

УДК 004.75

Болдак А. О., Лисенко Д. В., Ханін М. Є.

СПОСІБ ОПРАЦЮВАННЯ ПРОПУСКІВ В РЕЗУЛЬТАТАХ ЕКСПЕРТНИХ ОПИТУВАНЬ

METHOD OF MISSED ANSWERS PROCESSING IN RESULTS OF EXPERT EVALUATION

У статті запропонований новий спосіб опрацювання пропусків в результатах експертних опитувань, який мінімізує втрати об'єму вибірки при збереженні точності експертних оцінок. Програмна реалізація запропонованого способу використовується в системі експертних опитувань, розробленій на розподіленій платформі опрацювання даних.

Ключові слова: експертне опитування, заповнення пропусків, об'єм вибірки, програмне забезпечення, розподілене опрацювання даних.

Актуальність дослідження. Експертне оцінювання [1] як процедура оцінювання деякого об'єкта на основі сукупності суб'єктивних оцінок експертів сьогодні широко використовується для вирішення завдань підтримки прийняття рішень в області маркетингу, економічного аналізу [2] та таких нових областях, якими є, наприклад, прогнозування та передбачення розвитку соціо-економічних систем [3].

Постановка проблеми. Не зважаючи на те, що існує достатній спектр способів організації процедури експертного опитування (метод асоціацій, метод парних порівнянь, індивідуальне опитування, метод «Дельфі» та ін.) та відповідних методів об'єктивізації експертних оцінок, об'єм вибірки, тобто потужність множини відповідей експертів на запитання, є одним з найважливіших факторів, що впливають на точність оцінки. Тому на підготовчій стадії експертного опитування використовують один з підходів до вибору кількості експертів, наприклад таких, як [4]:

- довільний – об'єм вибірки обирається довільно без розрахунків точності такого оцінювання;
- довірчий інтервал – вказується необхідна точність оцінювання, після чого розраховується об'єм вибірки, за якої досягається задана точність;
- статистичний аналіз – вказується необхідна точність для окремих груп, які формуються на основі віку, статі, рівню освіти та ін., після чого розраховується необхідний загальний об'єм вибірки.

Але реальний об'єм вибірки зменшується в залежності від кількості пропусків у відповідях експертів та вибраного способу опрацювання цих пропусків, наприклад таких як [5]: методи зважування та заміщення, логічний метод, заповнення середніми або випадковими значеннями, заміна експертів. Використання вказаних методів опрацювання пропусків в реальних умовах викликає певні труднощі, пов'язані з необхідністю залучення занадто великої кількості експертів, організації повторних опитувань та недостатньої точності результатів експертизи.

Постановка завдання. Метою роботи є зменшення втрат об'єму вибірки при збереженні точності оцінок експертів за рахунок розробки та використання нового способу опрацювання пропусків в результатах експертних опитувань.

Викладення основного матеріалу. Пропонований спосіб заснований на гіпотезі, що якщо респонденти в рамках певної тематики дають близькі відповіді на

питання однієї множини, то їхні відповіді на питання тієї ж тематики з іншої множини теж будуть близькими. Для оцінки «близькості» між респондентами введемо формальну модель: респондент – точка багатовимірного простору відповідей.

Для прикладу візьмемо наступну матрицю відповідей:

```

1, n, n, 1, n, n, 1, 0, 1, 1, n, n, 0, 1, n, n, 0, n, n, n, 0, n, n, n, 0, 1, 0
0, 1, n, 0, n, 0, n, 1, 0, 1, n, 0, n, 1, 1, n, n, 1, n, 1, 1, n, 1, 0, 0, 0, 0
0, n, 0, n, 0, 1, 1, 1, 1, 1, 0, 0, 0, 1, n, 0, 0, 0, n, 0, 0, 0, 0, n, 0, 1, 0
1, 0, 0, n, 0, 1, 0, 0, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0
0, 0, 1, 0, 0, 1, 0, 0, 0, n, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0, 1
n, 0, 1, 0, n, 1, 0, 0, n, n, 1, 0, 0, 0, 0, 0, 0, n, 0, 0, 1, 1, 0, 1, 1, 1, n,

```

де кожний рядок – це відповіді одного експерта, а кожний стовпець – одне питання.

1 та 0 – представлення відповідей у просторі, n – означає відсутність відповіді.

