

Article

Enhancing Forest Fire Risk Assessment: An Ontology-Based Approach with Improved Continuous Apriori Algorithm

Yumin Dong^{1,2,3}, Ziyang Li^{1,2,*} and Changzuo Xie^{1,2,3}¹ Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China² Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, Chinese Academy of Sciences, Beijing 100094, China³ School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: zyli@aircas.ac.cn

Abstract: Forest fires are sudden and difficult to extinguish, so early risk assessment is crucial. However, there are currently a lack of suitable knowledge-mining algorithms for forest fire risk assessment. This article proposes an improved continuous Apriori algorithm to mining forest fire rules by introducing prior knowledge to classify input data and enhance its ability to process continuous data. Meanwhile, it constructs an ontology to provide a standardized expression platform for forest fire risk assessment. The improved continuous Apriori algorithm cooperates with ontology and applies the mining rules to the forest fire risk assessment results. The proposed method is validated using the forest fire data from the Bejaia region in Algeria. The results show that the improved continuous Apriori algorithm is superior to the raw Apriori algorithm and can mine the rules ignored by the raw Apriori algorithm. Compared to the raw Apriori algorithm, the number of generated rules increased by 191.67%. The method presented here can be used to enhance forest fire risk assessments and contribute to the generation and sharing of forest-fire-related knowledge, thereby alleviating the problem of insufficient knowledge in forest fire risk assessment.

Keywords: forest fires; risk assessment; Apriori algorithm; ontology; association rule mining



Citation: Dong, Y.; Li, Z.; Xie, C.

Enhancing Forest Fire Risk

Assessment: An Ontology-Based Approach with Improved Continuous Apriori Algorithm. *Forests* **2024**, *15*, 967. <https://doi.org/10.3390/f15060967>

Academic Editor: Manuel

Marey-Pérez

Received: 6 May 2024

Revised: 25 May 2024

Accepted: 28 May 2024

Published: 31 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Forest fires are sudden and difficult to extinguish, so conducting a risk assessment of forest fires is particularly important for forest fire management [1]. At present, the assessment of forest fire risk uses a wide range of data sources, including remote sensing data, in situ detection data, basic geographic data, etc. [2–4]. While a large number of data sources improve the potential of forest fire risk assessment, they also bring about issues such as data heterogeneity and semantic gaps [5]. The increase in data complexity has increased the knowledge level required for data users to judge fire risk.

In response, Ontology, a hot knowledge engineering technology, widely accepted as “an explicit specification of a conceptualization [6]”, has been introduced into forest fire risk assessment [7–9]. Ontology can express concepts and their relationships in a structured way that both humans and computers can understand. Rules are a way of expressing knowledge [10], using ontology as the basis for concept expression. Through rule reasoning, relationships between concepts can be found, and the combination of ontology and rules can achieve efficient knowledge expression and sharing [11,12].

Existing knowledge-based researches [7–9] has focused on using ontology and rules to assess forest fire risk, demonstrating the effectiveness of ontology and rules in forest fire management. However, in these researches, some rules rely on complex and highly manual fire index methods [7], some use heuristic algorithm-based methods [8], and some use experimental or illustrative rules [9]. The process of mining rules has been overlooked to a

certain extent, and the proposed methods are also difficult to meet the needs of interpretable and scalable knowledge mining.

Association rule mining is a type of data mining algorithm, and it can uncover important and reliable rules between attributes in databases [13]. Representative association rule algorithms include Apriori [14] and Eclat [15], where Apriori scans the database more times and occupies less memory, while Eclat only scans the database once but occupies more memory. Considering the volume of forest fire dataset, the Apriori algorithm [16–18] is used more frequently than the Eclat algorithm in forest fire rules mining. However, the Apriori algorithm is designed to mine the relationship between discrete data. When conducting forest fire risk assessment, continuous data, such as temperature, needs to be processed. At this point, the Apriori algorithm may not be able to effectively mine the rules that users are interested in, which is overlooked in previous research [16–18]. For example, continuous data, such as temperature, can increase the forest fire risk as the numerical values increase. It is assumed that if the results of data mining show a high fire risk in a certain scenario at 20 °C, then under other unchanged scenarios, a similar scenario at 40 °C should have a higher fire risk. However, due to the rarity of a temperature of 40 °C, the importance of this rule is much lower than similar rules at 20 °C (which is measured by the frequency of data in the Apriori algorithm), which may lead to the neglect of important rules under extreme scenarios. Therefore, it is necessary to improve the Apriori algorithm when using it for forest fire rule mining to avoid this kind of neglect.

Current researchers have performed some work to improve the Apriori algorithm's support for continuous data, such as using K-means [19], distribution probability [20], membership functions [21], etc. However, these works did not provide sufficient attention to rules in extreme scenarios, which is extremely important in forest fire risk assessment.

In summary, the combination of the powerful expression ability of ontology and the Apriori algorithm can effectively utilize the mined rules in the forest fire assessment process. A unified and standardized expression platform is conducive to the automated dissemination and sharing of knowledge. However, the potential of this combination is limited by the current association rule mining algorithm.

