Використання моделі для прогнозування транспортних потоків | 27 |
| Висновки до розділу 3                                          | 28 |
| Розділ 4. Практична частина                                    | 29 |
| 4.1 Опис реалізації                                            | 29 |
| 4.2 Використані технології                                     | 32 |
| Висновки до розділу 4                                          | 33 |
| Висновки                                                       | 34 |
| Джерела                                                        | 35 |

## **Анотація**

Дана робота стосується проблеми прогнозування потоків трафіку. Було розроблено соціально-орієнтований застосунок для прогнозування потоків трафіку. З цією метою будується бассовська мережа, яка здатна врахувати випадковий характер рівня загального середнього потоку та мінливість парних потоків джерело-призначення разом із випадковим порушенням рівнянь балансу для пар джерело-призначення. Бассовські мережі забезпечують спільну щільність усіх неспостережуваних змінних  $i$ , зокрема, відповідні умовні та граничні щільності, які дозволяють не лише спільні прогнози, але й ймовірнісні інтервали. Вплив перевантаженого трафіку також можна врахувати шляхом поєднання правил розподілу трафіку із запропонованою моделлю мережі Басса.

## Вступ

У роботі розглядається проблема оцінки потоків трафіку, включаючи потоки джерело-призначення і потоки каналів, і оновлення цієї інформації, коли нова інформація доступна.

Знання транспортних потоків зазвичай вимагає двох різних інструментів: модель призначення користувача та метод оцінки матриці поїздки. Модель призначення користувача має в якості вхідних даних потоки пар джерело-призначення і надає як вихідні дані ймовірності користувача вибрати різні маршрути пар джерело-призначення. Проблема матричної оцінки має як вхідні дані ці ймовірності та потоки зв'язків, а на виході дає потоки пар «початок-пункт призначення». Однак вони тісно пов'язані між собою і можуть бути об'єднані в одну задачу, в якій оцінка потоку матриці і задачі розподілу трафіку стають єдиною.

Баєсовські мережі забезпечують спільну щільність усіх неспостережуваних змінних і, зокрема, відповідні умовні та граничні щільності, які дозволяють не лише спільні прогнози, але й ймовірнісні інтервали. Вплив перевантаженого трафіку також можна врахувати шляхом поєднання правил розподілу трафіку із запропонованою моделлю мережі Баєса.

Отже, метою роботи є за допомогою баєсівської бази знань розробити модель, що вирішувала б проблему прогнозування потоків трафіку. Створити соціально-орієнтований застосунок, який дозволяв би прогнозувати трафік.

Текстова частина роботи складається з чотирьох розділів.

У першому описується поняття баєсівської мережі, історія виникнення баєсівської мережі, аналіз методів побудови та структура баєсівської мережі.

У розділі другому досліджується предметна область.

У розділі третьому запропонована модель баєсівської мережі описана у двох версіях, одна спрощена версія, якої достатньо для вирішення проблеми оцінки, і більш повна версія, яка дозволяє оцінити загальний середній рівень потоку та транспортний потік.

У четвертому розділі опис застосунку реалізованого в рамках роботи.

## **Розділ 1. Теоретична частина**

У першому розділі детально описується певні аспекти баєсівських мереж. Розділ структурований наступним чином. У розділі 1.1 йдеться про історію терміна баєсівської мережі. Поняття баєсівської мережі розкрито у розділі 1.2 разом із деякими прикладами. Методи побудови баєсівської мережі наведено в розділі 1.3, після чого детальніше розглядаємо ручну побудову в розділі 1.3.1 та автоматичну побудову в розділі 1.3.2. У розділі 1.4 розповідається про структуру баєсівської мережі.

### **1.1 Історія виникнення баєсівської мережі**

Термін Баєсова мережа був введений Йудою Перлом у 1985 році, щоб підкреслити:

- Часто суб'єктивної природи вхідної інформації.
- Покладання на баєсове обумовлювання як основу для уточнення інформації.
- Відмінність між причинними та доказовими способами міркування.

В кінці 1980-х років праці Йуди Перла «Імовірнісне міркування в інтелектуальних системах» та Річарда Неаполітана «Імовірнісне міркування в експертних системах» підсумували властивості баєсових мереж та утвердили баєсові мережі як область дослідження.

### **1.2 Поняття баєсівської мережі**

У данній роботі вирішили зосередитися на спеціальному класі моделей під назвою Баєсові мережі. Вони є графічними моделями, це означає, що вони містять частину, яку можна зобразити як граф.

Бассівська мережа - це ймовірнісна графова модель, яка може бути використана для побудови моделей з даних. Їх також називають мережами Басса, мережами переконань та іноді причинними мережами. Бассівські мережі можуть бути використані для широкого спектру завдань, включаючи прогнозування, виявлення аномалії, діагностику, автоматизоване розуміння, міркування, прогнозування часових рядів та прийняття рішень в умовах невизначеності, класифікація, машинне навчання, розпізнавання образів тощо. Бассівські мережі є ймовірнісними, оскільки вони побудовані з розподілу ймовірностей, а також використовують закони ймовірності для прогнозування та виявлення аномалії, для міркувань та діагностики, прийняття рішень в умовах невизначеності та прогнозування часових рядів.

Бассівська мережа - це граф, який складається з вузлів і спрямованих посилок між ними. У багатьох бассівських мережах кожен вузол являє собою ймовірнісну змінну, таку як чийсь зріст, вік чи стать. Змінна може бути дискретна, наприклад,  $\text{гендер} = \{\text{жінка}, \text{чоловік}\}$  або може бути безперервною, наприклад, якась відстань. Ми називаємо вузли з більш ніж одною змінною багатозмінними вузлами.

Вузли та зв'язки утворюють **структуру бассівської мережі**, і ми називаємо це структурною специфікацією.

Сервер Басса підтримує безперервні змінні з умовними лінійними розподілами Гаусса. Це означає, що безперервні розподіли можуть залежати один від одного (є багатоваріантними), а також можуть залежати від однієї або декількох дискретних змінних.

Ребра додаються між вузлами, щоб вказати, що один вузол умовно залежить від іншого. Якщо зв'язку між двома вузлами не існує, це не означає, що вони умовно незалежні, оскільки вони можуть бути з'єднані через інші вузли.

Бассівська мережа представляє причинно-ймовірнісну залежність між набором випадкових величин, їх умовними залежностями, і вона забезпечує компактне **зображення спільного розподілу ймовірностей**. БМ складається з двох основних частин: **спрямованого ациклічного графа та наборів умовних розподілів ймовірностей**. Спрямований ациклічний граф - це граф, в якому відсутні орієнтовані цикли, тобто шляхи, що починаються і закінчуються в одній і тій самій вершині.

