

БАЙЄСІВСЬКА МЕРЕЖА В ІОТ НА ПРИКЛАДІ ПОБУДОВИ КМІТЛИВОЇ КУР'ЄРСЬКОЇ СЛУЖБИ

Розглянуто задачу побудови байєсівської мережі – орієнтованого ациклічного графа, з прорачуванням умовної та безумовної ймовірності змінних на прикладі оцінювання рейтингу водія за допомогою OBD-II давача.

Ключові слова: байєсівська мережа, орієнтований ациклічний граф, умовна ймовірність, безумовна ймовірність, IoT, навчаємий глибинний аналіз даних, оцінювання водіїв.

Для наочного прикладу застосування байєсівської мережі було розроблено модель оцінювання якості кур'єрської служби. Ця модель є певним IoT-проектом, де attach dongle пристрій під час поїздки на машині обробляє і аналізує телеметричні дані автомобіля для отримання оперативних відомостей і довгострокових тенденцій поїздки. При цьому пропонується з певного набору характеристик обчислити коефіцієнт навичок водія, який потім можна застосовувати з різною метою, наприклад, для побудови рейтингу кур'єрів, визначення більш точного часу доставки тощо. Прикладом виявлення безпосередньої залежності між зчитуваними даними автомобіля під час поїздки і якістю водіння можна вважати My Driving [5], де навчальна вибірка становила понад 9000 поїздок, а за допомогою їх аналізу можна було сегментувати водіїв на категорії.

Доцільність проекту

У сервісах, де покупець оцінює продавців, наприклад, Uber або Airbnb, згодом обов'язково відбувається «інфляція репутації». Річ у тім, що середні оцінки ростуть і наближаються до максимуму, тому виокремити поганих постачальників послуг уже неможливо. Користувачам просто незручно ставити погані оцінки іншим людям.

Рейтинги – найважливіший інструмент саморегулювання маркетплейсів. Наприклад, середній продавець eBay має 100 % позитивний рейтинг. На «Убер» і «Лифт» менше, ніж 5 зірок, вважається «поганим» зворотним зв'язком: майже 90 % поїздок UberX у Чикаго на початку 2017 року мали ідеальний 5-зірковий рейтинг [4].

Байєсівська мережа

Байєсівська мережа – це ациклічний орієнтований граф, у якому кожна вершина (вузол

мережі) становить n -значну змінну, де дуги означають існування безпосередніх причинно-наслідкових залежностей між змінними, а сила цих змінних кількісно виражається у вигляді умовних ймовірностей [7].

У загальному вигляді байєсівська мережа – це зручний інструмент для опису подій з невизначеністю. Основною ідеєю побудови мережі є розкладання складної системи на прості елементи. Для об'єднання окремих елементів у систему використовується математичний апарат теорії ймовірностей. Такий підхід забезпечує можливість будувати моделі з безліччю взаємодіючих змінних для подальшого розроблення ефективних алгоритмів обробки даних і ухвалення рішень [2].

Зручність такої мережі – в наочності подання її у вигляді графа, що мінімізує помилку логічного проектування залежностей, та високої працездатності в абсолютно різних галузях.

Наприклад, у статті [9] показано байєсівську мережу вимірювання рівня компетенції студентів за допомогою їхніх відповідей на тестові завдання.

Моделювання мережі

Моделювання потрібно починати з визначення змінних, які безпосередньо належать до нашої галузі. У цьому разі змінними є показники давачів автівки. Визначимо класи цих змінних залежно від їхньої ролі [1].

1. *Цільові змінні.* Ці змінні використовуються для моделювання того, що становить для нас інтерес. Зазвичай цільові змінні відображають приховані (латентні) характеристики. Це означає, що їх неможливо виміряти безпосередньо, а тільки за допомогою певних пристроїв.

2. *Спостереження.* Їх використовують для того, щоб надати інформацію щодо цільових змінних. Тобто це висновки з тих даних, що ми виміряли.

3. *Фактори*. Це змінні, які моделюють джерела впливу на цільову змінну. Їх також називають контекстними змінними. Фактори ділять на чотири категорії відповідно до їхнього впливу на змінну: промоутери, сповільнювачі, вимоги, винятки. У статті [7] здійснено опис цих категорій чинників. Загалом, фактори – це певні змінні, які надають корисну інформацію та враховуються в підсумковій моделі незалежно від показників дачачів.

