

Бойко Б.,
аспірант кафедри менеджменту та міжнародного підприємництва,
Національний університет "Львівська політехніка"

Процик І.,
кандидат економічних наук, доцент,
доцент кафедри менеджменту та міжнародного підприємництва,
Національний університет "Львівська політехніка"

ПОШУК АСОЦІАТИВНИХ ПРАВИЛ В СОЦІОЛОГІЧНИХ ДОСЛІДЖЕННЯХ ПРАЦІВНИКІВ: АНАЛІЗ ТА ПЕРСПЕКТИВИ

В роботі розглядаються результати опитувань серед працівників компаній з різних галузей. Аналізуються асоціації, які виникають у відповідях на питання про різні аспекти роботи, взаємодію з колегами, сприйняття керівництва тощо. Результати дослідження допомагають зрозуміти, які асоціації важливі для працівників, їхні відмінності в залежності від контексту та вплив на організаційні процеси.

Ключові слова: соціологічне опитування, респондент, Алгоритм пошуку асоціативних правил.

Основним видом пошуку закономірностей методами Data Mining є афінтивний аналіз. Ціль даного методу - пошук асоціативних правил, тобто дослідження взаємного зв'язку між подіями, що відбуваються спільно [1].

З допомогою соціологічних досліджень здійснюється подальше пізнання об'єктивних законів розвитку і функціонування соціальних спільнот працівників і визначаються шляхи і форми використання нагромаджених знань на практиці [2].

Для роботи обрані результати соціологічного опитування працівників компаній про благополуччя, соціальне відчуження, релігію, сприйняття дискримінації, національної та етнічної ідентичності. Одним з питань було суб'єктивне відчуття задоволеності роботою, тому основним питанням дослідження - що об'єднує щасливих (нещасливих) людей в різних соціальних групах на підприємствах.

Нехай є деякий датасет D , який складається з унікальних транзакцій itemset. Тоді $D = d_0, d_1, \dots, d_j$. Всередині кожної транзакції d представлений набір items. Наприклад: $d_1 = [\{partner\ 1:1\}, \{gndr\ 2:1\}, \{blgetmg\ 2:0\}, \{crmvct\ 2:1\}, \dots, \{\dots\}]$, $d_2 = [\{partner\ 1:0\}, \{gndr\ 2:1\}, \{blgetmg\ 2:0\}, \{crmvct\ 2:0\}, \dots, \{\dots\}]$. Прийнято кожен ітемсет описувати кількістю ненульових значень (k -itemset) [1, 3]. Наприклад для d_1 — 2-itemset, а для d_2 — 1-itemset.

Також існує ряд інших базових понять необхідних для розуміння

Алгоритму пошуку асоціативних правил [4].

1. *Support* (підтримка)

Нехай X - itemset, який має в собі i -items; T - кількість транзакцій.

Тоді: $supp(X) = \{t \in T; X \in t\} \div |T|$.

У такій формі це лише частота цього ітемсету у всіх транзакціях. Але це стосується лише цього X . В реальності потрібно наприклад x_1 і x_2 . Тоді формула набуває наступного вигляду: $supp(x_1 \cup x_2) = \frac{\sigma(x_1 \cup x_2)}{|T|}$,

де σ - кількість транзакцій, які мають x_1 і x_2 .

2. *Confidence* (достовірність) [3]

Це показник того, наскільки часто правило працює для всього датасету.

Обчислюється подібно до теореми Байєса: $conf(x_1 \cup x_2) = \frac{supp(x_1 \cup x_2)}{supp(x_1)}$.

3. *Lift* (підтримка) [3]

lift - це відношення «залежності» items до їх «незалежності». Lift показує, наскільки items залежать один від одного:

$$lift(x_1 \cup x_2) = \frac{supp(x_1 \cup x_2)}{supp(x_1) \times supp(x_2)} = \frac{confidence}{expected\ confidence}$$

4. *Conviction* (переконливість) [3]

Переконливість - це частота помилок правила:

$$conv(x_1 \cup x_2) = \frac{1 - supp(x_2)}{1 - conf(x_1 \cup x_2)}$$

Чим вище цей показник за 1 - тим краще. Наприклад, якщо $conv = 1,2$, то це означатиме, що ймовірність, що це правило в 1,2 рази вища, а ніж що це є випадковість [5].