This article uses ICAA (Improved Continuous Apriori Algorithm), a new improved Apriori algorithm that can be used for continuous data, to mine forest fire data and achieve automated knowledge generation, thereby alleviating the problem of the shortcomings of rule mining algorithms neglecting extreme scenario rules in knowledge-based fire risk assessment. Furthermore, ontology technology is introduced as a unified and standardized expression platform to standardize the semantics of heterogeneous data, combining the constructed ontology and forest fire rules generated by algorithms to achieve knowledge reasoning, thereby improving the automation level of forest fire risk assessment, and reducing the knowledge requirements for data users.

The architecture of this article is shown in Figure 1.

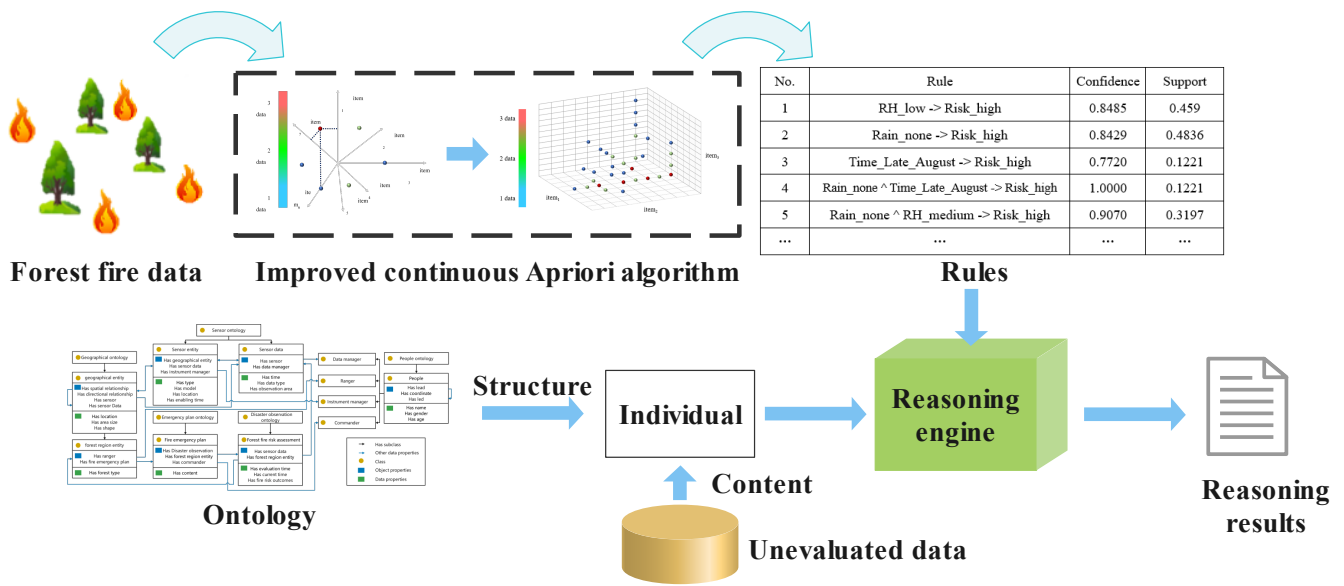


Figure 1. Architecture.

2. Materials and Methods

2.1. Research Area and Dataset

The research area is located in the Bejaia region of northeastern Algeria in northern Africa. Algeria is a country with a high incidence of forest fires, with an average of 31,300 ha of forest areas that were destroyed by fires annually [22]. The Bejaia region is rich in forest resources and is one of the most important forest areas in Algeria, with an area of 122,500 ha of forests, which includes 38% of the total forest surface in Algeria [23]. Meanwhile, the Bejaia region is located in the southern Mediterranean and is a hot spot for forest fires [24].

The dataset [25] used by the research includes temperature, humidity, wind speed, and rainfall in the Bejaia region at different times, as well as whether forest fires occurred on that day. The occurrence of forest fires is considered a high forest fire risk, while days without forest fires are considered to have low forest fire risk. Using this dataset for knowledge mining, the generated rules are combined with the ontology technology to achieve automatic semantic reasoning to support the assessment of forest fire risk.

2.2. Raw Apriori Algorithm and ICAA

2.2.1. Raw Apriori Algorithm

The raw Apriori algorithm considers the association mining of discrete data with a discrete dataset as input and a rule set as output. Assuming there is a discrete dataset:

$$DataSet = \{D_1, D_2, \dots, D_n\} \tag{1}$$

$$D_n = \{I_x, I_y, \dots, I_z\} \tag{2}$$

D_n is a piece of data in the dataset, I_x is an item in the data, and each piece of data contains several items. Assuming there are total of m items in the dataset, that is:

$$D_1 \cap D_2 \cap \dots \cap D_n = \{I_1, I_2, \dots, I_m\} \tag{3}$$

Then, D_n can be regarded as a point on an m -dimensional space, with a value range of $\{0, 1\}$ for each dimension, when:

$$D_{nm} = \begin{cases} 0 & I_m \notin D_n \\ 1 & I_m \in D_n \end{cases} \tag{4}$$

