

(چکیده توسعه یافته Extended Abstract)

عنوان:

بررسی روش‌های استخراج قانون در داده کاوی: مرور نظام‌مند مطالعات

نام نویسندگان:

سارا انصاری* دانشجوی کارشناسی ارشد مهندسی صنایع؛
دانشکده مهندسی صنایع و سیستم‌ها؛ دانشگاه
صنعتی اصفهان؛ اصفهان، ایران
saraansari@in.iut.ac.ir
صبا صارمی نیا استادیار؛ دانشکده مهندسی صنایع و سیستم‌ها؛
دانشگاه صنعتی اصفهان؛ اصفهان، ایران
s.sareminia@iut.ac.ir

مقدمه:

روش‌های استخراج قانون در داده کاوی به دلیل شفافیت، تفسیرپذیری و کاربرد گسترده در سیستم‌های هوشمند، مورد توجه روزافزون پژوهشگران قرار گرفته‌اند. این روش‌ها امکان شناسایی الگوهای نهفته در داده‌های ساختاریافته و پیچیده را فراهم می‌کنند. پژوهش حاضر با بهره‌گیری از چارچوب مرور نظام‌مند، به بررسی انتقادی استراتژی‌ها و معیارهای عملکردی استخراج قوانین پرداخته و شکاف‌های پژوهشی موجود را شناسایی می‌کند. پژوهش‌های بررسی شده، شامل ۱۹ مقاله منتخب از میان ۶۷۸ مقاله اولیه است. پرسش اصلی مطالعه، شناخت روش‌های مؤثر در استخراج قانون و تحلیل محدودیت‌ها و چالش‌های آن‌ها با هدف ارائه تصویری جامع از وضعیت کنونی و جهت‌دهی به مطالعات آینده است.

روش‌شناسی:

این پژوهش به صورت یک مطالعه مروری نظام‌مند با رویکرد تحلیلی و انتقادی انجام گرفت و هدف آن بررسی جامع، ساخت‌یافته و انتقادی روش‌های استخراج قانون در حوزه داده کاوی بود. جامعه آماری پژوهش را کلیه مقالات علمی منتشرشده در بازه زمانی ۲۰۱۵ تا ۲۰۲۴ تشکیل داد که از پایگاه‌های معتبر بین‌المللی از جمله ScienceDirect، IEEE، Springer، Emerald و Scopus بازیابی شدند. گردآوری داده‌ها از طریق جست‌وجوی نظام‌مند با بهره‌گیری از کلیدواژه‌های «Rule-based system» و «Rule induction» در عنوان، چکیده و کلیدواژه مقالات صورت گرفت. در گام نخست، ۶۷۸ مقاله شناسایی شد که با پیروی از چارچوب پریسما (PRISMA)، طی فرایندی چندمرحله‌ای شامل حذف موارد تکراری، ارزیابی تطابق محتوای موضوعی، بررسی روش‌شناسی و سنجش کیفیت منابع، در نهایت ۱۹ مقاله برای تحلیل نهایی انتخاب گردید.

برای تجزیه و تحلیل داده‌ها، از ترکیبی از روش‌های تحلیل کیفی و آماری بهره گرفته شد. تحلیل در سه محور اصلی شامل: (۱) ارزیابی انتقادی استراتژی‌های استخراج قانون و معیارهای عملکرد آن‌ها، (۲) شناسایی شکاف‌های پژوهشی، و (۳) تحلیل محتوای متنی مقالات منتخب صورت گرفت. ابزار تحلیل متن VOS viewer برای شناسایی مفاهیم کلیدی، ابرکلمات و روابط میان آن‌ها به کار گرفته شد. این فرایند به ترسیم تصویری جامع از وضعیت موجود و تبیین مسیرهای پژوهشی آتی در حوزه استخراج قانون کمک کرده است.

یافته‌های اصلی:

بر اساس تحلیل نظام‌مند انجام‌شده، یافته‌های این پژوهش در راستای پاسخ به سه پرسش اصلی به دست آمد. نخست، در خصوص این که «روش‌های رایج استخراج قانون از داده و معیارهای ارزیابی عملکرد آن‌ها در ادبیات پژوهش کدام‌اند و کدام‌یک بیشتر مورد توجه قرار گرفته‌اند؟» مشخص شد که پرکاربردترین استراتژی‌های استخراج قانون، رویکردهای «تقسیم و تسخیر» و «جدا و تسخیر» بوده‌اند. از میان معیارهای ارزیابی عملکرد، دقت (Accuracy) بیشترین تمرکز را به خود اختصاص داده بود، در حالی که پوشش (Coverage) و تفسیرپذیری (Interpretability) کمتر مورد توجه قرار گرفته‌اند.

