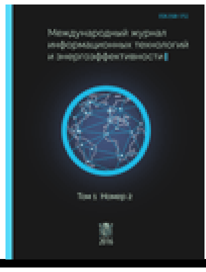
Воробьев В.В., Роза М.П. Интеллектуальный анализ процессов и автоматическая группировка на основе плотности для расширенного анализа бизнес-процессов: краткий обзор // Международный журнал информационных технологий и энергоэффективности.– 2024. –
Т. 9 № 1(39) с. 39–43

ОТКРЫТАЯ НАУКА
издательство

Международный журнал информационных технологий и энергоэффективности

Сайт журнала:
http://www.openaccessscience.ru/index.php/ijcse/



УДК 004.89

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ПРОЦЕССОВ И АВТОМАТИЧЕСКАЯ ГРУППИРОВКА НА ОСНОВЕ ПЛОТНОСТИ ДЛЯ РАСШИРЕННОГО АНАЛИЗА БИЗНЕС-ПРОЦЕССОВ: КРАТКИЙ ОБЗОР

[1]Воробьев В.В., Роза М.П.
ФГБОУ ВО "СИБИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ НАУКИ И ТЕХНОЛОГИЙ ИМЕНИ АКАДЕМИКА М.Ф. РЕШЕТНЕВА", Красноярск, Россия (660037,город Красноярск, пр-кт Им.Газеты "Красноярский Рабочий", д. 31), e-mail: [1]vvlad1997@mail.ru

Целью исследования является предоставление всестороннего обзора современных методов автоматической группировки на основе плотности при интеллектуальном анализе небольших данных, подчеркивая влияние более эффективных алгоритмов на эти методы. Основная цель – определить оптимальные алгоритмы для различных методов автоматической группировки на основе плотности, специально адаптированные для конкретных приложений. В ходе обсуждения в тексте освещаются вопросы алгоритмической точности, устойчивости к шуму и умения обрабатывать выбросы в контексте методов автоматической группировки на основе плотности.

Ключевые слова: Методы автоматической группировки на основе плотности, интеллектуальный анализ процессов, методы кластеризации, анализ бизнес-процессов, пространственная кластеризация приложений с шумом на основе плотности.

INTELLIGENT PROCESS ANALYSIS AND AUTOMATIC DENSITY-BASED GROUPING FOR ADVANCED BUSINESS PROCESS ANALYSIS: A BRIEF OVERVIEW

[1]Vorobyov V.V., Rosa M.P.
RESHETNEV SIBERIAN STATE UNIVERSITY OF SCIENCE AND TECHNOLOGY, Krasnoyarsk, Russia (660037, Krasnoyarsk, Newspaper "Krasnoyarsk Worker" Ave., 31), e-mail: [1]vvlad1997@mail.ru

Study aims to offer a comprehensive review of state-of-the-art automatic grouping methods based on density in small data mining, emphasizing the influence of more efficient algorithms on these techniques. The primary objective is to identify optimal algorithms for various automatic grouping methods based on density, specifically tailored to distinct applications. Through the discussion, the text highlights insights into algorithmic accuracy, robustness to noise, and proficiency in handling outliers within the context of automatic grouping methods based on density.

Keywords: Density-based automatic grouping methods, process mining, clustering methods, business process analysis, HDBSCAN.

Today, many software companies are actively collecting and analyzing data on user command usage in their applications. This provides the potential to extract valuable information about users' workflows and improve the efficiency of the software application. This article provides an overview of the possibilities of using clustering methods and process detection algorithms to analyze user workflows. The evaluation was carried out based on quantitative and qualitative parameters such as suitability, simplicity, interpretability and usefulness of the process models. Advances in density-based automatic grouping techniques have made it possible for organizations to efficiently collect and store data in much larger volumes than in the past. This shift has left many companies faced with massive amounts of unstructured data waiting to be processed and analyzed [1]. The benefits of successfully using data in the context of automatic grouping and process mining techniques are significant and can be applied to a variety of business areas. It is for this reason that many organizations are seeking to become more data-driven when making strategic business decisions.

One of the key sources of data typically obtained from applications is event logs, which record sequential user actions such as clicks and commands within the system. Applying automatic density-based clustering and process mining techniques to this data can reveal structures, patterns, and anomalies in users' workflows. The results of analyzing command usage data in event logs can shed light on patterns and trends in user behavior, providing a deeper understanding of their typical workflows.