Among them, D_{nm} is the value of D_n in dimension m . In the q -dimensional ($q \leq m$) subspace of the m -dimensional space, data intersects at certain points. The more data at the intersection point, the more important the information it contains. This importance is measured by support in the raw Apriori algorithm:

$$supp\{I_x, I_y, \dots, I_z\}_q = P(I_x, I_y, \dots, I_z) = \frac{N(I_x, I_y, \dots, I_z)}{n} \tag{5}$$

Among them, $\{I_x, I_y, \dots, I_z\}_q$ represents the item set with q items, $P(I_x, I_y, \dots, I_z)$ represents the probability of the occurrence of the item set, which is estimated by dividing the number of occurrences of the item set, $N(I_x, I_y, \dots, I_z)$, by the total number of data, n .

Figure 2, an example with $m = 7$, shows blue, green, and red dots representing 1, 2, and 3 data points, respectively.

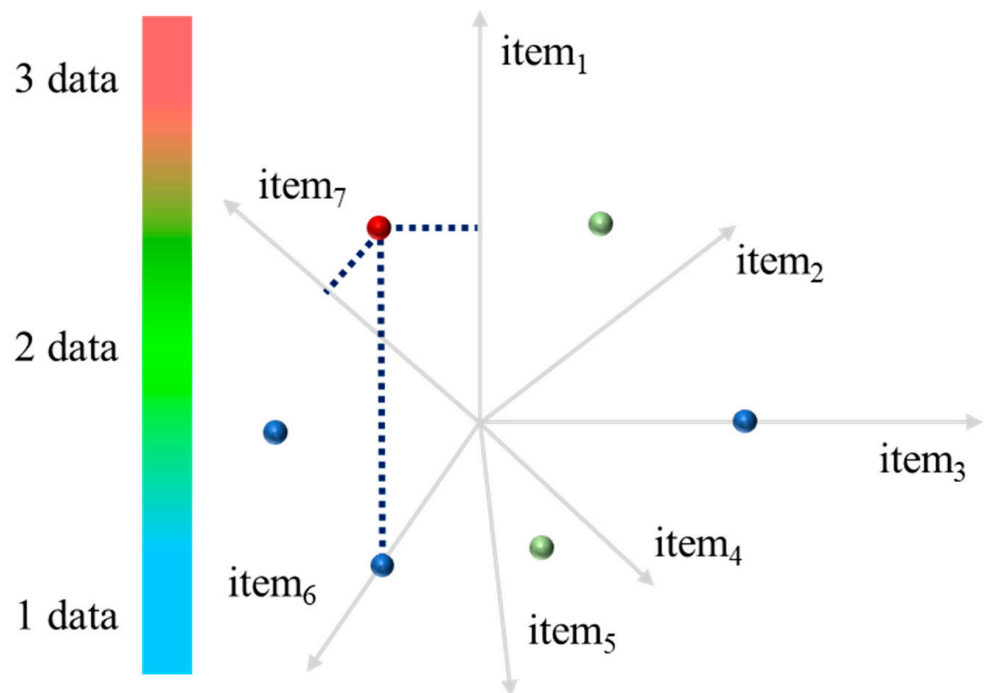


Figure 2. The raw Apriori algorithm represents discrete data as a set of points in a high-dimensional space.

There are total of ten data points in the Figure, and the intersection point with the most data is $\{item_1 = 1, item_6 = 1, item_7 = 1\}$. There are three intersection points. The support level for $\{item_1, item_6, item_7\}$ is:

$$supp\{item_1, item_6, item_7\} = \frac{N(item_1, item_6, item_7)}{n} = 0.3 \tag{6}$$

The raw Apriori algorithm measures the accuracy of generated rules by confidence:

$$conf(\{I_x, \dots\} \rightarrow \{I_y, \dots\}) = \frac{P(\{I_x, \dots\} \cup \{I_y, \dots\})}{P(\{I_x, \dots\})} = \frac{N(\{I_x, \dots\} \cup \{I_y, \dots\})}{N(\{I_x, \dots\})} \tag{7}$$

s.t. $\{I_x, \dots\} \cap \{I_y, \dots\} = \emptyset$

Unlike support, confidence is directional. $\{I_x, \dots\} \rightarrow \{I_y, \dots\}$ is called a rule, and P and N are the probabilities and number of occurrences of the item set, respectively. If the support and confidence are both greater than the set threshold, then the output the rule $\{I_x, \dots\} \rightarrow \{I_y, \dots\}$.

Taking the item set $\{item_1, item_6, item_7\}$ in Figure 2 as an example:

$$conf(\{item_6, item_7\} \rightarrow \{item_1\}) = \frac{N(\{item_6, item_7, item_1\})}{N(\{item_6, item_7\})} = \frac{3}{3+1} = 0.75 \quad (8)$$

The support and confidence of item set $\{item_1, item_6, item_7\}$ is 0.3 and 0.75, and if the set support and confidence threshold are to 0.1 and 0.7, respectively, then the Algorithm will output the rule $\{item_6, item_7\} \rightarrow \{item_1\}$.