در پاسخ به پرسش دوم مبنی بر این که «مشخصات و محدودیت‌های داده‌های ورودی چگونه بر استراتژی‌های استخراج قانون اثر می‌گذارند؟»، نتایج نشان داد که حدود ۸۹ درصد پژوهش‌ها صرفاً بر داده‌های ساختاریافته متمرکز بوده‌اند و تنها تعداد محدودی از مطالعات، داده‌های غیرساختاریافته مانند متن یا تصویر را در نظر گرفته‌اند. همچنین، چالش‌هایی نظیر داده‌های نامتعادل، ابعاد بالا و ویژگی‌های چندگانه به ندرت بررسی شده‌اند.

در پاسخ به پرسش سوم که به «شناسایی شکاف‌های پژوهشی در حوزه توسعه و استخراج قانون» می‌پردازد، مشخص شد که «نبود الگوریتم‌هایی که بتوانند به‌طور هم‌زمان دقت، پوشش و تفسیرپذیری را متعادل کنند، عدم توجه کافی به داده‌های پیچیده و ناهمگن، و کمبود رویکردهای مقیاس‌پذیر و قابل تعمیم»، از مهم‌ترین خلأهای موجود در ادبیات این حوزه هستند. این یافته‌ها، ضرورت طراحی و توسعه روش‌های پیشرفته‌تر و ترکیبی را برای پاسخ‌گویی به نیازهای واقعی تأیید می‌کند.

بحث و نتیجه‌گیری:

پژوهش حاضر با رویکردی نظام‌مند و تحلیلی، به بررسی جامع روش‌های استخراج قانون در حوزه داده‌کاوی پرداخته و سه یافته کلیدی و نوآورانه را برجسته می‌سازد. نخست آنکه، تمرکز غالب مطالعات بر بهبود دقت الگوریتم‌ها بوده، در حالی که دو معیار اساسی دیگر یعنی تفسیرپذیری و پوشش، به‌طور نسبی کمتر مورد توجه قرار گرفته‌اند. دوم، تحلیل مقالات نشان می‌دهد که اکثریت قریب به اتفاق پژوهش‌ها محدود به داده‌های

ساختاریافته بوده‌اند و چالش‌های مرتبط با داده‌های غیرساختاریافته، نامتعادل و با ابعاد بالا، کمتر در کانون توجه قرار گرفته است. سوم، اغلب روش‌های پیشنهادی در ایجاد توازن میان دقت، پوشش و تفسیرپذیری با محدودیت‌هایی مواجه هستند.

این یافته‌ها حاکی از آن است که برای پاسخ‌گویی به نیازهای پیچیده دنیای واقعی، توسعه رویکردهایی که بتوانند هم‌زمان سه معیار یادشده را برآورده سازند، امری ضروری است. پژوهش حاضر از طریق شناسایی شکاف‌های مفهومی و کاربردی موجود، به غنای دانش نظری در حوزه یادگیری قانون‌محور یاری رسانده و بستر مناسبی برای طراحی الگوریتم‌های ترکیبی و مقیاس‌پذیر فراهم می‌آورد.

در نهایت، پیشنهاد می‌شود پژوهش‌های آتی بر طراحی الگوریتم‌های تفسیرپذیر ویژه داده‌های غیرساختاریافته، و نیز بر ارتقای توان الگوریتم‌ها در مدیریت هم‌زمان معیارهای عملکردی در محیط‌های داده‌ای چندمنظوره متمرکز شوند.