Алгоритм пошуку асоціативних правил працює з даними у вигляді транзакцій, де в кожному рядку перелік ітемів, або з бінарними даними у вигляді розрідженої матриці, де вибрані ітеми позначені 1.

Для пошуку асоціативних правил використовуватимемо значення minimum support - мінімальна підтримка. Тобто для скількох відсотків респондентів справджуватиметься правило. Також вхідним параметром для функції вказуватимемо кількість ітемів - з скількох параметрів складатиметься правило.

В ході експериментів було отримано асоціативні правила для різної довжини ітемсетів (від 2 до 6, і кілька правил довжиною 7). Очевидно, що чим менша довжина, ти сильніші правила. Далі розберемо кілька правил для кожної довжини. При цьому варто враховувати, що кожне правило по замовчуванню довше на 1, адже пошук павил був між працівниками, які задоволені роботою.

Розглянемо розбір правил різної довжини:

1. *Правила довжиною 2.*

Загальна кількість згенерованих правил - 285

Підтримка найсильнішого правила = 0.81

Секція 1. Аналітичні методи та економіко-математичні моделі в задачах управління економічними системами

Середня підтримка перших 5 правил - 0.64

- "('crmvct 2', 'blgetmg 2')",0.8103975535168195

Респодент, або близьке оточення, НЕ були жертвами булінгу на робочому місці впродовж останніх 5 років.

Респодент НЕ належить до нацменшин

- "('rlgblg 1', 'blgetmg 2')",0.6261467889908257

Респодент НЕ належить до нацменшин

Респодент вважає себе приналежним до певної релігії

- "('crmvct 2', 'rlgblg 1')",0.6215596330275229

Респодент, або близьке оточення, НЕ були жертвами булінгу на робочому місці останніх 5 років.

Респодент вважає себе приналежним до певної релігії

- "('blgetmg 2', 'gndr 2')",0.5879204892966361

Респодент НЕ належить до нацменшин

Респодент жінка

- "('crmvct 2', 'gndr 2')",0.581039755351682

Респодент жінка

Респодент, або близьке оточення, НЕ були жертвами булінгу на робочому місці останніх 5 років.

- "('blgetmg 2', 'partner 1')",0.5298165137614679

Респодент НЕ належить до нацменшин

Респодент живе має сім'ю.

2. Правила довжиною 3.

Загальна кількість згенерованих правил - 695

Підтримка найсильнішого правила = 0.56

Середня підтримка перших 5 правил - 0.48

- "('crmvct 2', 'rlgblg 1', 'blgetmg 2')",0.5611620795107034

- "('crmvct 2', 'blgetmg 2', 'gndr 2')",0.5259938837920489

- "('crmvct 2', 'blgetmg 2', 'partner 1')",0.4694189602446483

- "('rlgblg 1', 'blgetmg 2', 'gndr 2')",0.4334862385321101

3. Правила довжиною 4.

Загальна кількість згенерованих правил - 741

Підтримка найсильнішого правила = 0.39

Середня підтримка перших 5 правил - 0.31

Далі пояснено лише ті значення, яких не було в минулих правлах.

- "('crmvct 2', 'rlgblg 1', 'blgetmg 2', 'gndr 2')",0.39067278287461776

- "('crmvct 2', 'rlgblg 1', 'blgetmg 2', 'partner 1')",0.3333333333333333

- "('crmvct 2', 'blgetmg 2', 'gndr 2', 'partner 1')",0.29357798165137616

- "('sclact 3', 'crmvct 2', 'rlgblg 1', 'blgetmg 2')",0.2874617737003058

Бере участь в соціальних активностях приблизно стільки ж, скільки

однолітки.

- 155,"('crmvct 2', 'rlgblg 1', 'blgetmg 2', 'hincfel 3')",0.2622324159021407
Складно жити з теперішнім доходом.

4. Правила довжиною 5.