2.2.2. ICAA

The raw Apriori algorithm calculates all item sets that exceed the support threshold, calculates the confidence level of each possible rule in each itemset, and finally outputs all rules that meet the confidence threshold.

When using the raw Apriori algorithm for rule mining on continuous data, it is necessary to discretize the continuous data. At present, most work extends the Apriori algorithm in terms of dimension quantity through discretization of continuous data, and a better method is to expand within a certain dimension. Based on this, this article proposes a new expression model for the ICAA, as shown in Figure 3, which improves the attention to the relationships within a certain class of continuous data dimensions by expanding the range of values in a single dimension to improve the neglect of important rules.

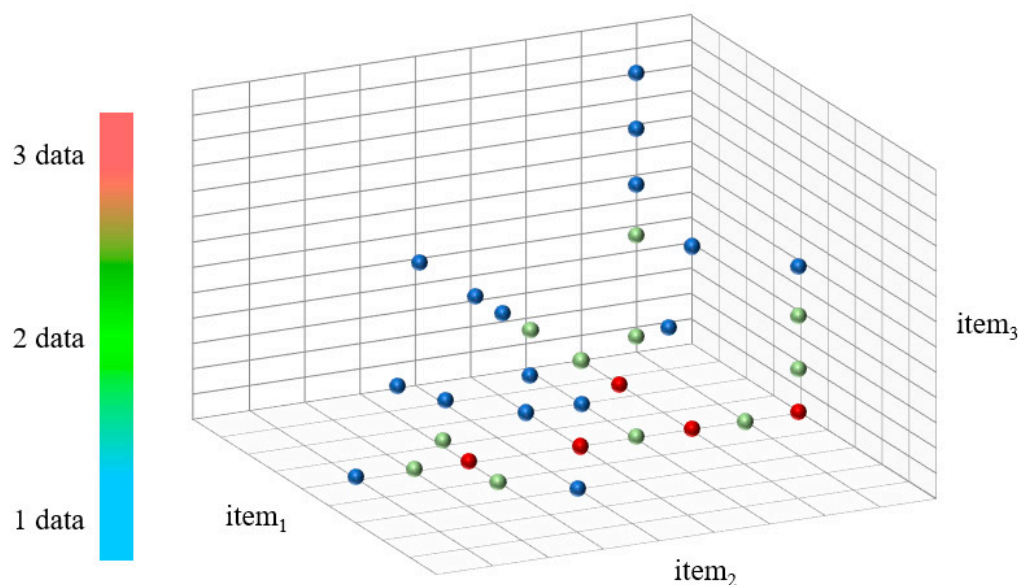


Figure 3. The expression model of the ICAA.

Considering that one continuous data only uses one data dimension for discretization, discretized continuous data should also be considered in one dimension, not in several dimensions like discrete data. Research [26] has introduced fuzzing to extend the Apriori algorithm, which has to some extent optimized the results of rule mining. However, the fuzzy discretization of continuous data can only solve the discretization problem of certain special continuous data.

To solve this problem, firstly, rule mining constraints [27] are added to the Apriori algorithm. By adding constraints, the right half of the generated rule (i.e., the reasoning result) is fixed. Based on the fixed reasoning result of the rule, prior knowledge based on the reasoning result are introduced into the left side (i.e., the data used for reasoning). In this article, considering that the mined rules are used for forest fire risk assessment, indicators reflecting forest fire risk are used as the result of association rules.

Then, we classify continuous data into the following four categories based on their relationship with the reasoning results, and process them separately:

1. Positive correlation

Continuous data, such as “temperature”, increases the likelihood of reasoning results from “temperature” to “fire risk” as numerical values increase. Only introducing fuzzing into Apriori algorithms, or fuzzed continuous data without consider direction, cannot solve this problem. Therefore, this method considers the direction, and expands this datatype upwards to correct the support of such rules.

2. Negative correlation

Similarly, there is a negative correlation between “rainfall” and “fire risk”, which means that previous discretization methods may ignore the important suppression of extreme rainfall on fire risk. Therefore, this method extends this type of data downward to correct the support of such rules.

3. Periodicity related

For example, the impact of “time” on “fire risk” has a clear periodicity. The introduction of fuzzing can effectively solve the discretization problem of this type of data. Therefore, this method adopts fuzzing to discretize and perform subsequent calculations on this type of data. Unlike other categories, this method defines fuzzy membership functions for mining periodic data, and defines:

$$supp\{I_x, I_y, \dots, I_c, \dots\}_q = \frac{N(I_x, I_y, \dots) + \sum \mu(I_p, \dots)}{n} \tag{9}$$

Among them, $\{I_p, \dots\}$ is the set of periodic data participating in the support calculation, and μ is the value of membership function which is defined according to the actual situation.