Якщо на графу існує причинно-ймовірнісна залежність між двома випадковими змінними, відповідні два вузли з'єднані спрямованим ребром, тоді як спрямований край від вузла А до вузла В вказує, що випадкова величина А викликає випадкову величину В. Оскільки спрямовані ребра являють собою статичну причинно-імовірнісну залежність, цикли у графі не дозволені. Для кожного вузла графа визначається умовний розподіл ймовірностей. Іншими словами, умовний розподіл ймовірності вузла (випадкова величина) визначається для кожного можливого результату попереднього причинного вузла.

Для ілюстрації розглянемо наступний тривіальний приклад. Припустимо, ми намагаємось увімкнути наш комп'ютер, але комп'ютер не запускається (спостереження / свідчення). Ми хотіли б дізнатися, яка з можливих причин комп'ютерних збоїв є більш імовірною. У цій спрощеній ілюстрації ми припускаємо лише дві можливі причини цього нещастя: збій електрики та несправність комп'ютера. Відповідний спрямований ациклічний граф зображений на рисунку 1.

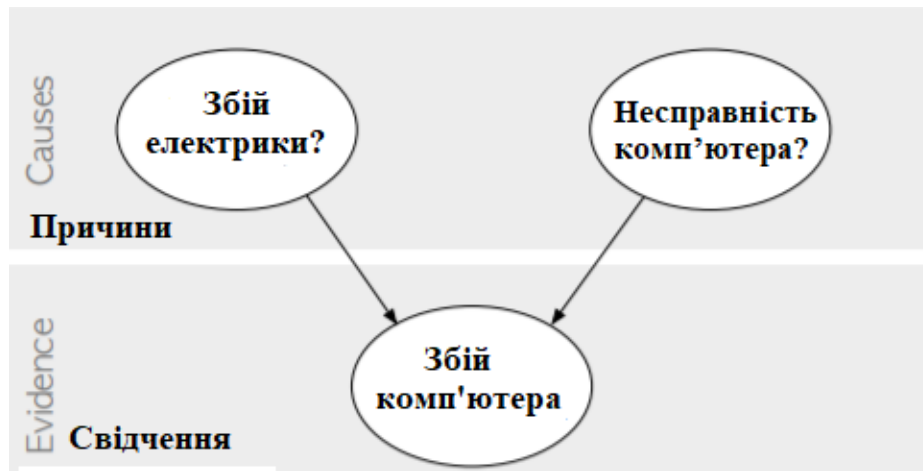


Рисунок 1: Спрямований ациклічний граф, що представляє дві можливі незалежні причини відмови комп'ютера.

Дві причини в цьому банальному прикладі вважаються незалежними (між двома причинними вузлами немає ребра. Якщо в графі немає циклу, баєсівські мережі здатні зафіксувати стільки причинно-наслідкових зв'язків, скільки їх потрібно для того, щоб достовірно описати ситуацію тієї частки зовнішнього середовища, для якої розв'язується задача.

Вся концепція баєсівських мереж побудована на **теоремі Баєса**, яка допомагає нам виразити умовний розподіл ймовірності причини з огляду на свідчення, використовуючи зворотну умовну ймовірність огляду свідчень з урахуванням причини:

$$P[\text{Cause} | \text{Evidence}] = P[\text{Evidence} | \text{Cause}] \cdot \frac{P[\text{Cause}]}{P[\text{Evidence}]}$$

Ми можемо побудувати повну ймовірнісну модель, лише вказавши умовний розподіл ймовірностей у кожному вузлі.

Повертаючись до нашого прикладу, ми припускаємо, що відмова електроенергії, позначена  $E$ , відбувається з вірогідністю 0,1,  $P[E = \text{так}] = 0,1$ , а несправність комп'ютера, позначена  $M$ , виникає з ймовірністю 0,2,  $P[M = \text{так}] = 0,2$ . Доцільно вважати збій електрики та несправність комп'ютера незалежними. Крім того, ми припускаємо, що якщо немає проблем з електрикою, а комп'ютер не працює, комп'ютер працює

нормально. Іншими словами, якщо  $C$  позначає збій комп'ютера, то  $P[C = \text{так} | E = \text{ні}, M = \text{ні}] = 0$ . Якщо немає проблеми з електрикою, але комп'ютер має несправність, ймовірність виходу з ладу комп'ютера становить 0,5,  $P[C = \text{так} | E = \text{ні}, M = \text{так}] = 0,5$ . Нарешті, якщо електроенергія вимкнеться, комп'ютер не запуститься незалежно від його потенційної несправності,  $P[C = \text{так} | E = \text{так}, M = \text{ні}] = 1$  і  $P[C = \text{так} | E = \text{так}, M = \text{так}] = 1$ . Ймовірність збою комп'ютера  $P[C = \text{так}]$  можна обчислити як

$$\begin{aligned} P[C = \text{так}] &= \sum_{E, M} P[C = \text{так}, E, M] \\ &= \sum_{E, M} (P[C = \text{так} | E, M] \cdot P[E] \cdot P[M]) \\ &= 0.19 \end{aligned}$$

Ми можемо вважати ймовірність  $P[C = \text{так}] = 0,19$  як апріорну ймовірність збою комп'ютера перед врахуванням свідчень. Графічна модель з апріорним розподілом ймовірностей, тобто перед спостереженням за будь-якими свідченнями, зображена на рисунку 2.



Рисунок 2: Спрямована графова модель.

Припустимо, що ми намагались увімкнути комп'ютер, але не вийшло. Іншими словами, ми спостерігаємо  $C = \text{ні}$  з вірогідністю 1 і задаємося питанням, як змінився розподіл ймовірностей відмови електроенергії  $E$  та

несправності комп'ютера  $M$  з огляду на спостережувані свідчення. Використовуючи формулу Баєса, знаходимо

$$\begin{aligned}
 P[E = \text{так} \mid C = \text{так}] &= \sum_M P[E = \text{так}, M \mid C = \text{так}] \\
 &= \sum_M \frac{P[C = \text{так} \mid E = \text{так}, M] \cdot P[E = \text{так}] \cdot P[M]}{P[C = \text{так}]} = 0.53 \\
 P[M = \text{так} \mid C = \text{так}] &= \sum_E P[E, M = \text{так} \mid C = \text{так}] \\
 &= \sum_E \frac{P[C = \text{так} \mid E, M = \text{так}] \cdot P[E] \cdot P[M = \text{так}]}{P[C = \text{так}]} = 0.58
 \end{aligned}$$

Графічна модель із апостеріорним розподілом ймовірностей, тобто після спостереження свідчення (відмова комп'ютера), зображена на рисунку 3.