Загальна кількість згенерованих правил — 331

Підтримка найсильнішого правила = 0.22

Середня підтримка перших 5 правил — 0.19

- "('crmvct 2', 'rlgblg 1', 'blgetmg 2', 'gndr 2', 'partner 1')",
0.22247706422018348
- "('sclact 3', 'crmvct 2', 'rlgblg 1', 'blgetmg 2', 'gndr 2')",
0.19954128440366972
- "('crmvct 2', 'rlgblg 1', 'blgetmg 2', 'gndr 2', 'hincfel 3')",
0.19495412844036697
- "('crmvct 2', 'health 3', 'rlgblg 1', 'blgetmg 2', 'gndr 2')", 0.1720183486238532
- "('crmvct 2', 'rlgblg 1', 'blgetmg 2', 'gndr 2', 'partner 2')",
0.1659021406727829

5. Правила довжиною 6.

Загальна кількість згенерованих правил - 62

Підтримка найсильнішого правила = 0.12

Середня підтримка перших 5 правил - 0.10

- "('crmvct 2', 'rlgblg 1', 'blgetmg 2', 'gndr 2', 'partner 1', 'hincfel 3')",
0.12155963302752294
- "('sclact 3', 'crmvct 2', 'rlgblg 1', 'blgetmg 2', 'gndr 2', 'partner 1')",
0.11314984709480122
- "('sclact 3', 'crmvct 2', 'rlgblg 1', 'blgetmg 2', 'gndr 2', 'hincfel 3')",
0.10397553516819572
- "('crmvct 2', 'health 3', 'rlgblg 1', 'blgetmg 2', 'gndr 2', 'partner 1')",
0.10091743119266056
- "('sclact 3', 'crmvct 2', 'health 3', 'rlgblg 1', 'blgetmg 2', 'gndr 2')",
0.08868501529051988

6. Правила довжиною 7

Для правил довжиною 7 вдалось згенерувати лише 3 правила. Це пов'язано з обмеженням часу виконання одного cell в Jupyter.

Загальна кількість згенерованих правил - 3

Підтримка найсильнішого правила = 0.06

Середня підтримка перших 3 правил - 0.05

- "('sclact 3', 'crmvct 2', 'rlgblg 1', 'blgetmg 2', 'gndr 2', 'partner 1', 'hincfel 3')",
0.06498470948012232
- "('crmvct 2', 'health 3', 'rlgblg 1', 'blgetmg 2', 'gndr 2', 'partner 1', 'hincfel 3')",

Секція 1. Аналітичні методи та економіко-математичні моделі в задачах управління економічними системами

0.0581039755351682

- "('sclact 3', 'crmvct 2', 'health 3', 'rlgblg 1', 'blgetmg 2', 'gndr 2', 'partner 1')",
0.054281345565749234

Представимо отримані результати у вигляді Рис. 1. Порівняємо рівень підтримки для найсильніших правил в залежності від їх довжини на графіку. Очікувано, що чим довше правило, тим менший максимальний рівень підтримки. Але разом з тим зменшується різниця між найсильнішим правилом і середнім значенням підтримки перших 5 правил. Це можна пояснити тим, що правила ніби зливаються і стають більше залежними одне від одного.