4. No obvious pattern

For example, for data with unclear patterns such as “wind speed” and “fire risk”, this method discretizes them into segments.

Based on prior knowledge and the proposed method, the continuous data in the dataset are classified as shown in Table 1.

Table 1. Data classification results.

Positive Correlation	Negative Correlation	Periodicity Related	No Obvious Pattern
Temperature	Rainfall, Relative humidity	Time	Wind speed

After introducing prior knowledge for classification, the data are discretized according to their classification results, as shown in Tables 2–5.

Table 2. Discretization of temperature after introducing prior knowledge.

Raw Apriori		Improved Continuous Apriori	
Temperature/°C	Discretized temperature	Temperature/°C	Discretized temperature
≥36	Temp_very_high	ALL	Temp_very_high
≥33 and <36	Temp_high	<36	Temp_high
≥30 and <33	Temp_medium	<33	Temp_medium
≥27 and <30	Temp_low	<30	Temp_low
<27	Temp_very_low	<27	Temp_very_low

Table 3. Discretization of rainfall after introducing prior knowledge.

Raw Apriori		Improved Continuous Apriori	
Rainfall/mm	Discretized rainfall	Rainfall/mm	Discretized rainfall
≥10	Rain_high	≥10	Rain_high
≥1 and <10	Rain_medium	≥1	Rain_medium
>0 and <1	Rain_low	>0	Rain_low
0	Rain_none	ALL	Rain_none

Table 4. Discretization of relative humidity after introducing prior knowledge.

Raw Apriori		Improved Continuous Apriori	
Relative humidity/%	Discretized relative humidity	Relative humidity/%	Discretized relative humidity
≥80	RH_very_high	≥80	RH_very_high
≥70 and <80	RH_high	≥70	RH_high
≥60 and <70	RH_medium	≥60	RH_medium
≥50 and <60	RH_low	≥50	RH_low
<50	RH_very_low	ALL	RH_very_low

Table 5. Discretization of time and wind speed.

Time		Wind Speed	
Time/day	Discretized time	Wind Speed/Km·h ⁻¹	Discretized wind Speed
1–10 June	Time_Early_June	≥19	Ws_very_high
11–20 June	Time_Mid_June	≥16 and <19	Ws_high
21–30 June	Time_Late_June	≥13 and <16	Ws_medium
1–10 July	Time_Early_July	≥10 and <13	Ws_low
.....	<10	Ws_very_low

According to Formula 9, this article sets the membership function of time data as shown in Figure 4.

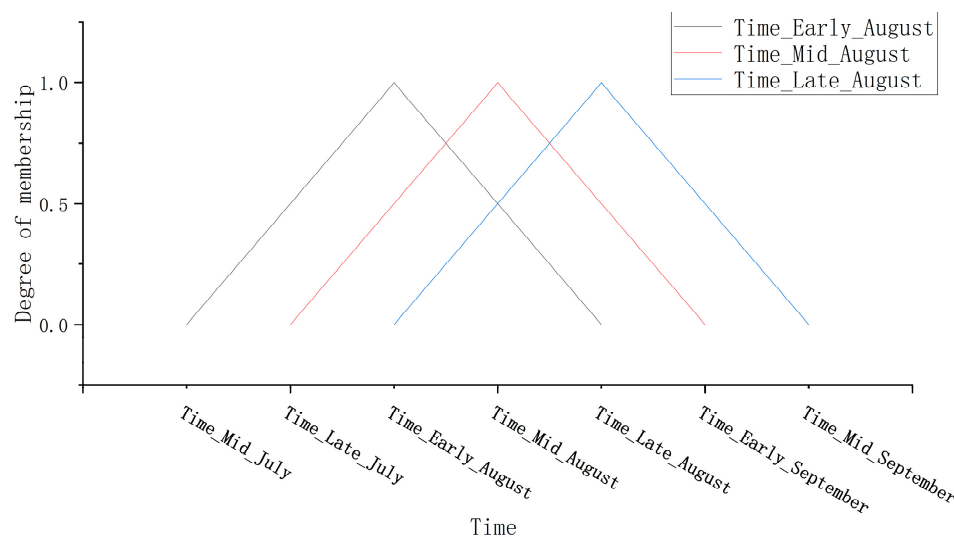


Figure 4. Membership function of time.

2.3. Ontology and Rule Reasoning

Ontology Web Language (OWL) is an ontology expression language recommended by The World Wide Web Consortium (W3C) [28], and Semantic web rule language (SWRL) is a rule expression language built on the OWL language [29]. By combining SWRL and OWL, rule-based automated reasoning has been achieved to support forest fire risk assessment.

