International Journal of Computer Science and Information Technology
ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 6, Number 2, Year 2025
DOI: https://doi.org/10.62051/ijcsit.v6n2.09
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Application of Time-Weighted Association Rule Mining in Urban Water Resources Management

Wenhao Jiang *, Gaohui Peng

School of Mathematics and Statistics, North China University of Water Resources and Electric Power, Zhengzhou, Henan, China

*Corresponding Author: Wenhao Jiang (Email: jiangwenh@foxmail.com)

ABSTRACT

Based on the water resource data of Zhengzhou City from 1999 to 2023, this study employs the time-weighted Apriori algorithm (with a decay factor of 0.85) combined with static-dynamic discretization preprocessing methods to mine dynamic association rules for three major water use types: integrated urban-rural water use, industrial water use, and agricultural water use. The results indicate that integrated urban-rural water use exhibits rigid growth, with its proportion rising from 18.4% to 67.8%; industrial water use has achieved "reduced volume with enhanced efficiency," with its proportion decreasing from 24.9% to 11.2%; and agricultural water use presents three coexisting patterns: stable, increasing, and decreasing. Water resource allocation strategies have shifted from "production-promotes-water-use" to "water-use-determines-production," with newly added water resources primarily allocated to urban-rural water use.

KEYWORDS

Water resource management; Time-weighted Apriori; Association rule mining; Water use structure; Zhengzhou City

1. INTRODUCTION

The water resources crisis has emerged as a fundamental constraint on global sustainable development. China experiences acute water stress, with per capita water resources substantially lower than the global average, leading to severe supply-demand imbalances. Traditional management strategies relying on static analysis are inadequate for capturing the dynamic evolution of water resource systems, necessitating the adoption of temporally-aware data mining methodologies. Among China's nine national central cities, Zhengzhou faces particularly severe water scarcity challenges, ranking as one of the most water-stressed urban centers. The selection of Zhengzhou as our case study is motivated by the typicality of its water resource constraints, ensuring that our findings can offer transferable insights for water management in similarly challenged urban contexts.

Since Agrawal et al. [1] proposed the Apriori algorithm, association rule mining has been widely applied in various domains. However, traditional methods treat historical data with equal weights, overlooking the dynamic evolutionary characteristics of data. As research has progressed, several improvements have been proposed: Wang et al. [2] introduced weighting mechanisms to address the differences in item importance; Han et al. [3] proposed segmented periodic patterns to capture temporal characteristics; Tanbeer et al. [4] constructed PF-tree structures to handle periodic-frequent patterns; and Zhuang et al. [5] explored association relationships in multidimensional time series. In recent years, Gan et al. [6] innovatively proposed the RHEWI (Recent High Expected Weighted Itemsets) concept, which simultaneously considers timeliness, weights, and uncertainty; while Yun et al. [7] developed the WSpan algorithm that incorporates weight constraints into sequential pattern mining. Although time-weighting mechanisms have emerged as an important research direction, existing methods predominantly employ fixed weight strategies that fail to reflect how data importance decays over time. This limitation is particularly evident in domains such as water resource management, where decision-making requires balancing historical trends with recent changes. Therefore, there is an urgent need for more flexible time-weighting mechanisms to enhance both the timeliness and accuracy of decision-making processes.

2. METHODOLOGY

2.1. Time-Weighted Apriori Algorithm

Given the limitations of traditional Apriori algorithm in processing temporal data, this study introduces a time-weighting mechanism. The decay factor is used to define the decay rate of historical data weights over time, serving as a core parameter affecting rule mining effectiveness.

Let α be the decay factor, $t_{\max}$ be the most recent time point, and t be a historical time point. The time weight function is defined as:

$$\omega(t) = \lambda^{(t_{max}-t)} \tag{1}$$

When α approaches 1, historical data weights decay slowly; when α approaches 0, historical data weights decay rapidly.

Based on the time weight function, weighted support is defined as the weighted average of support values at each time point, expressed mathematically as equation (2):

$$Support(X \Rightarrow Y)_{weighted} = \frac{\sum_t \omega(t) \cdot \frac{count_t(X \cup Y)}{|D_t|}}{\sum_t \omega(t)} \tag{2}$$

Where $count_t(X \cup Y)$ represents the number of transactions containing both itemsets X and Y in D_t at time point t, $|D_t|$ represents the total number of records in the dataset at time point t. Compared to traditional support, weighted support better reflects the importance of the recent occurrence frequency of itemsets.

