



соответствуют расчетным. Проведенные исследования показали достоверность результатов расчетов методом конечных элементов.

**Вывод.** Проведенные исследования показали, что ширина целиков  $b$  оказывает существенное влияние на напряжения в кровле камер только в том случае, если целик значительно меньше пролета камеры  $a$ . Это влияние сказывается только в том случае, когда  $b/a < 0,6$ . В практике рудников отношение  $b/a < 0,4$  не встречается. При таком отношении напряжения в кровле серии камер на 12-15% больше чем у одиночной камеры. Эта величина лежит в пределах точности подобных расчетов. Следовательно, в расчетах на прочность пролетов серии камер можно использовать данные о напряжениях, возникающих в кровле одиночной камеры, находящейся в аналогичных условиях, либо внести соответствующие поправки в эти расчеты в соответствии с данными исследований.

## ЛИТЕРАТУРА

1. Борисенко С.Г. Технология подземной разработки рудных месторождений: учеб. для вузов / С.Г. Борисенко. – К.: Вища школа, 1987. – 262 с.
2. Зенкевич О.К. Метод конечных элементов в теории сооружений и в механике сплошных сред / О.К. Зенкевич, И.К. Ченг. – М.: Недра, 1974. – 230 с.
3. Хмарский В.В. Выбор рациональной высоты камер при разработке мощных крутых рудных месторождений / Хмарский В.В. // Автореферат кандидатской диссертации. – Днепропетровск, 1981. – 20 с.
4. Борисенко С.Г. Расчет междукамерных целиков на устойчивость / С.Г. Борисенко, В.В. Хмарский // Горный журнал. – М., 1978. – № 10. – С. 21 – 25.

УДК 681.518.25

## ДОСЛІДЖЕННЯ ВІДСТАНІ КУЛЬБАКА-ЛЕЙБЛЕРА У ЗАДАЧАХ МОДЕЛЮВАННЯ У КРЕДИТНОМУ СКОРИНГУ

О.М. Солошенко<sup>1</sup>

<sup>1</sup>аспірант, Національний технічний університет України «Київський політехнічний інститут», м. Київ, Україна, e-mail: [soloshenko\\_s@ukr.net](mailto:soloshenko_s@ukr.net)

**Анотація.** В роботі проведено теоретичне дослідження пошуку взаємозв'язку класичної відстані Кульбака-Лейблера та загальноприйнятих статистичних показників, що відображають щонайменше два напрямки практичного застосування у задачах бінарної класифікації, зокрема у задачах кредитного скорингу. Наведено особливості включення відстані Кульбака-Лейблера у розрахунок ключових індикаторів кредитного скорингу.

*Ключові слова:* відстань Кульбака-Лейблера, похідна Радона-Нікодима, ентропія, кредитний скоринг, скорингова карта, значення інформації, індекс стабільності популяції.





## KULLBACK–LEIBLER DIVERGENCE RESEARCH FOR THE SIMULATION OF THE CREDIT SCORING

Oleksandr Soloshenko<sup>1</sup>

<sup>1</sup>postgraduate student, National Technical University of Ukraine “Kyiv Polytechnic Institute”, Kyiv, Ukraine, e-mail: [soloshenko\\_s@ukr.net](mailto:soloshenko_s@ukr.net)

**Abstract.** The theoretical research has been conducted to find the relationship between classic Kullback–Leibler divergence and generally accepted statistical indicators which reflect at least two practical application fields in the binary classification tasks, particularly in the credit scoring tasks. The peculiarities of the including the Kullback–Leibler divergence into the key statistical indicators estimation process are given.

*Keywords:* Kullback–Leibler divergence, Radon–Nikodym derivative, entropy, credit scoring, scorecard, information value, population stability index.

**Вступ.** У задачах моделювання, зокрема у задачах бінарної класифікації, до яких відноситься побудова моделей кредитного скорингу, поведінкового скорингу та інших скорингових моделей (скорингових карт), чільне місце посідають методи оцінювання предикативної сили вхідних параметрів відносно цільової бінарної змінної з метою вирішення доцільності включення проаналізованих параметрів у модель, та методи оцінювання стабільності параметрів моделі, а також результуючих скорингових груп, як у процесі побудови моделей з метою виключення нестабільних в часі факторів, так і в процесі використання моделей з метою постановки питання перебудови існуючих моделей задля забезпечення стабільності популяції відносно навчальної вибірки як на рівні результуючого скорингового балу, так і на рівні кожної окремо взятої характеристики, що включена в модель [1].