Рисунок 3: Спрямована графова модель.

Зауважте, що спостережуваний збій викликав сильну залежність між спочатку незалежними можливими причинами; наприклад, якщо одна причина могла бути виключена, інша повинна мати місце. Тим не менш, наведені вище результати все ще не дуже корисні.

Припустимо розширення прикладу, включивши в модель ще один доказ, зокрема світловий збій  $L$ . Ми припускаємо, що світловий збій не залежить від несправності комп'ютера. Як і раніше, якщо електроенергію

вимкнено, світло не буде світити ні за яких обставин,  $P[L = \text{так} \mid E = \text{так}] = 1$ . Якщо немає проблем з електрикою, ми припускаємо, що все-таки є 0,2 шансу, що світло згасне (розбита лампочка),  $P[L = \text{так} \mid E = \text{ні}] = 0,2$ . Використовуючи той же алгоритм, що і раніше, отримуємо, що попередня ймовірність  $P[L = \text{так}] = 0,28$ . Розширена графічна модель з апіорним розподілом ймовірностей, тобто перед спостереженням за будь-яким свідченням, зображена на рисунку 4.



Рисунок 4: Спрямована графова модель.

На рисунку 5 показані зміни в апостеріорному розподілі ймовірностей після спостереження свідчень для всіх чотирьох комбінацій результатів відмови від світла та комп'ютера. Наприклад, якщо ми спостерігаємо як збій комп'ютера, так і світловий збій, тобто спостерігаємо як  $C = \text{так}$ , так і  $L = \text{так}$  з вірогідністю 1 (правий верхній граф на рисунку 5), отримуємо  $P[E = \text{так} \mid C = \text{так}, L = \text{так}] = 0,85$  і  $P[M = \text{так} \mid C = \text{так}, L = \text{так}] = 0,33$ . Спостереження, що і світло, і комп'ютер не працюють, значно збільшили ймовірність збою електроенергії. Збій комп'ютера, таким чином, можна пояснити. У трьох інших випадках принаймні один з приладів (світло та комп'ютер) працює, і тому ми можемо стверджувати, що з електроенергією точно немає нічого поганого. Якщо світло працює, але комп'ютер не запускається (нижній лівий граф на рисунку 5), ми точно

знаємо, що з електрикою немає нічого поганого, тому несправність комп'ютера є єдиним можливим поясненням збою комп'ютера.

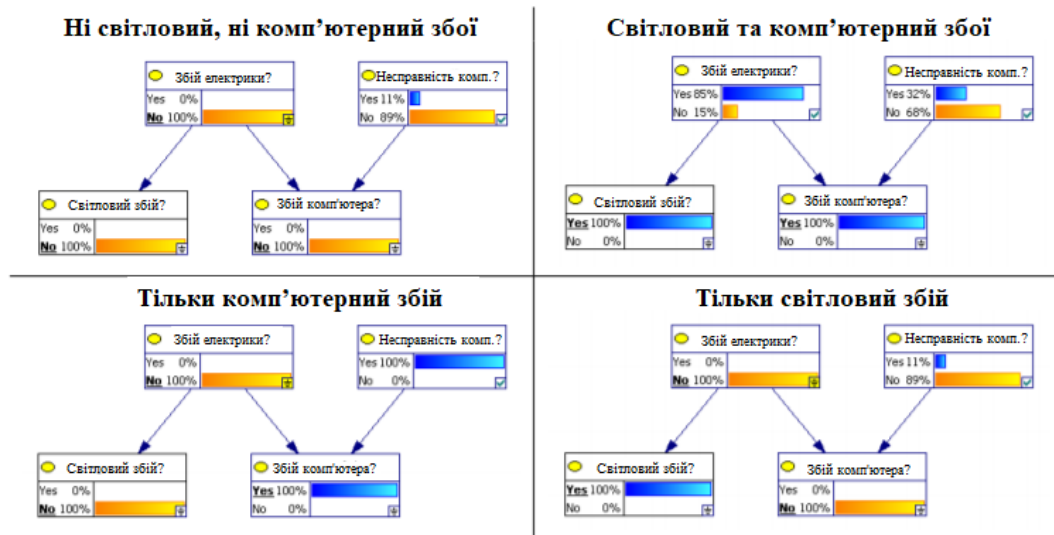


Рисунок 5: Спрямована графова модель.

На практиці баєсівські мережі значно складніші, використовують десятки чи навіть сотні вузлів. Важливо також зазначити, що кожен вузол у графі повинен бути пов'язаний принаймні з одним ребром до іншого вузла. В іншому випадку відокремлений вузол не залежить від усіх інших вузлів, і тому немає необхідності враховувати цей вузол.

Через їх візуальний вигляд, баєсівські мережі можуть плутати з марківськими моделями. Однак між цими двома поняттями є принципова різниця. Модель Маркова - приклад графу, який представляє лише одну випадкову змінну, і вузли представляють можливі реалізації випадкової змінної у різних часових точках. На противагу цьому, кожен вузол в баєсівській мережі являє собою одну випадкову змінну в одну мить. Іншими словами, моделі Маркова фіксують динаміку однієї випадкової величини, тоді як мережі Баєса фіксують статичний причинно-наслідковий зв'язок серед набору випадкових величин.

Баєсівські мережі можна використовувати як для якісного, так і кількісного моделювання, оскільки вони можуть поєднувати об'єктивні емпіричні ймовірності (частоти) з суб'єктивними оцінками. Баєсівські мережі також можуть бути використані як діаграми впливу замість дерев рішень. У порівнянні з деревами рішень, баєсівські мережі, як правило, більш компактні, простіші в побудові та їх зміні. Кожен параметр з'являється лише один раз у баєсівській мережі, і в разі потреби мережа може перетворитися на дерево рішень, тоді як зворотне не завжди можливо.

Основна слабкість полягає в тому, що баєсівські мережі вимагають апріорного розподілу ймовірностей, які можуть мати оманливий вплив на результати.

### **1.3 Аналіз методів побудови баєсівської мережі**

Існує два способи побудови баєсівської мережі: ручна побудова або автоматична побудова з баз даних. Обидва способи мають переваги та недоліки.

#### **1.3.1 Ручна побудова**

Ручна побудова баєсівської мережі передбачає попередні знання у цій галузі. Перший крок - побудова спрямованого ациклічного графа, після чого другий крок - оцінка умовного розподілу ймовірностей у кожному вузлі.