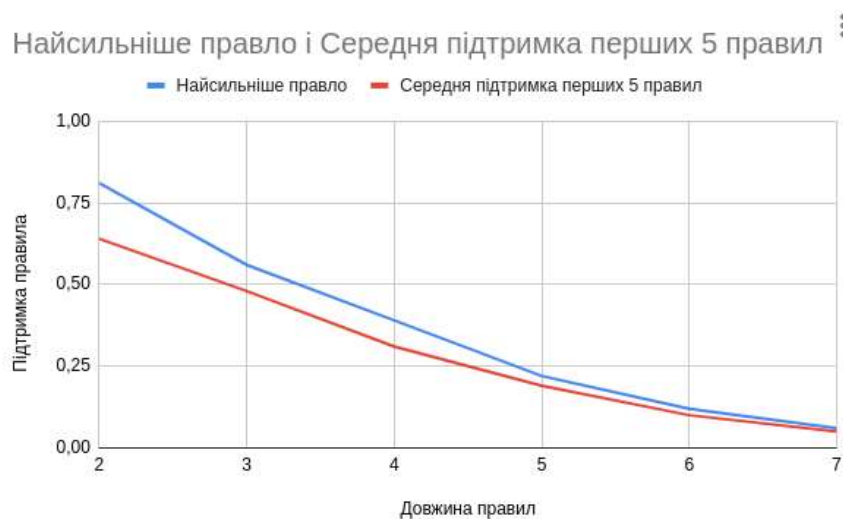


Рис. 1. Рівень підтримки правил різної довжини

Як вказувалось вище, для подібних до соціологічних досліджень наборів даних надзвичайно важливо правильно підготувати дані. Залишити лише ті дані, які справді важливі для аналізу. В ході експерименту залишились лише ті респонденти, суб'єктивне відчуття задоволеності роботою яких вище середнього. Разом з розміром набору даних зменшилась довжина часу виконання.

Розглянемо властивість адитивності в правилах. Вона означає, що якщо $A \rightarrow B$ та $A \rightarrow C$, то $A \rightarrow (B, C)$ [6].

В ході експерименту було знайдено сильні правила довжиною 2:

- "('crmvct 2', 'blgetmg 2')" - 0.8103975535168195
- "('rlgblg 1', 'blgetmg 2')" – 0.6261467889908257
- "('crmvct 2', 'rlgblg 1')" – 0.6215596330275229

Це 3 найсильніші правила довжиною 2. Вони складаються з 3 унікальних значень: 'crmvct 2', 'rlgblg 1', 'blgetmg 2'.

А тепер якщо переглянути найсильніше правило довжиною 3, воно складатиметься з усіх трьох найсильніших правил довжиною 2:

"('crmvct 2', 'rlgblg 1', 'blgetmg 2')" – 0.5611620795107034.

Аналогічно для перших 4 найсильніших правил довжиною 2 є 4 унікальні значення: 'crmvct 2', 'rlgblg 1', 'blgetmg 2', 'gndr 2'.

Найсильніше правило довжиною 4 складається з усіх цих значень:

"('crmvct 2', 'rlgblg 1', 'blgetmg 2', 'gndr 2')" – 0.39067278287461776

Те ж саме можна спостерігати і з іншими значеннями. Така "адитивність" не дає можливість дізнатись підтримку правил, але у випадку, коли потрібно знайти лише найсильніші – дуже пришвидшує обрахунки, адже не потрібно повністю перевіряти весь набір даних, а лише вже отримані правила.

Висновок. Здійснення аналізу асоціативних правил вимагає використання різноманітних методів, таких як аналіз тексту, статистичні методи та алгоритми машинного навчання. Однак, результати цього дослідження можуть принести значний внесок у покращення управління персоналом, розвитку організаційних процесів та створенні сприятливого робочого середовища.

Розуміння асоціативних правил може бути корисним для розробки персоналізованих програм навчання та розвитку персоналу, що дозволить підвищити ефективність роботи працівників та задоволеність їхніх потреб. Таким чином, дослідження асоціативних правил має великий потенціал для вдосконалення організаційних практик та сприяння підвищенню результативності та конкурентоспроможності компаній.

Список використаних джерел

1. Davison H. K., Maraist C., Bing M. N. Friend or foe? The promise and pitfalls of using coworker ratings to assess performance. *Journal of Applied Psychology*. 2011. 96(6). p. 1258–1274.
2. Galesic M., Bosnjak M. Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly*. 2009. 73(2). p. 349–360.
3. Kozlowski S. W., Bell B. S. Work groups and teams in organizations. *Handbook of psychology: Industrial and organizational psychology*. 2003. 3. p. 333-375.
4. Lee S. M., Lee D. Sentiment analysis of employee feedback to improve organizational management. *Journal of Organizational Computing and Electronic Commerce*. 2008. 28(3). p. 197-215.
5. Miao C. F., Qian S., Ma D. S., Xin T. Exploring Employees' Performance Evaluation Based on Data Analysis of Multi-source Heterogeneous Information. *IEEE Access*. 2020. 8. p. 228660-228674.
6. Reynolds W. M., Reynolds W. M. Development of reliable and valid short forms of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology*. 2001. 57(12). p. 1579-1596.