The weighted confidence is calculated based on the weighted support, as shown in Equation (3):

$$confidence(X \Rightarrow Y)_{weighted} = \frac{\sum_t \omega(t) \cdot \frac{count_t(X \cup Y)}{|D_t|}}{\sum_t \omega(t) \cdot \frac{count_t(X)}{|D_t|}} \tag{3}$$

The lift is also redefined based on the weighted support, specifically given by Equation (4):

$$Lift(X \Rightarrow Y)_{weighted} = \frac{confidence(X \Rightarrow Y)_{weighted}}{Support(Y)_{weighted}} \tag{4}$$

2.2. Evaluation Metrics

To comprehensively evaluate the quality of association rule mining results under different decay factors, this study constructs a multi-dimensional evaluation metric system. In addition to the traditional support, confidence, and lift measures, the following supplementary metrics are introduced:

Coverage: This metric measures the explanatory power of the rule set on the original data, calculated as the ratio of transactions covered by the rules to the total number of transactions. High coverage indicates that the rules possess strong generalizability.

Diversity: This metric assesses the degree of difference between rules to avoid mining numerous similar redundant rules. It is quantified by calculating the overlap between the antecedents and consequents of rules.

In practical applications, radar chart visualization is employed to comprehensively display various metrics, facilitating intuitive comparison of rule quality under different decay factors. By setting minimum thresholds for each metric, high-quality rule sets that simultaneously satisfy multi-dimensional requirements are selected.

3. RESULTS AND DISCUSSION

3.1. Data Description and Preprocessing

The data for this study were obtained from the Zhengzhou Water Resources Bulletin, spanning from 1999 to 2023. The dataset encompasses three major categories: urban-rural integrated water use, industrial water use, and agricultural water use, along with related natural factors and socioeconomic indicators. The original data are organized on an annual basis, containing multi-dimensional information such as absolute water consumption values, total water resources, and precipitation.

Data preprocessing employs a combined approach of static and dynamic discretization. Static discretization captures the overall distribution characteristics of the data by dividing continuous variables into categories such as "high," "medium," and "low" based on statistical quantiles. Dynamic discretization focuses on temporal trends by converting annual change rates into five levels: "significant increase," "increase," "stable," "decrease," and "significant decrease." This dual discretization strategy preserves both the absolute level information and highlights dynamic evolution characteristics, laying the foundation for subsequent association rule mining. After preprocessing, a structured dataset containing 25 annual observations was formed.

3.2. Optimal Decay Factor Selection

To comprehensively evaluate the quality of mining rules under different decay factors, this study conducts analysis from five dimensions: average support, average confidence, average lift, rule coverage, and rule diversity. Figure 1 presents radar charts of these five metrics under different decay factors. The radar chart analysis reveals that when the decay factor is 0.75 or 0.8, the algorithm performs well in terms of average support and average confidence, but exhibits low rule diversity. When the decay factor is 0.95, rule diversity reaches its maximum, while average support and average confidence are relatively low. From an overall perspective, the radar chart corresponding to $\alpha = 0.85$ demonstrates the most balanced performance, with all metrics falling within ideal ranges and no apparent weaknesses. The selection of $\alpha = 0.85$ as the optimal decay factor holds practical significance. This value implies that data from the previous year receives a weight of 0.85, data from two years ago receives 0.72, and data from three years ago receives 0.61, and so forth. This weight distribution ensures the dominance of recent data while reasonably preserving the reference value of medium- and long-term historical information. From the perspective of water resource management practice, this precisely aligns with the temporal scale of water resource planning—responding to recent supply-demand changes while considering the continuity of medium-term trends. Therefore, all subsequent analyses are conducted based on association rule mining with $\alpha = 0.85$.