**Спрямований ациклічний граф:** побудова спрямованого ациклічного графа починається з ідентифікації відповідних вузлів (випадкових змінних) та структурної залежності між ними. Не всі змінні потрібно розглядати; насправді деякі випадкові величини можуть визначати непомічені величини, які, як вважають, впливають на результати спостереження. Дані, приховані змінні та параметри всі однаково розглядаються як вузли на графові. Баєсівський підхід заснований на тому,

що всі невідомі величини вважаються випадковими змінними, а отже, природно включати параметри як вузли у графі, так само як і приховані змінні та потенційно спостережувані величини. Наступний крок - накреслити мережу з урахуванням зв'язків між випадковими змінними. Структура графу, як правило, ґрунтується на предметних знаннях, хоча критика та перегляд моделі часто є важливими. Незважаючи на свою назву, баєсівські мережі не обов'язково передбачають вплив баєсівської статистики. Справді, для оцінки параметрів умовного розподілу ймовірностей зазвичай застосовують методи частот. Звичайно, можна реалізувати й баєсівський підхід використовуючи гіперпараметри замість параметрів умовного розподілу ймовірностей, що лежать в основі графа, які можна розглядати як вузли в моделі.

**Умовний розподіл ймовірностей:** побудований спрямований ациклічний граф повинен містити умовні розподіли ймовірностей для кожного вузла в графі. Якщо змінні дискретні, це може бути представлено у вигляді таблиці (мультиноміального розподілу), в якій перерахована ймовірність того, що дочірній вузол приймає кожне з різних його значень для кожної комбінації значень батьків. Якщо умовний розподіл ймовірності недоступний, для отримання цього умовного розподілу з даних (наприклад, емпіричного умовного розподілу ймовірності / частоти) можуть застосовуватися інші статистичні методи. На даний момент, Баєсова мережа повністю визначена. Однак перед використанням мережі в реальному застосуванні необхідно провести аналіз чутливості. Аналіз чутливості може виконуватися або як односторонній детермінований аналіз чутливості (тобто, змінюючи один параметр відразу у визначеному діапазоні), або як імовірнісний аналіз чутливості (тобто, змінюючи всі параметри мережі відразу за визначеним розподілом ймовірностей) .

### **1.3.2 Автоматична побудова**

На відміну від ручної побудови, автоматична не вимагає експертних знань в даній області. Баєсівські мережі можуть автоматично навчатися безпосередньо з баз даних, використовуючи алгоритми, засновані на досвіді, часто вбудовані у відповідне програмне забезпечення. Однак недоліком є те, що автоматична конструкція ставить більше вимог до даних. Крім того, повинно бути достатньо даних, щоб задовольнити алгоритм щодо надійних оцінок умовного розподілу ймовірностей. Для ручної побудови умовні розподіли ймовірностей вважаються завідома відомими. В той час як автоматична побудова передбачає як створення структури мережі, так і оцінку умовних розподілів ймовірностей.

### **1.4 Структура баєсівської мережі**

Основна мета структури баєсівської мережі - графічно узагальнити ряд умовних незалежних зв'язків. Простими словами, структура передбачає набір умовних незалежних зв'язків серед залучених змінних. Крім незалежностей, структура графів баєсівської мережі може також використовуватися в певних областях для представлення взаємозв'язку причинних ефектів через ребра та їх напрямки. У цих випадках батьки вузла вважаються "прямими причинами" кількості, представлені цим вузлом. Однак це справедливо лише за певних припущень, найбільш важливими є наступні:

- чи є, чи немає загальних неспостережуваних причин (змінних) двох або більше спостережуваних змінних. Якщо таких змінних немає, вважається, що властивість причинної достатності відповідає дійсності. Неспостережувані змінні називають також латентними або прихованими змінними.

- враховуючи причинну достатність, чи можливо для більш ніж однієї мережевої структури відповідати обмеженням. Ці обмеження є статистичними залежностями, які спостерігаються в даних.

Баєсовські мережі можуть використовуватися як причинно-наслідкові моделі у звичайній “причинно-наслідковій” інтерпретації, коли ми можемо сміливо припускати наведені вище припущення (що рідко, особливо для причинної достатності). В даний час існує великий інтерес та дослідження щодо виявлення причинної недостатності, наприклад, з використанням інструментальних змінних.

## **Висновки до розділу 1**

Баєсовські мережі є дуже природними інструментами. Зокрема, звичайні баєсовські мережі дуже прості та потужні. Вони можуть бути використані для широкого спектру завдань, включаючи прогнозування, виявлення аномалії, діагностику, автоматизоване розуміння, міркування, прогнозування часових рядів та прийняття рішень в умовах невизначеності.

## **Розділ 2. Дослідження предметної області**

У даній роботі за допомогою баєсівської бази знань буде створено модель, що передбачатиме трафік.

Для початку потрібно розглянути проблему розподілу трафіку детальніше, щоб зрозуміти які риси можуть бути виділені як важливі для моделі, та як їх можна формалізувати. Після цього, розглядається також задача матричного оцінювання та комбіновані методи.

### **2.1 Проблема розподілу трафіку**

Існує багато альтернатив, щоб знайти потоки зв'язків, враховуючи потоки джерело-призначення. Однак проблема розподілу трафіку традиційно розглядалася в основному двома різними методами: моделлю рівноваги користувача і моделлю стохастичної рівноваги користувача.

У цій роботі використовується підхід стохастичної рівноваги користувача, оскільки запропонований метод не дуже добре застосовується до моделі рівноваги користувача через його проблему не унікальності. Однак проблему рівноваги користувача можна модифікувати, щоб отримати унікальність.

Проблема рівноваги користувача базується на припущенні, що всі водії мають досконале знання мережі і, отже, часу в дорозі, і вони вибирають маршрути на основі цієї інформації.

Метод стохастичної рівноваги користувача складається з розгляду як випадкових варіацій у сприйнятті та перевагах водіїв. Оскільки всі ці елементи вважаються випадковими, користувачі вибирають відповідні маршрути випадковим чином. Іншою цікавою моделлю є модель

стохастичного призначення, яка має перевагу в тому, що не вимагає перерахування маршрутів.

Стохастичний термін використовується, оскільки передбачається, що користувачі сприймають час у випадковій формі.

Різні автори вважали статистичні розподіли сприйнятого часу поїздки за маршрутом належним до різних сімей. Наприклад, Vardi (1996) і Hazelton (2000) розглядали потоки джерело-призначення як незалежні випадкові величини Пуассона, хоча перший пізніше використовує нормальне наближення.