Актуальною є задача математичного та статистичного обґрунтування ключових загальноприйнятих індикаторів [2] з використанням класичних статистичних понять та показників – тут за допомогою відстані Кульбака–Лейблера [3].

**Мета роботи.** Виявити та довести взаємозв'язок класичної відстані Кульбака–Лейблера з подібними до індексу ентропії [4] показниками значення інформації (Information Value, IV) та індексу стабільності популяції (Population Stability Index, PSI), а також з вагою значення змінної (Weight Of Evidence, WOE) як складової зокрема значення інформації [1, 2].

Об'єктом дослідження є класичний показник значення інформації та класичний індекс стабільності популяції.



Предметом дослідження є метод підрахунку нерівності дискретних розподілів за допомогою відстані Кульбака-Лейблера.

Суть імплементації показників скорингу як обчислення функції від відстаней Кульбака-Лейблера. Формули для обчислення показників IV та WOE для певної дискретної або дискретизованої вхідної змінної, щоб оцінити предикативну силу характеристики, на множині прийнятих заявок на надання кредиту з чітко визначеним майбутнім бінарним результатом (позитивним або негативним) мають такий вигляд [1, 2]:

$$WOE_i = \ln\left(\frac{g_i}{b_i}\right) \quad (1)$$

$$IV = \sum_{i=1}^c (g_i - b_i) \ln\left(\frac{g_i}{b_i}\right) = \sum_{i=1}^c (g_i - b_i) WOE_i \quad (2)$$

Тут  $g_i$  – відносна кількість позитивних заявок у сегменті (кластері) до загальної кількості позитивних заявок:

$$g_i = \frac{G_i}{\sum_{i=1}^c G_i} = \frac{G_i}{G} \quad (3)$$

Тобто оперуємо розподілом позитивних заявок по дискретних або дискретизованих значеннях змінної, тому має місце рівність:

$$\sum_{i=1}^c g_i = 1 \quad (4)$$

Складова  $b_i$  – це відносна кількість негативних заявок у сегменті (кластері) до загальної кількості негативних заявок:

$$b_i = \frac{B_i}{\sum_{i=1}^c B_i} = \frac{B_i}{B} \quad (5)$$

Тобто оперуємо розподілом негативних заявок по дискретних або дискретизованих значеннях змінної, тому має місце рівність:

$$\sum_{i=1}^c b_i = 1 \quad (6)$$



Щодо індексу стабільності популяції (PSI) на рівні розподілу скорингових груп та на рівні розподілу характеристик, має місце подібна формула його оцінювання [1, 2]:

$$PSI = \sum_{i=1}^c (r_i - d_i) \ln \left( \frac{r_i}{d_i} \right) \quad (7)$$

Тут  $r_i$  відображає дискретний розподіл по поточній вибірці (recent sample) відносно скорингових груп або відносно дискретних значень вхідної змінної, що аналізується на стабільність, а  $d_i$  відображає аналогічний дискретний розподіл, але по навчальній вибірці (development sample). Виконуються наступні тотожності:

$$\sum_{i=1}^c r_i = 1 \quad (8)$$

$$\sum_{i=1}^c d_i = 1 \quad (9)$$

Сутність формул IV та PSI однакова – оцінити ступінь розходження двох дискретних розподілів [2].

Відстань Кульбака-Лейблера є спрямованою дивергенцією (directed divergence) для двох розподілів [5]:

$$D_{KL}(P \parallel Q) = \sum_i \ln \left( \frac{P(i)}{Q(i)} \right) P(i) \quad (10)$$

Для щільностей двох неперервних розподілів відстань Кульбака-Лейблера обчислюється таким чином [6]:

$$D_{KL}(P \parallel Q) = \int_{-\infty}^{+\infty} \ln \left( \frac{p(x)}{q(x)} \right) p(x) dx \quad (11)$$

У більш загальному випадку при виконанні необхідних умов, відстань Кульбака-Лейблера можна переписати з використанням похідної Радона-Нікодіма (Radon-Nikodym derivative) [7], використовуючи кумулятивні функції двох розподілів:



$$D_{KL}(P \parallel Q) = \int_X \ln\left(\frac{dP}{dQ}\right) dP = \int_X \ln\left(\frac{dP}{dQ}\right) \frac{dP}{dQ} dQ \quad (12)$$