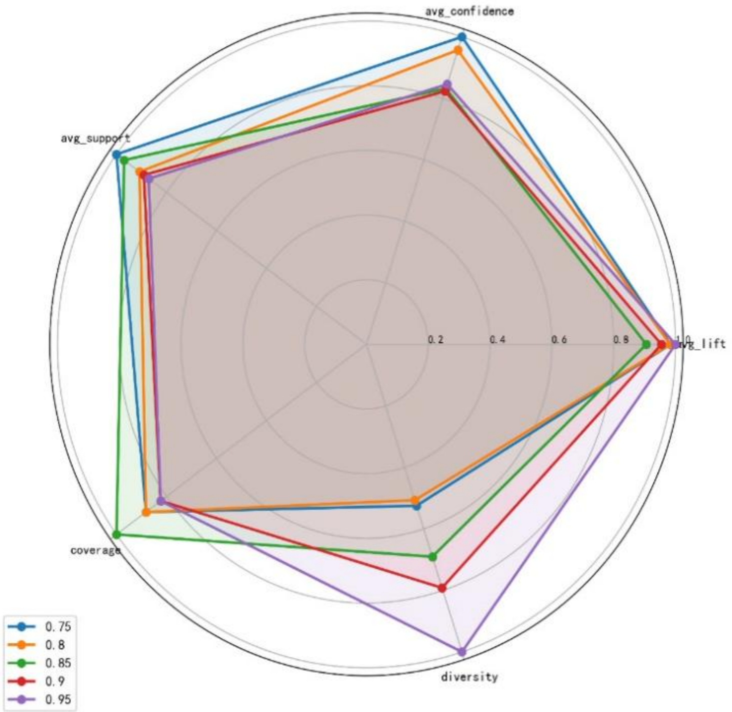

Figure 1. Radar Chart of Evaluation Indicators for Different Attenuation Factors

3.3. Association Rule Analysis

3.3.1. Urban-Rural Water Use Patterns

The urban-rural integrated water consumption increased approximately 5.4-fold during 1999-2023, with its proportion of total water consumption surging from 18.4% to 67.8%, establishing it as the primary water resource consumer in Zhengzhou City. Table 1 demonstrates that urban-rural integrated water use exhibits two distinct patterns: "gradual growth" and "rapid growth."

In the "gradual growth" pattern, the decrease in total water consumption demonstrates the strongest association (confidence 0.9753, lift 2.4018), revealing the rigid demand characteristics of urban-rural water use—urban-rural water consumption maintains growth even when total water use declines. The associations with medium water production coefficient (confidence 0.8938) and low effective irrigation area (confidence 0.7495) respectively reflect the supporting role of water resource availability and the substitution relationship between sectoral water uses.

Under the "rapid growth" pattern, low surface water resources exhibit high correlation (confidence 0.8629), indicating that cities may intensify groundwater extraction or seek external water transfers under resource constraints. The close relationship between increased total water consumption and rapid growth (confidence 0.8543) confirms that urban-rural water use has become the core driver of overall water consumption growth.

Table 1. Association Rules Analysis Results for Urban-Rural Water Use

Urban-Rural Water Use	Influencing Factors (Antecedent)	Support	Confidence	Lift
Gradual Growth	Total water consumption change → Decrease	0.3325	0.9753	2.4018
Gradual Growth	Water production coefficient → Medium	0.1877	0.8938	2.2010
Gradual Growth	Effective irrigation area → Low	0.2701	0.7495	1.8458
Gradual Growth	Surface water-Groundwater resource overlap → Low	0.1568	0.7336	1.8064
Gradual Growth	Surface water resource quantity → Medium	0.3562	0.6603	1.6260
Gradual Growth	Water resource total change → Slight decrease	0.1568	0.6123	1.5080
Rapid Growth	Surface water resource quantity → Low	0.2690	0.8629	1.9239
Rapid Growth	Total water consumption change → Increase	0.3466	0.8543	1.9046
Rapid Growth	Effective irrigation area → Medium	0.1662	0.6246	1.3925

3.3.2. Industrial Water Use Patterns

Industrial water consumption exhibits a distinctive trend of "water reduction with output growth." During 1999-2023, industrial GDP increased nearly 13-fold, while water consumption proportion declined from 24.9% to 11.2%, reflecting the effectiveness of industrial transformation and upgrading. Table 2 shows that industrial water use primarily manifests in two patterns: "significant decrease" and "decrease."