4. *Допоміжні змінні*. Ці змінні використовуються для зручності. Наприклад, якщо у вузла є багато батьківських вузлів, проміжні допоміжні змінні можна використати для їх групування. Завдяки цьому структура мережі спрощується, а кількість параметрів зменшується.

Опис змінних для моделювання мережі

У нашому прикладі головним дачачем значень змінних є **вбудована діагностика** (далі – OBD-II) – пристрій для налаштування автомобіля та діагностики непарних шумів і попереджувальних ламп. За допомогою нього можна отримати показники ефективності водіння та здійснити певні обчислення. Пристрій OBD можна придбати в магазинах, він працює за допомогою Bluetooth або Wi-Fi у додатку телефону.

Пристрій OBD-II може давати безліч параметрів, але ми обрали ті, що можуть надати нам доцільну інформацію для врахування в рейтингу (табл. 1).

Визначення структури мережі

Після визначення змінних потрібно побудувати структуру самої мережі. Її визначають відношенням між змінними за допомогою

направлених ребер. Як уже зазначалось раніше, байєсівська мережа – це ациклічний граф, де ребра є спрямованими і зміна напрямку ребра має значення. Сенс ребра полягає в тому, що змінна у вихідній вершині безпосередньо впливає на змінну в цільовій. У деяких джерелах байєсівську мережу ще називають причинно-наслідковою (casual).

Визначення параметрів

Розподіли для кореневих вузлів задаються апріорно, за допомогою експертної оцінки, або отримуються з даних. Це безумовне розподілення, яке не змінюється в подальшому, незалежно від отриманих показників. Для всіх інших вузлів визначається умовна ймовірність.

Як ми бачимо на рисунку, з обраних характеристик ми експертно позначили розподілення цільових змінних та визначили структуру графа. Для розподілення умовних імовірностей інших змінних потрібно використати теорему Байєса, найточніше зробити це можна, наприклад, використавши бібліотеку gRain, на мові R, попередньо побудувавши байєсівську мережу (задавши апріорні ймовірності та відношення між ними). Приклад побудови такої програми наведено у статті [3].

У нашому прикладі ми розглянемо лише один зв'язок із нашої мережі (див. табл. 2): А – сила натискання на газ, В – дистанція, подолана з увімкненою лампочкою про проблеми з двигуном, С – швидкість, a1 – рейтинг водія «плюс», a2 – рейтинг водія «мінус». У вершинах А, В змінна може набувати значень «висока» або «норма», а в змінній С – «вища за дозволену», «нижча за дозволену», «норма».

Таблиця 1. Дані з OBD-II (On-Board Diagnostics Parameter Identification)

Ідентифікатор параметра	Назва	Одиниці вимірювання	Показники норми
04	Навантаження на двигун	Н/м	>200 – високе 100–200 – норма
0C	Оберти двигуна	об/хв	500–1000 – низькі >3000 – високі 1000–3000 – норма
0D	Швидкість транспортного засобу	км/год	
10	Масова витрата повітря	л/с	<10 – низька >15 – висока 10–15 – норма
11	Положення дросельної заслінки	градуси	>10 – високе 3–10 – норма
21	Дистанція, подолана з увімкненою лампочкою «перевір двигун»	км	>10 – забагато 0–10 – норма
5E	Витрата палива	л/100 км	>10 – висока <10 – норма