Для розбиття на групи зі схожими відповідями використаємо проекцію на питання, на які всі експерти дали відповідь:

```

0, 1, 0, 0, 1
1, 1, 1, 0, 0
1, 1, 0, 0, 1
0, 1, 0, 0, 1
0, 0, 1, 1, 0
0, 0, 1, 1, 1

```

Тоді можна ввести деяку функцію відстані $l(x_i, x_j)$ між точками x_i та x_j . В разі ортогоналізації цього простору за допомогою метода головних компонент [6] в якості $l(x_i, x_j)$ можна використати евклідову відстань:

$$l(x_i, x_j) = \sqrt{\sum_{k=1}^n (p_{ik} - p_{jk})^2} \quad (1)$$

де x_i, x_j – точки простору, n – його розмірність, p_{ik} – k -та координата відповідної точки.

Тепер необхідно розбити експертів на групи зі схожими відповідями. Для цього використаємо метод k -середніх [7] (англ. k -means) та функцію відстані (1).

Для визначення кількості груп використовується *elbowmethod*[8]. Ідея методу полягає в тому, що ми обраховуємо k -means [7] для набору даних з різними k в якомусь діапазоні, та для кожного k обчислюємо *sumofsquarederrors* (SSE). Тоді обирається мінімальне k таке, при якому SSE насадовольняє.

Отримані кластери (останній стовпець – номер кластеру):

```

1, n, n, 1, n, n, 1, 0, 1, 1, n, n, 0, 1, n, n, 0, n, n, n, 0, n, n, n, 0, 1, 0, 0
0, 1, n, 0, n, 0, n, 1, 0, 1, n, 0, n, 1, 1, n, n, 1, n, 1, 1, n, 1, 0, 0, 0, 0, 1
0, n, 0, n, 0, 1, 1, 1, 1, 1, 0, 0, 0, 1, n, 0, 0, 0, n, 0, 0, 0, 0, n, 0, 1, 0, 0
1, 0, 0, n, 0, 1, 0, 0, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0
0, 0, 1, 0, 0, 1, 0, 0, 0, n, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0, 1, 2
n, 0, 1, 0, n, 1, 0, 0, n, n, 1, 0, 0, 0, 0, 0, 0, n, 0, 0, 1, 1, 0, 1, 1, 1, n, 2

```

Розглянемо кластер **0**:

1, n, n, 1, n, n, 1, 0, 1, 1, n, n, 0, 1, n, n, 0, n, n, n, 0, n, n, n, 0, 1, 0, **0**
 0, n, 0, n, 0, 1, 1, 1, 1, 1, 0, 0, 0, 1, n, 0, 0, 0, n, 0, 0, 0, 0, n, 0, 1, 0, **0**
 1, 0, 0, n, 0, 1, 0, 0, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, **0**

Для кожного питання, у якому є пропущенні відповіді, розбиваємо експертів на 2 групи: ті, які відповіли, і ті, які пропустили дане питання.

Візьмемо друге питання. На нього відповів тільки третій респондент цього кластера, у матриці ця відповідь представлена нулем, який і є середнім значенням.

Розглянемо третє питання. На нього відповіли 2 та 3 респонденти, тоді середнє значення $\frac{0+0}{2} = 0$. При необхідності це значення округлюється.

Отримане середнє значення присвоюється іншим експертам цього кластера для цього питання.

Може бути випадок, коли жоден респондент даного кластеру не відповів на питання. Тоді ці поля залишаються незаповненими. Такий випадок є, наприклад, у другому кластері:

0, 0, 1, 0, 0, 1, 0, 0, 0, **n**, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0, 1, **2**
 n, 0, 1, 0, n, 1, 0, 0, n, **n**, 1, 0, 0, 0, 0, 0, 0, n, 0, 0, 1, 1, 0, 1, 1, 1, n, **2**

Зробивши аналогічні дії з кожним питанням кожного кластеру, отримуємо наступну матрицю:

1, 0, 0, 1, 0, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0
 0, 1, n, 0, n, 0, n, 1, 0, 1, n, 0, n, 1, 1, n, n, 1, n, 1, 1, n, 1, 0, 0, 0, 0
 0, 0, 0, 1, 0, 1, 1, 1, 1, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0
 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0
 0, 0, 1, 0, 0, 1, 0, 0, 0, n, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0, 1
 0, 0, 1, 0, 0, 1, 0, 0, 0, n, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 1, 1, 1, 1

Якщо відкидати експертів, які відповіли не на всі питання, то об'єм вибірки буде становити $\frac{3}{6} * 100\% = 50\%$ від початкового об'єму. Якщо відкидати питання, на які відповіли не всі експерти, то об'єм вибірки буде становити $\frac{17}{27} * 100\% \approx 63\%$.