Table 2. Association Rules Analysis Results for Industrial Water Use

Industrial Water Use	Influencing Factors (Antecedent)	Support	Confidence	Lift
Significant Decrease	Water resource total change → Significant increase	0.2655	0.9738	2.3059
Significant Decrease	Total water consumption change → Decrease	0.2890	0.8478	2.0077
Significant Decrease	Precipitation change → Significant increase	0.2465	0.7641	1.8095
Significant Decrease	Effective irrigation area → Low	0.2701	0.7495	1.7749
Significant Decrease	Water production coefficient → Medium	0.1568	0.7468	1.7684
Significant Decrease	Surface water-Groundwater resource overlap → High	0.1568	0.7336	1.7371
Significant Decrease	Groundwater resource quantity → Medium	0.3090	0.6231	1.4756
Significant Decrease	Total electricity sales → High	0.4033	0.6195	1.4671
Significant Decrease	Water supply structure → Groundwater dominant	0.4033	0.6195	1.4671
Significant Decrease	Water resource total change → Significant decrease	0.1568	0.6123	1.4501
Decrease	Surface water resource quantity → Low	0.2209	0.7085	2.6135
Decrease	Total water consumption change → Stable	0.1748	0.6901	2.5454

The "significant decrease" pattern correlates with 10 factors, among which the strong associations with significant increase in total water resources (confidence 0.9738, lift 2.3059) and decrease in total water consumption (confidence 0.8478) indicate that industrial water use continues to decline even with abundant water resources. Particularly noteworthy is the association between high total electricity sales and significant water reduction (confidence 0.6195), confirming the decoupling phenomenon between industrial economic growth and water consumption.

A key finding is that both significant increases and decreases in total water resources are associated with marked industrial water reduction. This reflects a shift in water resource allocation strategy: under varying resource conditions, industrial water use priority consistently decreases, with new water resources primarily allocated to urban-rural integrated use, embodying the transition from "production-driven consumption" to "water-constrained production."

3.3.3. Agricultural Water Use Patterns

Agricultural water consumption decreased from 813 million cubic meters in 1999 to 441.2 million cubic meters in 2023, a reduction of 45.7%, while effective irrigation area remained relatively stable, reflecting significant improvement in irrigation efficiency. Table 3 shows that agricultural water use exhibits three concurrent patterns: "stable," "increase," and "decrease."

In the "stable" pattern, the strongest association occurs at medium-high levels of surface-groundwater resource overlap (confidence 0.9044, lift 2.4805), indicating that complementary water resource utilization contributes to agricultural water stability. Under the "increase" pattern, significant decrease in total water resources associates with increased water use (confidence 0.6123), reflecting the rigid demand for food security. The "decrease" pattern is primarily influenced by normal precipitation years (confidence 0.6461) and groundwater resource constraints (confidence 0.6015).

The multi-pattern coexistence characteristic of agricultural water use contrasts sharply with the sustained growth of urban-rural water use and the continuous decline of industrial water use, highlighting its buffer role in the water resource allocation system. This flexibility provides adjustment space for optimal water resource allocation but also requires differentiated management strategies while ensuring food security.

Table 3. Association Rules Analysis Results for Agricultural Water Use

Agricultural Water Use	Influencing Factors (Antecedent)	Support	Confidence	Lift
Stable	Surface water-Groundwater resource overlap → Medium-High	0.1778	0.9044	2.4805
Stable	Water production coefficient → High	0.1521	0.7457	2.0452
Stable	Effective irrigation area → Medium	0.1682	0.6323	1.7342
Stable	Total water consumption change → Increase	0.2508	0.6182	1.6954
Increase	Water production coefficient → Medium	0.1568	0.7468	2.5318
Increase	Surface water-Groundwater resource overlap → High	0.1568	0.7336	2.4870
Increase	Water resource total change → Significant decrease	0.1568	0.6123	2.0761
Decrease	Precipitation change → Normal year	0.2637	0.6461	1.8978
Decrease	Groundwater resource quantity → Low	0.2647	0.6015	1.7670

4. CONCLUSION

4.1. Summary of Findings

This study improved the traditional Apriori algorithm by introducing a time-weighting mechanism to conduct dynamic association rule mining on Zhengzhou City's water resource supply-demand system from 1999-2023, revealing the evolution patterns and driving mechanisms of three major water use types. The main conclusions are as follows:

First, the optimal decay factor $\alpha=0.85$ was determined, achieving a balance between timeliness and stability. The improved algorithm successfully identified dynamic association patterns that traditional methods struggle to discover.