Based on the disaster ontology in article [30], and referring to references [31–33], the ontology structure used in this article has been determined. Considering the powerful expressive power of ontology technology, the ontology constructed in this article retains the possibility of expanding the natural disaster management process beyond forest fires and risk assessment. The ontology constructed in this article based on the previous work is divided into five parts:

1. Geographical ontology, which is mainly used to express geographical entities and relationships between entities.
2. Sensor ontology, which is used to express sensor entity and sensor data. As the basis for risk assessment, the information is semantically unified in the subclass sensor data under the sensor ontology.
3. Disaster observation ontology, which is the superclass of forest fire risk assessment. It is also built to ensure the scalability of the ontology.
4. Emergency plan ontology, which is connected to the geographic entity and commander. It is used for high-risk assessment results response processes.
5. People ontology, which is an important entity involved in natural disaster management. People participate in the entire assessment and emergency response process. Meanwhile, people are introduced into the ontology to improve the stability and robustness of the system.

The structure of the ontology is shown in Figure 5.

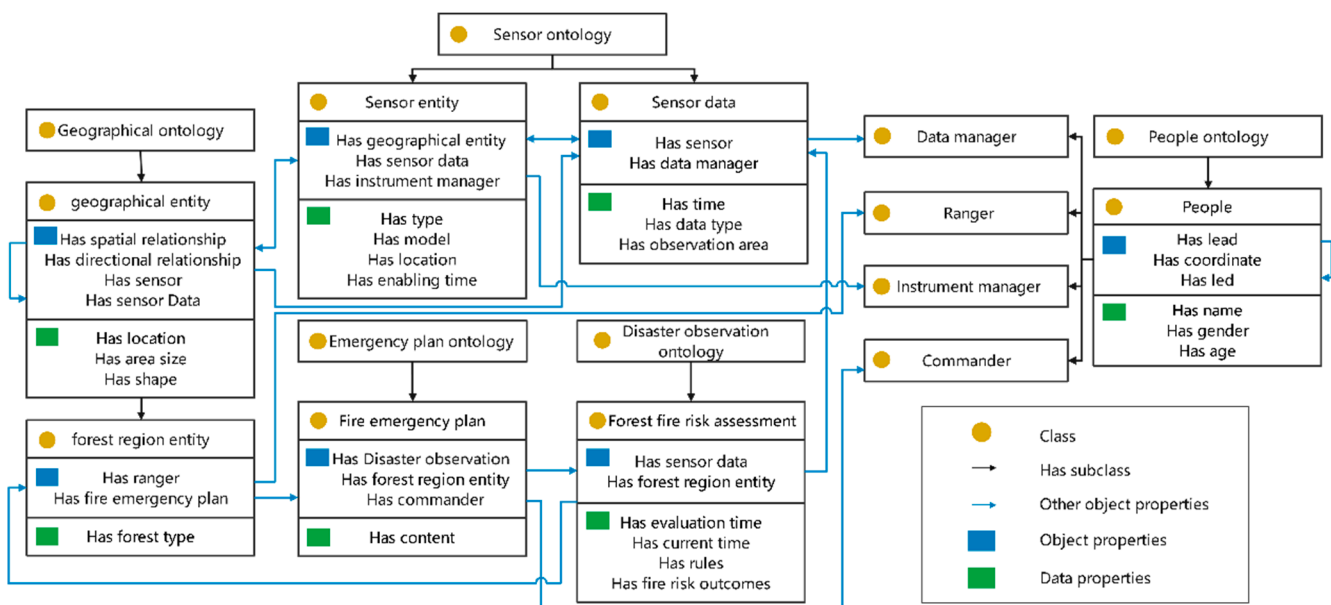


Figure 5. The structure of ontology.

Among them, “Class” is used to represent abstract concepts, which includes “Has subclass”, “Object properties”, “Data properties”, and “Individual”; “Object properties” describe the relationship between classes, with names and directions, but without specific numerical values, and “Has subclass” is also a special type of “Object properties”; and “Data properties” describe the characteristics of a class itself, which have no direction but specific numerical values. Because a class is an abstract concept, it can also be a superclass of an “Individual”, which is not shown in Figure 5. It represents a specific thing under the class, such as a specific piece of land under a geographical entity (rather than an abstract conceptual land). The “Individual” of the ontology constructed in this article are defined in Section 3, with specific examples, and are used in the real reasoning process.

By utilizing the concepts expressed by the ontology, we can achieve automated reasoning with SWRL rules. The basic syntax of SWRL rules is as follows:

$$a_1 \hat{a}_2 \cdots - > b_1 \hat{b}_2 \cdots \quad (10)$$

where “*a*” represents the input, “*b*” represents the output, “ $\hat{}$ ” stands for “and”, and “ $->$ ” is used to separate the input from the output. When conducting rule-based reasoning based on ontology, both “*a*” and “*b*” represent individuals, object properties, or data properties defined in the ontology. In this article, these rules are obtained through knowledge mining using the ICAA.

3. Results

3.1. Results of the Rule Mining Based on Raw Apriori Algorithm and ICAA

To process data using the raw Apriori algorithm, it is necessary to set the support threshold and confidence threshold. The setting of thresholds usually considers factors such as data domain, dataset scale, and data correlation degree. In this article, referring to expert knowledge and experience, the support threshold is set to 0.1 and the confidence threshold is set to 0.7. The algorithm outputs twelve rules, and the first six are shown in Table 6.

