

Анализ устойчивости алгоритмов отбора значимых факторов**Научный руководитель – Булинский Александр Вадимович****Захаров Даниил Вадимович***Студент (специалист)*

Московский государственный университет имени М.В.Ломоносова,
 Механико-математический факультет, Кафедра теории вероятностей, Москва, Россия
E-mail: Zakharov.daniil@gmail.com

Для обработки больших объёмов данных применяются разнообразные алгоритмы отбора значимых факторов (признаков). Они позволяют выявить набор признаков, оказывающих существенное влияние на исследуемую величину. Важным показателем качества таких алгоритмов является устойчивость. Она показывает, насколько меняется результат работы алгоритма при небольшом изменении искомым данных. Для измерения устойчивости алгоритма используются различные функции Φ (меры устойчивости). Нашей целью является исследование свойств мер устойчивости.

В настоящее время для анализа устойчивости применяются меры схожести, основанные на попарном сравнении результатов работы алгоритма, и меры, связанные с частотными характеристиками признаков [3]. Пусть в стохастической модели изучаемая величина зависит от признаков X_1, \dots, X_n . В результате работы алгоритма на k -ом наборе данных выбирается подмножество $s_k \subset (1, \dots, n)$, описывающее индексы значимых факторов. Нами рассматривается новый вариант меры схожести, основанный на сравнении троек результатов работы алгоритма:

$$\Phi_3(Z) = \frac{1}{M(M-1)(M-2)} \sum_{i=1}^M \sum_{\substack{j=1 \\ j \neq i}}^M \sum_{\substack{l=1 \\ l \neq i \\ l \neq j}}^M \varphi_3(s_i, s_j, s_l),$$

где $Z = \{s_k\}_{k=1}^M$ – набор результатов работы алгоритма, $\varphi_3(s_i, s_j, s_k)$ – мера схожести каждой конкретной тройки из набора Z .

Меры устойчивости, вовлекающие попарные сравнения, обычно удовлетворяют некоторым важным свойствам. К ним относятся полная определённость, симметричность, ограниченность, монотонность и поправка на случайность [1]. В нашей работе получены аналоги упомянутых свойств для меры Φ_3 , а также построены примеры φ_3 , опирающиеся на известные меры [2, 4].

Источники и литература

- 1) Kuncheva, L.I. A stability index for feature selection. Proceedings of the IASTED International Conference on Artificial Intelligence and Applications, AIA 2007.
- 2) Naik, A.K., Kuppili, V., Edla, D.R. A new hybrid stability measure for feature selection. Appl Intell 50, 3471–3486 (2020).
- 3) Nogueira, Sarah et al. On the Stability of Feature Selection Algorithms. J. Mach. Learn. Res. 18 (2017): 174:1-174:54.
- 4) Sen, R., Mandal, A.K., Chakraborty, B. A Critical Study on Stability Measures of Feature Selection with a Novel Extension of Lustgarten Index. Mach. Learn. Knowl. Extr. 2021, 3, 771–787.