Суть узагальненого виявленого взаємозв'язку та імплементації показників скорингу як обчислення функції від відстаней Кульбака-Лейблера описується наступними формулами:

$$IV = \sum_i \ln\left(\frac{g_i}{b_i}\right) g_i - \sum_i \ln\left(\frac{g_i}{b_i}\right) b_i = D_{KL}(G \parallel B) + D_{KL}(B \parallel G) \quad (13)$$

$$WOE_i = \ln\left(\frac{g_i}{b_i}\right) \quad (14)$$

$$PSI = \sum_i \ln\left(\frac{r_i}{d_i}\right) r_i - \sum_i \ln\left(\frac{r_i}{d_i}\right) d_i = D_{KL}(R \parallel D) + D_{KL}(D \parallel R) \quad (15)$$

**Висновки.** Виявлено та доведено взаємозв'язок класичної відстані Кульбака-Лейблера та загальноприйнятих статистичних показників IV, WOE, PSI, що відображають щонайменше два напрямки практичного застосування у задачах класифікації, зокрема у задачах кредитного скорингу: оцінювання прогнозої (предикативної) сили вхідних змінних та оцінювання стабільності характеристик та скорингових груп. Наведено особливості включення відстані Кульбака-Лейблера у розрахунок ключових індикаторів кредитного скорингу – значення інформації (IV) та індексу стабільності популяції (PSI) через суму двох відстаней Кульбака-Лейблера, що відображають два можливі варіанти спрямованої дивергенції (directed divergence), що підтверджує теоретичну і практичну цінність дослідження.

Перспективи подальших досліджень включають дослідження статистичної суті інших загальноприйнятих показників кредитного скорингу та пропозицію альтернативних метрик для вирішення ключових задач.

## ЛІТЕРАТУРА

1. Credit Scoring and its Applications [Текст]: SIAM monographs on mathematical modeling and computation. / Lyn C. Thomas, David B. Edelman, Jonathan N. Crook. — University City Science Center, Philadelphia, SIAM, 2002. — 248 p.
2. Credit risk scorecards: developing and implementing intelligent credit scoring [Текст]: навч. посібник / Naeem Siddiqi. — Hoboken: John Wiley & Sons, Inc., 2006. — 196 p.



3. On Information and Sufficiency [Текст] / S. Kullback, R.A. Leibler // Annals of Mathematical Statistics – Bethesda: Institute of Mathematical Statistics, 1951. – № 1. – 79 – 86 p.p.
4. Теорія інформації та кодування [Текст]: навч. посібник / В.Л. Кожевников, А.В. Кожевников. — Д.: Національний гірничий університет, 2011. — 108 с.
5. Information theory and statistics [Текст]: навч. посібник / S. Kullback. — NY: John Wiley and Sons, 1959. — 416 p.
6. Pattern Recognition and Machine Learning [Текст]: навч. посібник / С. Bishop. — New York: Springer, 2006. — 738 p.
7. Элементы теории функций и функционального анализа [Текст]: навч. посібник / А.Н. Колмогоров, С.В. Фомин. — Москва: Наука, 1976. — 544 с.

УДК 519.81

## **ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ СИСТЕМЫ ОЦЕНИВАНИЯ УРОВНЯ КОМПЕТЕНТНОСТИ ЭКСПЕРТОВ**

**В.И.Корсун<sup>1</sup>, Э.Ю. Прокуда<sup>2</sup>**

<sup>1</sup>доктор технических наук, профессор, заведующий кафедры метрологии и информационно-вычислительных технологий, Государственное высшее учебное заведение «Национальный горный университет», г. Днепропетровск, Украина

<sup>2</sup>аспирант кафедры метрологии и информационно-вычислительных технологий, Государственное высшее учебное заведение «Национальный горный университет», г. Днепропетровск, Украина, e-mail: [elinka9891@mail.ru](mailto:elinka9891@mail.ru)

**Аннотация.** В статье выполнен анализ формирования экспертной группы для оценки состояния базовых элементов карьерных автосамосвалов. Рассмотрены многоэтапный процесс формирования экспертной группы. Представленная программа оценки уровня компетентности экспертов.

*Ключевые слова:* эксперт, уровень компетентности, экспертная группа, автосамосвал.

## **INFORMATION SUPPORT FOR THE EVALUATION SYSTEM OF EXPERTS' COMPETENCE LEVEL**

**Valeriy Korsun<sup>1</sup>, Elina Prokuda<sup>2</sup>**

<sup>1</sup>Doctor of Engineering Science, Professor, Head of the Department of Metrology and Information Computation Technologies, State Higher Educational Institution "National Mining University", Dnepropetrovsk, Ukraine

<sup>2</sup>Postgraduate student of Department of Metrology and Information Computation Technologies, State Higher Educational Institution "National Mining University", Dnepropetrovsk, Ukraine, e-mail: [elinka9891@mail.ru](mailto:elinka9891@mail.ru)