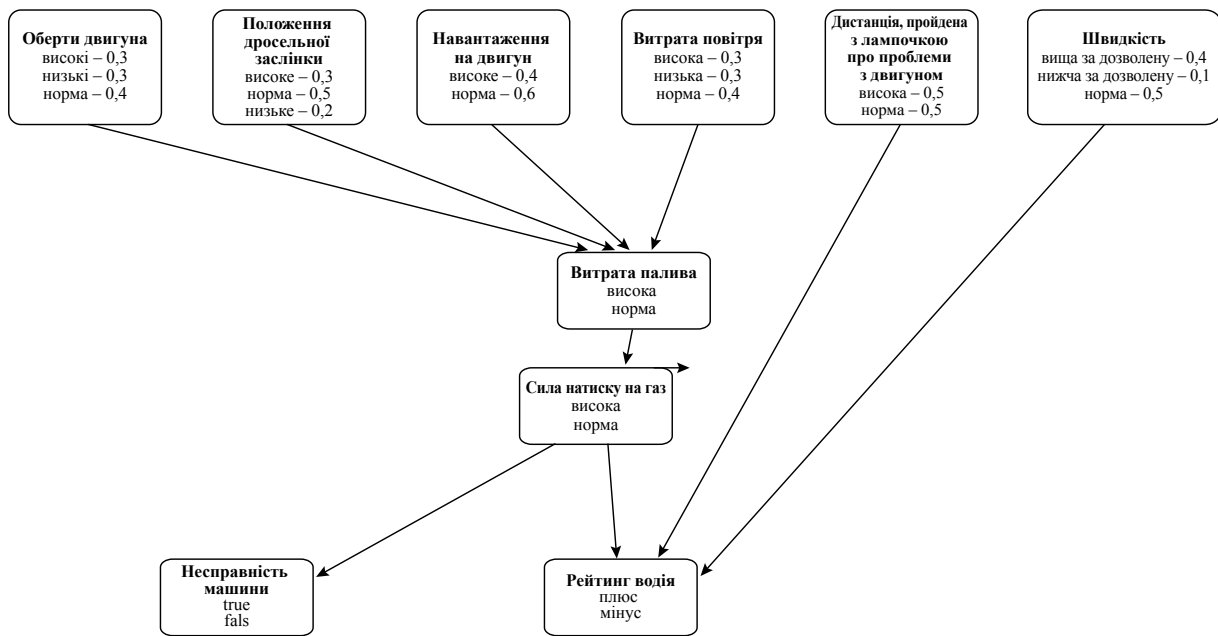


Рисунок. Побудована байєсівська мережа для IoT проекту

Таблиця 2. Показник «Рейтинг водія»

Сила натискання на газ	Дистанція, подолана з увімкненою лампочкою	Швидкість	Рейтинг водія – плюс	Рейтинг водія – мінус
висока	висока	висока	0	1
висока	висока	норма	0,4	0,6
висока	низька	норма	0,6	0,4
низька	висока	норма	0,7	0,3
низька	висока	висока	0,3	0,7
низька	низька	висока	0,6	0,4
низька	низька	норма	1	0
висока	низька	висока	0,1	0,9
висока	висока	низька	0	1
висока	низька	низька	0,2	0,8
низька	висока	низька	0,9	0,1
низька	низька	низька	0,5	0,5

На початку змінні розподілені рівномірно, тобто всі значення мають однакову ймовірність. Для того щоб більш точно визначити розподіл, потрібно скористатися даними, отриманими з навчального процесу, в нашому випадку визначимо це експертно, спираючись на такі характеристики:

1. За нормальної сили натискання на газ та одночасно високої витрати палива можна казати про несправність машини, тож у всіх інших варіантах машина працює нормально.

2. Якщо водій проїхав певну дистанцію з увімкненою лампочкою попередження про проблеми з двигуном, то це можна розглядати як ігнорування проблеми з машиною та погіршення рейтингу.

3. Враховується дозволена швидкість на подоланій ділянці для визначення, чи їхав водій за правилами.

Тобто, якщо змінна А буде низькою, В – високою, а С – низькою, то з ймовірністю 0,9 ми можемо сказати, що це кваліфікований водій.

Такі самі таблиці побудовано для показників «Витрата палива», «Сила натискання на газ» та «Несправність машини».

Висновки

У статті розглянуто алгоритм побудови мережі Байєса. Описано послідовність дій побудови мережі на прикладі «кмітливої кур'єрської служби» з урахуванням показників давача OBD-II для визначення рейтингу водія. Цю технологію можна застосовувати не тільки як критерій оцінювання працівників кур'єрської служби, а й у будь-якій галузі, де важлива якість водіння: таксі, особисті водії, далекобійники тощо.