Manually modeling workflows is often a labor-intensive and subjective process, and the models produced this way may differ from actual processes. By using density-based automatic grouping techniques, process mining becomes a powerful tool for automating the workflow modeling process of process mining. These methods open up new opportunities for effective analysis and optimization of business processes in conditions of large volumes of data, and help improve the efficiency of organizations in various fields of activity.

Scientific discourse surrounding the application of clustering methods and process detection algorithms in user workflow analysis has seen a proliferation of research endeavors, each contributing a nuanced layer to the understanding of these intricate processes.

Moreover, the application of clustering techniques in user workflow analysis is gaining momentum due to studies that explore the intricacies of identifying user behavior patterns [2]. The research employs a combination of clustering algorithms and process mining techniques to uncover hidden structures within event logs. The findings elucidate the potential of these methods to discern meaningful patterns, offering valuable insights into user interactions and aiding in the optimization of software applications.

In the domain of workflow modeling automation, density-based automatic grouping techniques have been subject to rigorous examination. Some work integrates the transformative potential of these methods into optimizing the analysis process, especially when compared to manual methodologies [3]. The research presents empirical evidence showcasing not only the accuracy of the automated models but also the substantial time savings achieved, affirming the practical viability of density-based clustering in expediting workflow modeling.

In the realm of automatic grouping techniques, the advances in density-based clustering algorithms have been a focal point of investigation. Some studies examine the effectiveness of these algorithms when processing large amounts of data [4]. The study reveals that density-based approaches, particularly DBSCAN, exhibit superior performance in scalability compared to other clustering methods. This empirical evidence aligns with the practical challenges faced by organizations dealing with massive datasets, reaffirming the relevance of density-based clustering techniques in real-world scenarios.

Since many separation methods are based on measuring the distance between objects, they are limited to finding clusters that have a spherical shape and have difficulty identifying clusters of arbitrary shapes. In contrast, density-based approaches focus on identifying areas with high density of objects and adding new objects to the cluster if the density in the specified area exceeds a specified threshold. Density can be defined, for example, as the number of other objects within a certain radius of a given object. Commonly used density-based algorithms include DBSCAN, OPTICS, and DENCLUE.

In the context of process mining, where the focus is on the analysis and modeling of business processes, similar techniques can be applied to highlight structural features in process trace data, allowing the discovery and description of various patterns and relationships between process events.

In the literature there are methods aimed at processing data representing business processes. In the context of process mining, these methods are applied in two key directions: first, using clustering methods; secondly, through business process analysis. Clustering, as an important stage in analysis, allows you to divide an extensive business process into more detailed components. One of the main clustering methods for traces in this context is HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise), which is based on the agglomerative clustering principle [5]. What's unique is that, rather than defining a predetermined number of clusters to divide the data into, the algorithm starts by assigning each data point its own cluster. Then, at each iteration, clusters are merged according to certain criteria, such as the distance between data points. Over the course of iterations, the algorithm keeps track of the current clusters and corresponding data points until all clusters merge into one overall cluster. The result is a clustering tree with one cluster at the top, created in the last iteration, and a cluster for each data point at the bottom, created at the first iteration. Intermediate levels represent clusters formed after each iteration of the algorithm. Such clustering methods provide the ability to group traces or clusters of similar sequences of actions. This stage of data preprocessing provides the basis for subsequent modeling of each cluster separately.

This approach allows you to create individual process models for each cluster, treating it as a separate set of event logs. In the context of clustering methods based on agglomerative approaches, such as HDBSCAN, the selection of a cut point in the clustering tree is based on various criteria, such as the number of clusters or the minimum distance between clusters. HDBSCAN is an improvement on the predecessor DBSCAN method, which also uses a density-based agglomerative method. However, HDBSCAN differs in that the selection of the cut point in the clustering tree is done at various locations rather than at a fixed level. This allows more data points to be included in clusters and noise detected at lower levels of the tree to be more effectively discarded.

Such hierarchical clustering methods furnish the ability to group traces or clusters exhibiting similar sequences of actions, unraveling hidden patterns within the labyrinth of business processes. This critical stage of data preprocessing lays the foundation for subsequent endeavors, providing a structured framework for modeling each cluster individually. One novel feature of this method is the separation of distinct process models for every cluster, handling them as separate sets of event logs. This granularity makes it possible to explore various behaviors that are contained inside each cluster in a nuanced manner, which facilitates a more focused and perceptive investigation of the underlying processes.