Еквівалентні математичні формулювання в термінах задач оптимізації були отримані для багатьох проблем трафіку моделі рівноваги користувача та стохастичної рівноваги користувача. Коли еквівалентні математичні функції знаходяться в замкнутій формі, а отримані цільові функції є опуклими, можна використовувати більш ефективні методи розв'язання.

## **2.2 Задача матричного оцінювання**

У задачі оцінки матриці поїздок намагаються оцінити матриці джерело-призначення на основі деяких спостережуваних потоків каналів. На жаль, це недостатньо визначена проблема через те, що кількість пар джерело-призначення зазвичай набагато більша за кількість зв'язків, і існує нескінченна множина рішень, які задовольняють закони збереження. Для того, щоб мати унікальне рішення, потрібно надати більше інформації, і вона повинна бути такою, щоб отримане рішення було близьке до фактичного чи реального. Зазвичай люди використовують попередню матрицю джерело-призначення, яка може надходити з багатьох різних джерел, включаючи стару застарілу або суб'єктивно вгадану матрицю

джерело-призначення, іншу матрицю джерело-призначення, отриману альтернативним методом тощо. На основі спостережень і цього попередньо, оцінка матриці джерело-призначення може бути отримана багатьма різними методами:

1. Метод найменших квадратів і узагальнений метод найменших квадратів. Ці методи мінімізують суму квадратів відмінностей між прогнозованою та попередньою матрицями поїздок джерело-призначення, коригуючи дисперсії кожного потоку, а в деяких випадках вони додають деякі обмеження. Хороші властивості функції суми квадратів і її зв'язку з нормальним розподілом роблять ці методи дуже зручними зі статистичної точки зору.

2. Ентропійні або інформаційні методи. Як і в методах найменших квадратів, вони оптимізують цільову функцію з урахуванням обмежень, які намагаються виміряти, наскільки розумними та близькими до реальності є різні матриці джерело-призначення. У цьому випадку обрана міра пов'язана з поняттям ентропії або кількості інформації. Таким чином, вони максимізують ентропію або мінімізують інформацію щодо попередньої матриці.

3. Статистичні методи. Вони являють собою найбільшу групу, і серед їхніх головних переваг одна з найважливіших полягає в мінливості інформації, яку вони надають щодо оцінок потоку. Зазвичай інші методи дають лише точкову інформацію, тобто конкретні значення джерело-призначення і потоків зв'язку, але не містять інформації про випадкову змінність, тоді як статистичні методи забезпечують відповідні інтервали ймовірності чи достовірності або регіони. Їх можна розділити на дві групи:

(а) Класичні методи. Потоки трафіку передбачають багатовимірні випадкові величини, що належать заданим параметричним сімействам, які задовольняють деяким зручним властивостям (відтворюваність, нескінченна подільність, позитивний діапазон тощо). Потім проблема

зводиться до оцінки параметрів і стає стандартною статистичною задачею, використовуючи багаторічний статистичний досвід.

(b) Баєсовські методи. Це добре відомі статистичні методи, які також розглядають параметричні сімейства розподілів, але параметри яких самі розглядаються як випадкові величини.

## **2.3 Комбіновані методи**

Як зазначено вище, якщо проблеми моделі рівноваги користувача і моделі стохастичної рівноваги користувача розглядаються окремо, зазвичай виникають деякі неузгодженості в розв'язках обох проблем. Таким чином, зручно розглядати їх разом, що призводить до нової проблеми, в якій загальна матриця потоку поїздок, пропорції вибору маршруту та неспостережувані потоки каналів стають результатом. Ці проблеми нещодавно були вирішені за допомогою дворівневих підходів, у яких верхній рівень вирішує проблему моделі рівноваги користувача, а нижній рівень має справу з проблемою моделі стохастичної рівноваги користувача.

Насправді будь-який існуючий метод розв'язання задач рівноваги користувача і моделі стохастичної рівноваги користувача можна використовувати для нижнього і верхнього рівнів відповідно.

Проблема рівноваги користувача визначає потоки джерело-призначення для заданих пропорцій вибору маршруту, а задача стохастичної рівноваги користувача визначає потоки каналів або пропорції вибору маршруту для заданих потоків джерело-призначення. Таким чином, обидві проблеми пов'язані, і необхідно використовувати методи декомпозиції. Під методами декомпозиції мається на увазі: розширений Лагранж, розклад Бендерса, релаксацію Лагранжа, методи розсікаючої гіперплощини або будь-яку іншу дворівневу техніку.

У роботі маємо справу з проблемою прогнозування джерело-призначення і потоку каналів зв'язку та оновлення, коли дані трафіку каналів стають доступними за допомогою баєсівських мереж. Це означає, що доступні умовні розподіли потоків каналу джерело-призначення, а потім можна отримати інтервали ймовірності або області, щоб дати уявлення про пов'язані невизначеності.

## **Висновки до розділу 2**

Існує багато альтернатив, щоб знайти потоки зв'язків, враховуючи потоки джерело-призначення. Однак проблема розподілу трафіку традиційно розглядалася в основному двома різними методами: моделлю рівноваги користувача і моделлю стохастичної рівноваги користувача. У роботі маємо справу з проблемою прогнозування джерело-призначення і потоку каналів зв'язку та оновлення, коли дані трафіку каналів стають доступними за допомогою баєсівських мереж.

### Розділ 3. Запропонована модель прогнозування трафіку

У цьому розділі представлено запропонований метод прогнозування трафіку разом із детальним описом та обґрунтуванням його основних припущень.

Основна ідея полягає в тому, щоб прийняти джерело-призначення і потоки зв'язку як багатовимірну випадкову величину, здатну відтворити свою реальну поведінку, і побудувати мережу Баєса Гауса з використанням спеціальних характеристик змінних потоку трафіку. З цією метою розглядаємо потоки джерело-призначення як батьків, а потоки зв'язку як синів і відтворюємо обмеження рівноваги в точній або статистичній (тобто з випадковими помилками) формі. Оскільки умовні ймовірності форми, дуже легко отримати, пропонуємо використовувати баєсівську мережу для відтворення ймовірнісної структури та пов'язаного з нею методу поширення доказів.

#### 3.1 Побудова простої моделі в практичних випадках

Нормальну баєсівську модель можна побудувати, якщо знати такі дані:

1. Топологія мережі, тобто матриця інцидентності  $\delta_{ak}^i$ .
2. Матриця  $\mu_{(T,V)}$  середніх значень усіх змінних, включаючи  $T$  і  $V$ , і матриця дисперсії–коваріації всіх змінних  $\Sigma_{(T,V)}$ , або альтернативно:
  - а. Очікуване значення  $E[U]$  та його стандартне відхилення  $\sigma(U)$ .