Table 6. Output results of the raw Apriori algorithm (partial).

No.	Rule	Confidence	Support
1	RH_low -> Risk_high	0.7083	0.1393
2	Rain_none -> Risk_high	0.8333	0.4508
3	Time_Late_August -> Risk_high	0.7778	0.1148
4	Rain_none ^ Time_Late_August -> Risk_high	1.0000	0.1148
5	Rain_none ^ RH_medium -> Risk_high	0.9583	0.1885
6	Temp_medium ^ Rain_none -> Risk_high	0.9091	0.1639

The rule “Rain_none ^ Time_Late_August -> Risk_high” means that in late August without rainfall, the fire risk is high, with a confidence level of 1.0000 and a support level of 0.1148.

Using the same threshold, with a support of 0.1 and a confidence of 0.7, the ICAA proposed in this article was used to mine rules in the dataset. The algorithm output a total of 35 rules, and the first 16 are shown in Table 7.

Table 7. Output results of ICAA (partial).

No.	Rule	Confidence	Support
1	RH_low -> Risk_high	0.8485	0.459
2	Rain_none -> Risk_high	0.8429	0.4836
3	Time_Late_August -> Risk_high	0.7720	0.1221
4	Rain_none ^ Time_Late_August -> Risk_high	1.0000	0.1221
5	Rain_none ^ RH_medium -> Risk_high	0.9070	0.3197
6	Rain_none ^ RH_low -> Risk_high	0.9492	0.459
7	RH_very_low ^ Rain_none -> Risk_high	0.9833	0.4836
8	Temp_medium ^ Rain_none -> Risk_high	0.8125	0.2131
9	Temp_high ^ Rain_none -> Risk_high	0.9259	0.4098
10	Rain_none ^ Temp_very_high -> Risk_high	0.9833	0.4836
11	RH_very_low -> Risk_high	0.9516	0.4836
12	Temp_very_high -> Risk_high	0.9833	0.4836
13	RH_medium ^ Temp_very_high -> Risk_high	0.9750	0.3197
14	Rain_none ^ Ws_medium -> Risk_high	0.8529	0.2377
15	Temp_high ^ RH_low -> Risk_high	0.9412	0.3934
16	RH_high ^ Temp_very_high -> Risk_high	0.9286	0.1066

3.2. Results of the Ontology based Rule Reasoning

After obtaining the mining rules, the rules are combined with the designed ontology for rule reasoning to verify the effective support of rule mining algorithms for disaster management.

Firstly, based on the ontology structure designed in the previous section, construct the ontology using Protégé [34], which is an ontology management software developed by the Stanford University research team, and the reasoning based on the rules mined in Section 3.1. Create individuals of Sensor_data based on the data and set data properties for each individual, such as time, rainfall, temperature, humidity, etc.

Express the rules mined in the previous section in the form of SWRL syntax, as follows: The output form of mining rules:

$$\text{Rain_none} \wedge \text{Time_Late_August} \rightarrow \text{Risk_high} \tag{11}$$

SWRL syntax:

$$\text{Sensor_data}(?x) \wedge \text{Rain}(?x, ?a) \wedge \text{Time}(?x, ?b) \wedge \text{swrlb:stringEqualIgnoreCase}(?a, \text{"Rain_none"}) \wedge \text{swrlb:stringEqualIgnoreCase}(?b, \text{"Time_Late_August"}) \rightarrow \text{Risk}(?x, \text{"Risk_high"}) \tag{12}$$

Among them, "Sensor_data" represents an individual under the "Sensor_data" class, "Rain", "Time", and "Risk" are data properties defined in the ontology, and "swrlb:stringEqualIgnoreCase" is a built-in rule function in SWRL used for string-type comparisons.

The reasoning process is shown in Figure 6.