Due to its complex hierarchical and adaptive clustering methodology, HDBSCAN is a powerful tool for process mining and data analysis. The dynamic nature of its algorithmic journey, coupled with the ability to finely tune clustering outcomes, amplifies its efficacy in unraveling the complexities embedded in business processes, ultimately paving the way for more informed decision-making and operational optimization.

The scientific discourse surrounding density-based algorithms extends to their application in business process analysis. A comprehensive review of the work of previous researchers nuances the application of clustering methods to trace trace data, highlighting the role of density-based algorithms in identifying hidden structural features [6]. The study contributes a comprehensive overview of the landscape, providing a foundation for subsequent research endeavors in the evolving field of business process analysis.

As organizations increasingly recognize the transformative potential of becoming more data-driven, the scientific underpinnings of clustering methods and process detection algorithms become paramount. The empirical findings gleaned from diverse research studies underscore the versatility, scalability, and practical efficacy of these methods in unraveling the complexities of user workflows. By leveraging the collective insights of these scientific inquiries, businesses can navigate the evolving terrain of data analytics with a nuanced understanding of the tools that drive innovation and efficiency.

In summary, the complex subject of user workflow analysis, supported by process detection algorithms and clustering techniques, is characterized by a wide range of research interests. Time-series analysis is a critical component of clustering techniques, and studies show that comprehending user processes requires an awareness of temporal dynamics. The usefulness of these analyses is demonstrated by real-world applications, especially when it comes to optimizing software interfaces depending on user behavior.

Collectively, these diverse research directions contribute to a comprehensive and evolving understanding of user workflow analysis. The integration of various perspectives, from ensemble techniques to ethical considerations, enriches the scholarly discourse and sets the stage for continued advancements in the field.

Список литературы

1. Хафизур М. и др. Сокращение накладных расходов на автономную реляционную базу данных за счет классификации данных: dis. – Дуэт, 2023.
2. Рен Р. и др. Моделирование дорожно-транспортных происшествий в целях безопасности: внедрение подключенных и автоматизированных транспортных средств на дорогах //Анализ и предотвращение дорожно-транспортных происшествий. – 2023. – Т. 186. – С. 107021.
3. Лю Ю., Ким К. С. Фреймворк проектирования продукта на основе искусственного интеллекта с синергетическим сочетанием генетического алгоритма и оптимизации роя частиц //Мягкие вычисления. – 2023. – С. 1-18.
4. Чжан С., Патрас П., Хаддади Х. Глубокое обучение в мобильных и беспроводных сетях: обзор //IEEE Communications surveys & tutorials. – 2019. – Т. 21. – №. 3. – С. 2224-2287.
5. Хадер М., Аль-Наймат Г. Алгоритмы кластеризации больших данных на основе плотности с использованием фреймворка MapReduce: комплексное исследование //ACM Computing Surveys (CSUR). – 2020. – Т. 53. – №. 5. – С. 1-38.

6. Амини А., Вах Т. Ю., Сабухи Х. Об алгоритмах кластеризации потоков данных на основе плотности: обзор //Журнал компьютерных наук и технологий. – 2014. – Т. 29. – С. 116-141.

References

1. Hafizur M. et al. Reducing overhead on standalone relational database through data classification: дис. – Duet, 2023.
2. Ren R. et al. Vehicle crash simulations for safety: Introduction of connected and automated vehicles on the roadways //Accident Analysis & Prevention. – 2023. – Т. 186. – С. 107021.
3. Liu Y., Kim K. S. An artificial-intelligence-driven product design framework with a synergistic combination of Genetic Algorithm and Particle Swarm Optimization //Soft Computing. – 2023. – С. 1-18.
4. Zhang C., Patras P., Haddadi H. Deep learning in mobile and wireless networking: A survey //IEEE Communications surveys & tutorials. – 2019. – Т. 21. – №. 3. – С. 2224-2287.
5. Khader M., Al-Naymat G. Density-based algorithms for big data clustering using MapReduce framework: A Comprehensive Study //ACM Computing Surveys (CSUR). – 2020. – Т. 53. – №. 5. – С. 1-38.
6. Amini A., Wah T. Y., Saboohi H. On density-based data streams clustering algorithms: A survey //Journal of Computer Science and Technology. – 2014. – Т. 29. – С. 116-141.