- b. Матриця  $K$  відносних ваг пар джерело-призначення відносно загального транспортного потоку (включаючи всі джерело-призначення пари).
- c. Матриця  $\beta$  коефіцієнтів регресії змінних  $V$ , заданих змінними  $T$ , або матриця  $\rho_{ik}$  пропорцій вибору маршруту.
- d. Дисперсії  $\varepsilon$  незалежних випадкових величин, які дозволяють записати діагональну дисперсію – коваріаційна матриця  $D_\varepsilon$ .
- e. Матриця дисперсії–коваріації  $D_\eta$  випадкових змінних.

### 3.2 Повний варіант моделі

У цьому розділі включаємо всі змінні в модель на випадок, якщо хтось забажає виконати подальший аналіз. Це дозволяє оцінити змінні  $U, \eta_i; i = 1, 2, \dots, m$ ; і  $\varepsilon_i; i = 1, 2, \dots, n$ , тобто оцінити  $U$  та умови помилки, які в деяких випадках можуть бути корисними. В повній моделі, яка розглядається усі змінні (включаючи змінні помилок), тоді як у простій моделі лише джерело-призначення та потоки зв'язку включені.

Структура дисперсійно-коваріаційної матриці ускладнюється, але в деяких випадках необхідно мати таке ускладнення, коли потрібно контролювати всі змінні.

Цілі набори випадкових величин, задіяних у нашій моделі, пов'язані лінійним виразом:

$$\begin{pmatrix} U \\ - \\ \boldsymbol{\eta} \\ - \\ \boldsymbol{\varepsilon} \\ - \\ \mathbf{T} \\ - \\ \mathbf{V} \end{pmatrix} = \begin{pmatrix} 1 & | & \mathbf{0} & | & \mathbf{0} \\ - & + & - & + & - \\ \mathbf{0} & | & \mathbf{I} & | & \mathbf{0} \\ - & + & - & + & - \\ \mathbf{0} & | & \mathbf{0} & | & \mathbf{I} \\ - & + & - & + & - \\ \mathbf{K} & | & \mathbf{I} & | & \mathbf{0} \\ - & + & - & + & - \\ \boldsymbol{\beta}\mathbf{K} & | & \boldsymbol{\beta} & | & \mathbf{I} \end{pmatrix} \begin{pmatrix} U \\ - \\ \boldsymbol{\eta} \\ - \\ \boldsymbol{\varepsilon} \end{pmatrix}$$

з чого випливає, що середнє  $E[(\mathbf{U}, \boldsymbol{\eta}, \boldsymbol{\varepsilon}, \mathbf{T}, \mathbf{V})]$  та матриця дисперсії–  
коваріації всіх змінних  $\boldsymbol{\Sigma}_{(\mathbf{U}, \boldsymbol{\eta}, \boldsymbol{\varepsilon}, \mathbf{T}, \mathbf{V})}$  є

$$\begin{pmatrix} E(U) \\ - - - \\ E(\boldsymbol{\eta}) \\ - - - \\ E(\boldsymbol{\varepsilon}) \\ - - - \\ E(\mathbf{T}) \\ - - - \\ E(\mathbf{V}) \end{pmatrix} = \begin{pmatrix} E(U) \\ - - - - - \\ \mathbf{0} \\ - - - - - \\ E(\boldsymbol{\varepsilon}) \\ - - - - - \\ E(U)\mathbf{K} \\ - - - - - \\ E(U)\boldsymbol{\beta}\mathbf{K} + E(\boldsymbol{\varepsilon}) \end{pmatrix}$$

Та

$$\Sigma = \left( \begin{array}{c|c|c|c|c} \sigma_U^2 & \mathbf{0} & \mathbf{0} & \sigma_U^2 \mathbf{K}^T & \sigma_U^2 \mathbf{K}^T \boldsymbol{\beta}^T \\ \hline \mathbf{0} & \mathbf{D}_\eta & \mathbf{0} & \mathbf{D}_\eta & \mathbf{D}_\eta \boldsymbol{\beta}^T \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{D}_\varepsilon & \mathbf{0} & \mathbf{D}_\varepsilon \\ \hline \sigma_U^2 \mathbf{K} & \mathbf{D}_\eta & \mathbf{0} & \sigma_U^2 \mathbf{K} \mathbf{K}^T + \mathbf{D}_\eta & (\sigma_U^2 \mathbf{K} \mathbf{K}^T + \mathbf{D}_\eta) \boldsymbol{\beta}^T \\ \hline \sigma_U^2 \boldsymbol{\beta} \mathbf{K} & \boldsymbol{\beta} \mathbf{D}_\eta & \mathbf{D}_\varepsilon & \boldsymbol{\beta} (\sigma_U^2 \mathbf{K} \mathbf{K}^T + \mathbf{D}_\eta) & \boldsymbol{\beta} (\sigma_U^2 \mathbf{K} \mathbf{K}^T + \mathbf{D}_\eta) \boldsymbol{\beta}^T + \mathbf{D}_\varepsilon \end{array} \right).$$

### 3.3 Використання моделі для прогнозування транспортних потоків

Коли створили модель, то можемо використовувати її

$$f(t_1, t_2, \dots, t_m, v_1, v_2, \dots, v_n) = f_{N(\mu_T, \Sigma_T)}(t_1, t_2, \dots, t_m) \prod_{a=1}^n p(v_a | \pi_a)$$

для прогнозування пар джерело-призначення і зв'язування потоків трафіку, коли деяка інформація стає доступною. Ідея полягає у використанні спільного розподілу пар джерело-призначення і потоків трафіку каналів, залежно від наявної інформації. Насправді, оскільки решта змінних (ті, що невідомі) є випадковими, найбільш інформативним елементом, який ми можемо отримати, є його умовний спільний розподіл, і це те, що дає методологія мережі Баєса. Однак, як правило, хтось зацікавлений у наданні точкових прогнозів та/або інтервалів ймовірності для прогнозів. З цією метою можемо використовувати умовні середні (моди) невідомих змінних як прогнози, тобто

$$\arg \text{Maximize}_{t_i \in \bar{T}_0; v_a \in \bar{V}_0} f_{N(\mu_T, \Sigma_T)}(t_1, t_2, \dots, t_m) \prod_{a=1}^n p(v_a | \pi_a) \Big|_{t_i \in T_0; v_a \in V_0},$$

Де  $T_0$  і  $V_0$  є підмножинами спостережуваних  $T$  і  $V$ , відповідно, і граничні нормальні умови для ймовірнісних інтервалів.

### **Висновки до розділу 3**

У цьому розділі представлено запропонований метод прогнозування трафіку разом із детальним описом та обґрунтуванням його основних припущень.