Список використаної літератури

1. Атанов Г. А. Обучение и искусственный интеллект, или Основы современной дидактики высшей школы / Г. А. Атанов, И. Н. Пустынникова. – Донецк : Изд-во ДООУ, 2002. – 504 с.
2. Бидюк П. И. Построение и методы обучения байесовских сетей / П. И. Бидюк, А. Н. Терентьев // Таврійський вісник інформатики і математики. – 2004. – № 2. – С. 139–154.
3. Хлопотов М. В. Применение байесовской сети при построении моделей обучающихся для оценки уровня сформированности компетенций / М. В. Хлопотов // Науковедение : интернет-журнал. – 2014. – № 5/сентябрь – октябрь. – С. 19. – Режим доступа: <https://naukovedenie.ru/PDF/20TVN514.pdf>. – Загл. с экрана.
4. Athey S. Generalized Random Forests / Susan Athey, Julie Tibshirani, Stefan Wager // *Annals of Statistics*, to appear, 2018.
5. Filippas A. Reputation Inflation [Electronic resource] / A. Fillipas, J. Horton, J. Golden. – March 7, 2018. – Mode of access: <http://john-joseph-horton.com/papers/longrun.pdf>. – Title from the screen.
6. Microsoft Corporation, My Driving Reference Guide [Electronic resource]. – Redmond, Washington, 2016, p. 2. – Mode of access: <https://download.microsoft.com/download/F/E/4/FE484B73-061C-4171-B95B-D16A500BDAC6/MyDriving%20Reference%20Guide.pdf>. – Title from the screen.
7. Millán E. Bayesian networks for student model engineering / E. Millán, T. Loboda, J. L. Pérez-de-la-Cruz // *Computers & Education*. – 2010. – Vol. 55(4). – P. 1663–1683.
8. Pearl J. Probabilistic Reasoning in Expert Systems: Networks of Plausible Inference / J. Pearl. – San Francisco : Morgan Kaufmann, 1988. – 552 p.
9. VanLehn K. Student modeling from conventional test data: a Bayesian approach with out priors / K. VanLehn, Z. Niu, S. Siler, A. Gertner // *Proc. of 4th Int. Conf. ITS'96*. – 1996. – P. 29–47.

Ladyslava Klishchenko, Yuliia Kudriavtseva

BAYESIAN NETWORK IN IOT FOR CREATING A SMART DELIVERY SYSTEM

Bayesian network is an acyclic oriented graph in which each vertex (node of a network) represents an n -valued variable, where arcs mean the existence of direct causal relationships between variables, and the strength of these variables is quantified in the form of conditional probabilities. Bayesian networks are useful for processing data with uncertainties. In this article we define the necessary variables and build relations between them. Also we show how to count the probability of depending variables according to side factors. The information to consider can be defined by experts or based on the previous experience. Our structure of the Bayesian network represents the structure of competences of the driver.

Nowadays, many companies are based on reputation systems. In online marketplaces, “reputations” are typically calculated from the numerical feedback scores left by past trading partners [5]. However, according to human psychology, it becomes uncomfortable for people to put a low grade for any kind of facilities. Therefore, it is important to get an objective system for rating counting. We have designed a Bayesian network that provides us with a method of getting a score for driving experience. Getting raw data from OBD-II and analyzing it can help us to estimate the driver. Such a system can be used in any industry where we take care about the quality of driving: taxis, personal drivers, truckers, etc.

Creating a network should begin with the definition of variables. We define the classes of these variables. Depending on their role, they could be Target variables, Observations, Factors, and Auxiliary variables. Then we can build relations between them, using directed ribs. The meaning of the rib lies in the fact that the variable in the initial vertex directly affects the change in the target. Then, while counting the probabilities, we must include all possible factors. Finally, the main purpose of the network is to define the right components and the relations between them and to count conditional and unconditional probability for each variable. In our case, we have described the general scheme of analyzing data and working with the algorithm of building the Bayesian network. In conclusion, the purpose is to get the objective driver rating and implement it in other structures.

Keywords: Bayesian network, oriented acyclic graph, conditional probability, unconditional probability, IoT, Educational Data Mining, drivers evaluation.

Матеріал надійшов 09.04.2018