کلیدواژه‌ها:

مرور نظام‌مند، استخراج قانون، داده‌کاوی، بهینه‌سازی قوانین، یادگیری ماشین

Extended Abstract

Title:

Review of Rule Extraction Methods in Data Mining: A Systematic Review of Studies

Names of Authors:

Sara Ansari*	Master's student in Industrial Engineering; Industrial and Systems Engineering Department; Isfahan University of Technology; Isfahan, Iran. saraansari@in.iut.ac.ir
Saba Sareminia	Associate Professor; Industrial and Systems Engineering Department; Isfahan University of Technology; Isfahan, Iran. s.sareminia@iut.ac.ir

Introduction:

Rule extraction methods in data mining have received increasing attention from scholars due to their inherent transparency, interpretability, and broad applicability in intelligent systems. These techniques enable the uncovering of patterns that may be hidden within both structured and complex datasets. This study employs a systematic review framework to critically assess the strategies and performance metrics relevant to rule extraction while identifying existing gaps in the research. Reviewed research includes 19 selected articles from an initial pool of 678. The primary question of this study is to identify effective methods for rule extraction and to analyse their limitations and challenges. The aim of the research is to provide a comprehensive overview of the current state and to guide future studies in this domain.

Methodology:

This study was conducted as a systematic literature review employing an analytical and critical approach, aiming to provide a comprehensive, structured, and in-depth examination of rule extraction methods within the field of data mining. The study population included all peer-reviewed scientific articles published between 2015 and 2024, retrieved from reputable international databases such as ScienceDirect, IEEE, Springer, Emerald, and Scopus. Data collection was carried out through a systematic search using the keywords “Rule-based system” and “Rule induction” across article titles, abstracts, and keywords.

Initially, 678 articles were identified. Following the PRISMA framework, a multi-stage screening process including duplicate removal, evaluation of thematic relevance, methodological review, and quality assessment was applied, resulting in the selection of 19 articles for final analysis.

A mixed-method approach, combining qualitative and quantitative analytical techniques, was employed for data analysis. The analysis focused on three key dimensions: (1) critical evaluation of rule extraction strategies and their performance metrics, (2) identification of research gaps, and (3) textual content analysis of the selected studies. The VOS viewer tool was utilized to extract key concepts, high-frequency terms, and the interrelationships among them. This process contributed to a comprehensive understanding of the current research landscape and helped illuminate future directions in rule extraction research.

Main findings:

Based on the systematic analysis conducted, the findings of this study were structured around addressing three core research questions.

First, in relation to the question “What are the prevalent rule extraction methods and performance evaluation criteria in the literature, and which have received the most attention?”, the analysis revealed that the most widely applied strategies were the divide-and-conquer and separate-and-conquer approaches. Among performance metrics, accuracy received the highest emphasis, whereas coverage and interpretability were comparatively underexplored.

Second, in response to “How do the characteristics and limitations of input data influence rule extraction strategies?”, the findings indicated that approximately 89% of the studies exclusively focused on structured data, with only a few incorporating unstructured formats such as text or images. Additionally, issues like imbalanced data, high dimensionality, and multivariate features were rarely examined.

Third, in addressing the question concerning “What are the research gaps in the development and application of rule extraction methods?”, it was found that key limitations included the lack of algorithms capable of simultaneously balancing accuracy, coverage, and interpretability, insufficient attention to complex and heterogeneous data, and a general absence of scalable and generalizable approaches.

These insights underscore the urgent need for designing and developing more advanced, hybrid methods to meet real-world challenges in rule extraction.

Discussion and conclusions:

This study adopts a systematic and analytical approach to provide a comprehensive review of rule extraction methods in the domain of data mining and presents three key and novel findings. First, the majority of existing studies have focused predominantly on enhancing algorithmic accuracy, while the two other fundamental criteria interpretability and coverage have received comparatively less attention. Second, the analysis reveals that an overwhelming proportion of the reviewed literature is confined to structured data, with limited exploration of challenges associated with unstructured, imbalanced, or high-dimensional datasets. Third, most of the proposed methods exhibit notable limitations in achieving a balance among accuracy, coverage, and interpretability.

These findings underscore the necessity of developing advanced approaches capable of simultaneously addressing all three performance criteria to meet the complex demands of real-world applications. By identifying both conceptual and practical gaps, this study contributes to the advancement of theoretical knowledge in the area of rule-based learning, and lays the groundwork for the development of scalable and hybrid algorithms.

Finally, it is recommended that future research focus on designing interpretable algorithms tailored for unstructured data, as well as enhancing the capacity of such algorithms to effectively manage multiple performance metrics in complex, multi-purpose data environments.

Keywords:

systematic review, rule extraction, data mining, rule optimization, machine learning