Тепер розглянемо початкову матрицю:

1, n, n, 1, n, n, 1, 0, 1, 1, n, n, 0, 1, n, n, 0, n, n, n, 0, n, n, n, 0, 1, 0
 0, 1, n, 0, n, 0, n, 1, 0, 1, n, 0, n, 1, 1, n, n, 1, n, 1, 1, n, 1, 0, 0, 0, 0
 0, n, 0, n, 0, 1, 1, 1, 1, 1, 0, 0, 0, 1, n, 0, 0, 0, n, 0, 0, 0, 0, n, 0, 1, 0
 1, 0, 0, n, 0, 1, 0, 0, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0
 0, 0, 1, 0, 0, 1, 0, 0, 0, n, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0, 1
 n, 0, 1, 0, n, 1, 0, 0, n, n, 1, 0, 0, 0, 0, 0, 0, n, 0, 0, 1, 1, 0, 1, 1, 1, n

При відкиданні експертів у даному прикладі об'єм вибірки буде становити 0%, а при відкиданні питань $\frac{5}{27} * 100\% \approx 18,5\%$.

Висновки. Таким чином, запропонований спосіб опрацювання пропусків у відповідях експертів дозволяє зменшити втрати об'єму вибірки та забезпечити точність більшу у порівнянні з заповненням пропусків середніми значеннями. Обмеженням запропонованого методу є недостатня кількість запитань, на які відповіли всі респонденти (кількість таких питань повинна перевищувати кількість респондентів).

Програмна реалізація запропонованого способу використовується в системі експертних опитувань на розподіленій платформі опрацювання даних DJ [9].

В подальшому планується розширення запропонованого способу на експертні оцінювання у номінальних та ординальних шкалах.

Список використаних джерел

1. А.И.Орлов (2002) Экспертные оценки. Учебное пособие
2. Damianos P. Sakas, Nikolaos Konstantopoulos (2010) Marketing and Management Sciences: Proceedings of the International Conference on ICMMS 2008
3. M. Zgurovsky (2016) Foresight and construction of the strategies of socio-economic development of Ukraine
4. Н.Г.Каменева, В.А.Поляков (2011) Маркетинговые исследования ст. 78-87.
5. Л.Р. Мухаметова (2009) Методы выборочных обследований ст. 55-57.
6. Jonathon Shlens (2014) A Tutorial on Principal Component Analysis
7. Steinhaus H. (1956). Sur la division des corps materiels en parties. Bull. Acad. Polon. Sci., C1. III vol IV: 801-804.
8. Using the elbow method to determine the optimal number of clusters for k-means clustering [Електронний ресурс]. - Режим доступу: bl.ocks.org (дата звернення 24.04.2018) – Назва з екрана.
9. ЕФЕКТИВНІСТЬ СУСПІЛЬНИХ ПЕРЕТВОРЕНЬ. II ТУР [Електронний ресурс]. - Режим доступу: <https://dj-next.herokuapp.com/app/Social%20Transformations%20II/> (дата звернення 24.04.2018) – Назва з екрана.

ДОВІДКА ПРО АВТОРІВ

Болдак Андрій Олександрович – доцент, кафедра обчислювальної техніки, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського».

Boldak Andriy – associate professor, Department of Computer Engineering, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”.

E-mail: boldak.andrey@gmail.com

Лисенко Дмитро Вадимович – студент, кафедра обчислювальної техніки, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського».

Lysenko Dmytro – student, Department of Computer Engineering, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”.

E-mail: dimalysenko1206@gmail.com

Ханін Максим Євгенійович – студент, кафедра обчислювальної техніки, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського».

Khanin Maxim – student, Department of Computer Engineering, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”.

E-mail: Maksimkhanin.kpi@gmail.com

A.O. Boldak, D.V. Lysenko, M.E. Khanin

METHOD OF MISSED ANSWERS PROCESSING IN RESULTS OF EXPERT EVALUATION

Relevance of the research topic. Expert evaluation today is widely used to solve decision-making tasks in the field of marketing, economic analysis and such new areas as prediction of the development of socio-economic systems.

Target setting. The actual sample size is reduced, depending on the number of passes in the expert's responses and the chosen method for processing these passes, such as: weighing and substitution methods, logical method, filling in average or random values, replacing experts. The purpose of the work is to reduce the sample size loss while maintaining the accuracy of expert assessments by developing and using a new way of processing passes in the results of expert evaluation.

The statement of basic materials. This method suggests filling passed answers in expert evaluation by the average of groups. For a breakdown into groups, a formal model is introduced: the respondent is the point of a multidimensional response space. In this space, the k-means algorithm is used. To determine the value of k, use the elbow method.

Conclusions. The proposed method for processing passes in the expert's answers allows to reduce the loss of sample size and to provide greater accuracy compared with the filling of the passes with average values.

Key words: expert evaluation, filling of passes, sample volume, software, distributed data processing.