Основна ідея полягає в тому, щоб прийняти джерело-призначення і потоки зв'язку як багатовимірну випадкову величину, здатну відтворити свою реальну поведінку, і побудувати мережу Баєса Гауса з використанням спеціальних характеристик змінних потоку трафіку. З цією метою розглядаємо потоки джерело-призначення як батьків, а потоки зв'язку як синів і відтворюємо обмеження рівноваги в точній або статистичній (тобто з випадковими помилками) формі.

## Розділ 4. Практична частина

У рамках даної роботи була реалізована запропонована модель для прогнозування транспортних потоків. Ціль даної роботи – на базі розробленого додатку продемонструвати роботу моделі для прогнозування транспортних потоків, який надалі можна було б використовувати для подальших досліджень в цій галузі.

Дана частина роботи описує технічні деталі та особливості реалізації алгоритму в певному середовищі.

### 4.1 Опис реалізації

Ідея полягає у використанні спільного розподілу пар джерело-призначення і потоків трафіку каналів, залежно від наявної інформації. Насправді, оскільки решта змінних (ті, що невідомі) є випадковими, найбільш інформативним елементом, який ми можемо отримати, є його умовний спільний розподіл, і це те, що дає методологія мережі Баєса. Однак, як правило, хтось зацікавлений у наданні точкових прогнозів та/або інтервалів ймовірності для прогнозів. З цією метою можемо використовувати умовні середні (моди) невідомих змінних як прогнози, тобто

$$\arg \underset{t_i \in \bar{T}_0; v_a \in \bar{V}_0}{\text{Maximize}} \quad f_{N(\mu_T, \Sigma_T)}(t_1, t_2, \dots, t_m) \prod_{a=1}^n p(v_a | \pi_a) \Big|_{t_i \in T_0; v_a \in V_0},$$

Де  $T_0$  і  $V_0$  є підмножинами спостережуваних  $T$  і  $V$ , відповідно, і граничні нормальні умови для ймовірнісних інтервалів.

#### Опис основних файлів програми:

TrafficPredictions.java – файл за допомогою якого реалізується запропонована модель прогнозування трафіку.

LocationProcess.java – файл за допомогою якого відбувається прорахунок відстані.

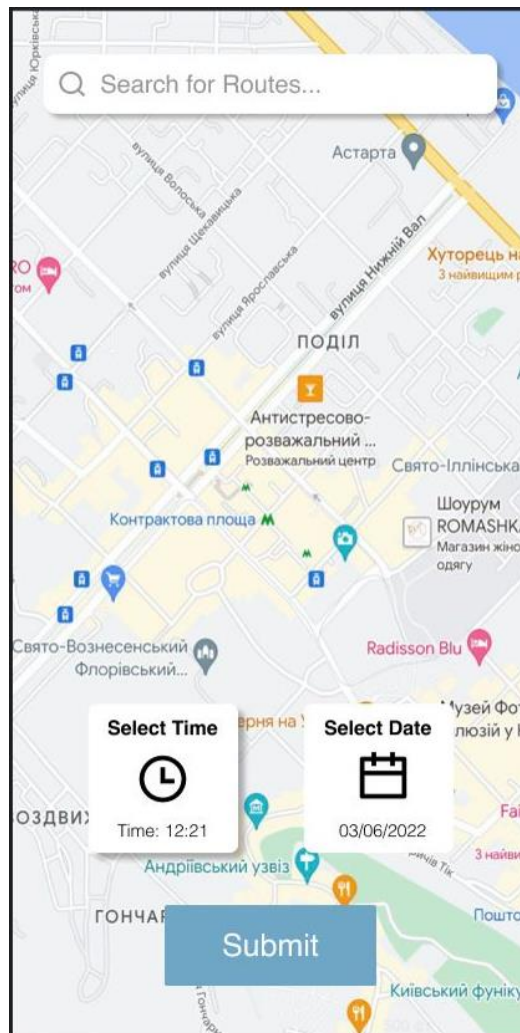
MainAppController.java – файл, який відповідає за роботу всього застосунку.

Коли застосунок завантажується, то видно таке вікно.



*Рисунок 6: Вікно завантаження застосунку.*

Головне вікно застосунку:



*Рисунок 7: Головне вікно застосунку.*

Вікно пошуку потрібної адреси:

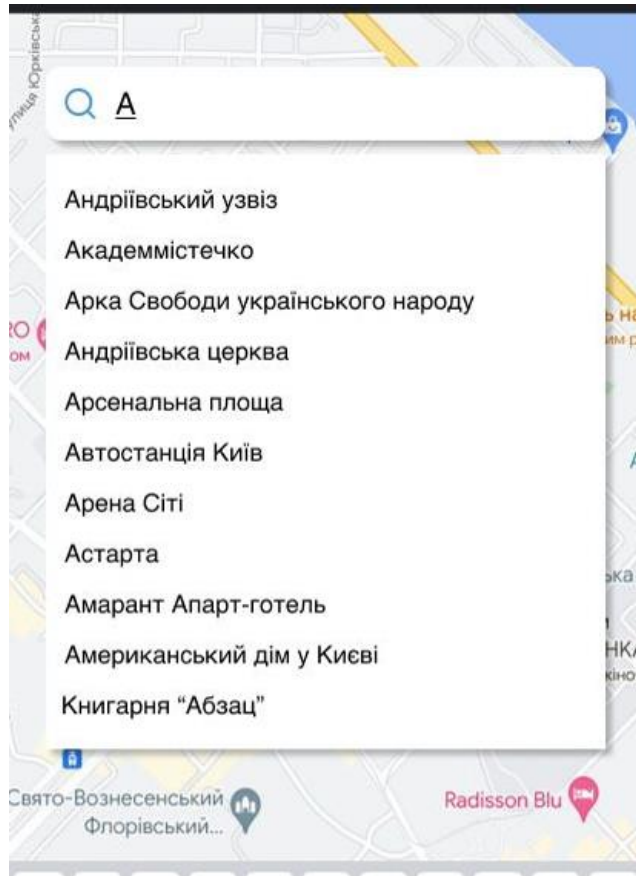


Рисунок 8:Вікно пошуку застосунку.