Second, urban-rural integrated water use shows rigid growth, with its proportion rising from 18.4% to 67.8%; industrial water use achieves "reduction with efficiency gains," with its proportion dropping to 11.2% while output value increased 13-fold; agricultural water consumption decreased 45.7% overall, exhibiting three concurrent patterns.

Third, association rule analysis reveals that regardless of water resource abundance or scarcity, industrial water use priority consistently decreases, with new water resources primarily allocated to urban-rural integrated use. This transition from "production-driven consumption" to "water-constrained production" reflects an important transformation in water resource management philosophy from supporting economic growth to ensuring livelihood and ecological security.

Fourth, urban-rural water use is influenced by total quantity changes and water production coefficients; industrial water use correlates with total resources and electricity consumption; agricultural water use is comprehensively influenced by multiple factors. These association patterns provide a basis for precise policy implementation.

4.2. Implications for Water Resource Management

Based on the research findings, the following water resource management recommendations are proposed:

To address the rigid growth in urban-rural water consumption, strengthening water conservation measures becomes paramount. Given the sustained growth trend and priority status of urban-rural integrated water use, efforts should accelerate water-saving city construction, focusing on enhancing residents' water conservation awareness, promoting water-saving appliances, and improving tiered water pricing mechanisms.

For the industrial sector, the focus should shift to consolidating existing water conservation achievements while promoting deeper transformation. While maintaining the momentum of industrial water use efficiency improvement, policies should further optimize industrial structure, develop low water consumption and high value-added industries, and establish mechanisms linking water use efficiency with industrial access requirements.

Agricultural water management presents unique opportunities due to its flexibility characteristics. Establishing dynamic allocation mechanisms based on water resource conditions can optimize agricultural water use while maintaining its regulatory function. This includes moderately adjusting planting structures while ensuring food security and promoting efficient water-saving irrigation technologies.

4.3. Research Prospects

This study has achieved preliminary results in time-weighted association rule mining, but there remains room for improvement:

From an algorithmic perspective, future research could explore adaptive decay factor mechanisms that dynamically adjust weight allocation based on data characteristics. Additionally, introducing spatial dimensions to construct spatiotemporal association rule mining frameworks and integrating machine learning methods could substantially enhance rule predictive capabilities.

In terms of application expansion, the methodology developed here shows promise for broader implementation. The research framework could be extended to other water resource management domains such as water quality evolution and aquatic ecological protection. Furthermore, expanding to larger spatial scales for watershed or regional comparative studies, combined with climate change scenarios, would provide valuable insights for assessing future water resource supply-demand patterns.

REFERENCES

[1] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," in Proc. 1993 ACM SIGMOD Int. Conf. Management of Data, 1993, pp. 207-216.

[2] W. Wang, J. Yang, and P. S. Yu, "Efficient mining of weighted association rules (WAR)," in Proc. 6th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2000, pp. 270-274.

[3] J. Han, W. Gong, and Y. Yin, "Mining segment-wise periodic patterns in time-related databases," in Proc. 4th Int. Conf. Knowledge Discovery and Data Mining (KDD), 1998, pp. 214-218.

[4] S. K. Tanbeer, C. F. Ahmed, B.-S. Jeong, et al., "Discovering periodic-frequent patterns in transactional databases," in Proc. 13th Pacific-Asia Conf. Knowledge Discovery and Data Mining (PAKDD), Bangkok, Thailand, Apr. 27-30, 2009, pp. 242-253.

[5] D. E. Zhuang, G. C. Li, and A. K. Wong, "Discovery of temporal associations in multivariate time series," IEEE Trans. Knowl. Data Eng., vol. 26, no. 12, pp. 2969-2982, Dec. 2014.

[6] W. Gan, J. C.-W. Lin, P. Fournier-Viger, et al., "Extracting recent weighted-based patterns from uncertain temporal databases," Eng. Appl. Artif. Intell., vol. 61, pp. 161-172, May 2017.

[7] U. Yun and J. J. Leggett, "WSpan: Weighted sequential pattern mining in large sequence databases," in Proc. 3rd Int. IEEE Conf. Intelligent Systems, Sep. 2006, pp. 512-517.