Статистика:

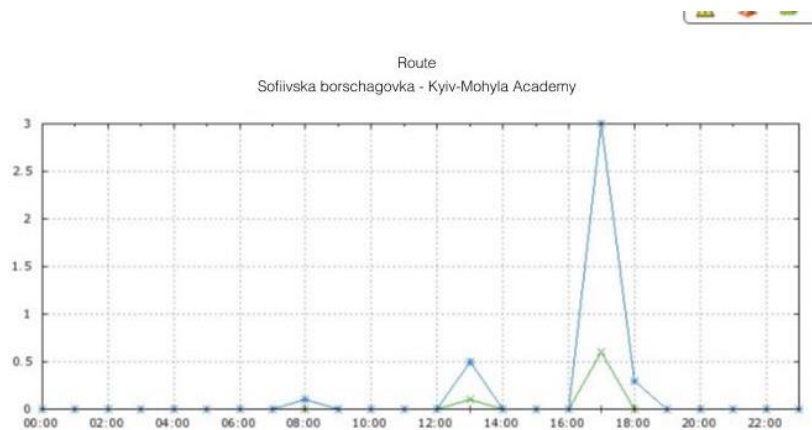


Рисунок 9:Статистика.

## 4.2 Використані технології

- AndroidStudio 3.4.2

- Gradle 5.1.1

#### **Висновки до розділу 4**

Баєсовські мережі є дуже природними інструментами для відтворення випадкової структури залежностей потоків трафіку в парах джерело-призначення і потоків трафіку зв'язку. Зокрема, звичайні баєсовські мережі дуже прості та потужні, і їх параметри можна легко дізнатися з топології мережі та інших даних.

Докази, тобто знання парних потоків ланки та/або джерело-призначення, суттєво змінюють середні значення та зменшують дисперсію решти змінних, що призводить до більш точних прогнозів. Баєсовські мережі є потужними методами для прогнозування потоків трафіку.

## Висновки

Як перший внесок, надали ретельний огляд та опис літератури навчання баєсівської структури мережі. Другий внесок - розробка соціально-орієнтованого застосунку. Представлений метод баєсівської мережі для оцінки матриці джерело-призначення, у якому графічне представлення надходить безпосередньо з мережі трафіку. Баєсовські мережі є дуже природними інструментами для відтворення випадкової структури залежностей потоків трафіку в парах джерело-призначення і потоків трафіку зв'язку. Зокрема, звичайні баєсовські мережі дуже прості та потужні, і їх параметри можна легко дізнатися з топології мережі та інших даних.

Докази, тобто знання парних потоків ланки та/або джерело-призначення, суттєво змінюють середні значення та зменшують дисперсію решти змінних, що призводить до більш точних прогнозів. Баєсовські мережі є потужними методами для прогнозування потоків трафіку.

## Джерела

1. Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, California.
2. David Heckerman's tutorial on learning with Bayesian networks.
3. Ф. Дженсен. "Вступ до баєсівських мереж". UCL Press. 1996.
4. Anderson, T. W. (1984). *An introduction to multivariate statistical analysis*, John Wiley and Sons, New York. Castillo, E., Gutiérrez, J. M. & Hadi, A. S. (1997a). *Expert systems and probabilistic network models*, Springer Verlag, New York.
5. Castillo, E., Gutiérrez, J. M. & Hadi, A. S. (1997b). Sensitivity analysis in discrete Bayesian networks, *IEEE Transactions on Systems, Man and Cybernetics* 26(7): 412–423.
6. [uk.wikipedia.org/wiki/Баєсова\\_мережа](http://uk.wikipedia.org/wiki/Баєсова_мережа) – 09.11.2019.
7. Hazelton, M. L. (2003). Some comments on origin-destination matrix estimation, *Transportation Research* 37A: 811–822.
8. Mínguez, R., Sánchez-Cambronero, S., Castillo, E. & Jiménez, P. (2010). Optimal traffic plate scanning location for od trip matrix and route estimation in road networks, *Transportation Research B* 44: 282–298.
9. Praskher, J. N. & Bekhor, S. (2004). Route choice models used in the stochastic user equilibrium problem: a review, *Transportation Reviews* 24: 437–463
10. Moninder Singh and Marco Valtorta. Construction of bayesian network structures from data: a brief survey and an efficient algorithm. In *International Journal of Approximate Reasoning*, pages 259–265. Morgan Kaufmann, 1995.
11. Kohsuke Yanai and Hitoshi Iba. Estimation of distribution programming based on Bayesian network. In Ruhul Sarker, Robert Reynolds, Hussein

- Abbass, Kay Chen Tan, Bob McKay, Daryl Essam, and Tom Gedeon, editors, Proceedings of the 2003 Congress on Evolutionary Computation CEC2003, pages 1618–1625. IEEE Press, 8-12 December 2003.
12. <http://www.opensourceinitiative.net/edu/bnintro/> - 19.01.2020.
  13. С. Лаурицен. "Графічні моделі", Оксфорд. 1996. Остаточне математичне виклад теорії графічних моделей.
  14. Pedro Larrañaga, Mikel Poza, Yosu Yurramendi, Roberto H. Murga, and Cindy M.H. Kuijpers. Structure learning of bayesian networks by genetic algorithms: A performance analysis of control parameters. IEEE Transactions on Pattern Analysis and Machine Intelligence, 18:912–926, 1996.
  15. J. M. Peña, J. A. Lozano, and P. Larrañaga. An improved bayesian structural em algorithm for learning bayesian networks for clustering. Pattern Recogn. Lett., 21(9):779–786, July 2000.
  16. M. Teyssier and D. Koller. Ordering-based search: A simple and effective algorithm for learning bayesian networks. In Proceedings of the Twenty-first Conference on Uncertainty in AI (UAI), pages 584–590, Edinburgh, Scotland, UK, July 2005
  17. M. Papageorgiou, J-M. Blosseville and H. Haj-Sale, “Dynamic modeling, assignment, and route guidance in traffic networks”, Transportation Research part B, vol. 24B, n° 6, pp. 471-495, 1990.
  18. D. Middleton, D. Gopalakrishna and M. Raman, “Advances in traffic data collection and management”, Technical Report BAT- 02-006, Texas Transportation Institute, Cambridge systematics, Inc., Washngton, DC, December, 2002.
  19. D.H. Nam and D.R Drew, “Automatic measurement of traffic variables for intelligent transportation systems applications”, in Transportation Research Part B: Methodological, n° 33, pp. 437- 457, 1999.

20. J.S. Oh, R. Jayakrishan and W. Rechker, “ Section travel time estimation from point detection data”, in proceedings of the 82th Annual Meeting of Transportation Research Board, Washington, D.C., U.S.A, 2003.
21. F. Bonvalet and Y. Robin-Prévallée, “Mise au point d’un indicateur permanent des conditions de circulation en Île-de- France”, Transport Environnement Circulation, pp. 84-85, 1987.
22. C. Buisson and JB. Lessort, “Comprendre le trafic routier, méthodes et calculs”, Edition Certu, 2010.