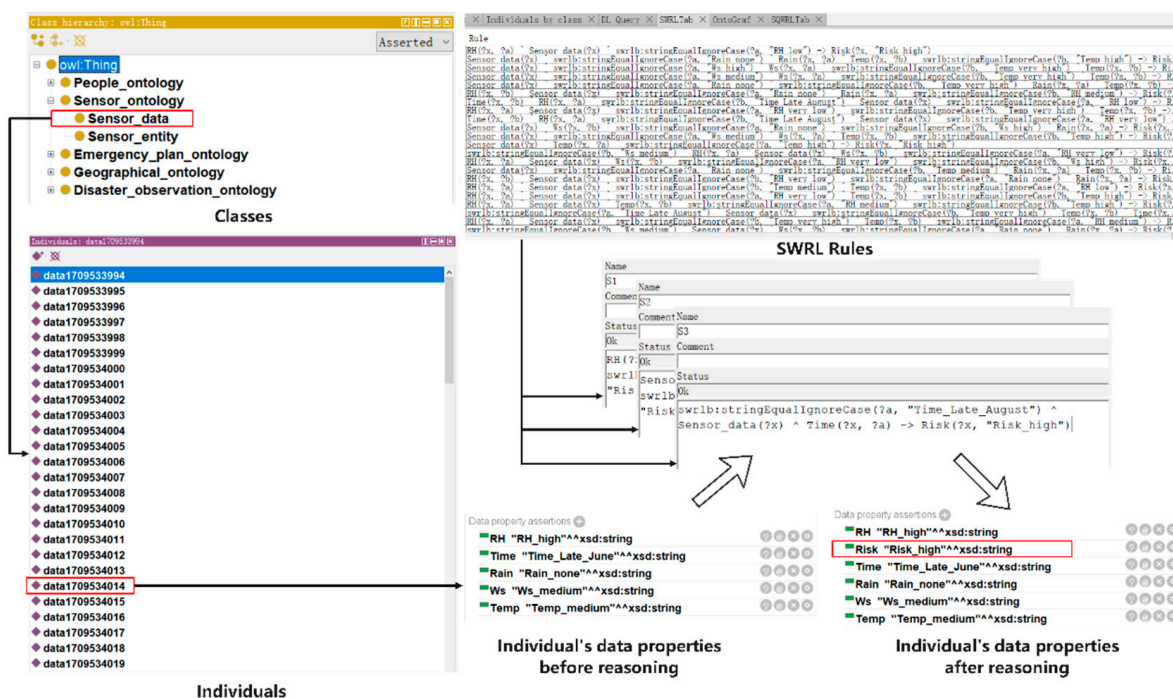


Figure 6. Rule reasoning process.

The rules output by the ICAA are entered into the rule library in the SWRL syntax, and the data collected by sensors are managed as individuals by the ontology. By combining the first two, reasoning based on SWRL rules can be achieved. The partial results of reasoning are shown in Table 8.

Table 8. Result of the rule reasoning (partial).

Data	Rule Reasoning Input					Result
	Time	Temp	RH	Ws	Rain	Risk
data1709534004	Time_Mid_June	Temp_medium	RH_medium	Ws_medium	Rain_none	Risk_high
data1709534005	Time_Mid_June	Temp_very_low	RH_very_high	Ws_very_high	Rain_none	Risk_high
data1709534006	Time_Mid_June	Temp_low	RH_very_high	Ws_very_high	Rain_medium	\
data1709534007	Time_Mid_June	Temp_medium	RH_high	Ws_very_high	Rain_low	\
data1709534008	Time_Mid_June	Temp_low	RH_very_high	Ws_high	Rain_high	\

The reasoning results are reflected in the form of data properties on the individual, as shown in the Table above.

4. Discussion

In this paper, we propose the ICAA, an Improved Apriori algorithm to enhancing forest fire risk assessment. Through rule mining and ontology-based reasoning of forest fire instance data cases in the Bejaia region of Algeria, the advantages and usability of the entire methodology proposed in this article has been fully demonstrated.

The ontology constructed in this article can provide a scalable and standardized expression platform for forest fire risk assessment. The ICAA rules can be combined with SWRL rules to achieve automated reasoning based on the constructed ontology, and the results are written back to the ontology. This improves the automation level of knowledge in forest fire risk assessment, reduces the knowledge requirements for users, and enables semantic knowledge and observation data to better support forest fire risk assessment work.

Compared with the raw Apriori algorithm, the ICAA can better handle continuous data association rule mining for forest fire risk assessment. The ICAA has the following advantages:

1. The mining results of the ICAA include all the rules mined with the raw Apriori algorithm, which proves that the ICAA is an incremental extension of the raw Apriori algorithm.
2. Due to the increased support for small probability events such as "Temp_very_high", the number of candidate rules that meet the support increased, resulting in a 191.67% increase in the number of generated rules.
3. The raw Apriori algorithm discovered "Temp_medium ^ Rain_none -> Risk_high", but did not find "Temp_high ^ Rain_none -> Risk_high" and "Rain_none ^ Temp_very_high -> Risk_high". These three rules were all discovered in the ICAA, and their support increased sequentially, which is also in line with the expectation of prior knowledge. Similarly, there are also the combinations of "Temp_high" and "RH", and "Rain_none" and "RH". Compared with the raw Apriori algorithm, the ICAA has improved the number of mining rules and support with confidence and support, as shown in Figure 7.

The subfigure (a) shows the combination of "Rain_none" and "Temp". In the absence of rainfall, the raw Apriori algorithm only outputs two rules: "Temp_medium", and "Temp_high". Although the confidence level of "Temp_high" is higher than that of "Temp_medium", its support is lower. This is because "Temp_high" appears fewer times, and as for "Temp_very_high", it appears too few times to be found with the raw Apriori algorithm. This situation has been improved in the ICAA. Due to the extension of temperature in the ICAA, a large amount of low-risk data have reduced the support of "Temp_medium" in the low-temperature region (by increasing the denominator of support calculation), while in the high-temperature region, the support and confidence of "Temp_high" and "Temp_very_high" have been increased. The subfigures (b) and (c) also reflect the same situation, and these data represent the possible fire risk under extreme scenarios. Considering their extremely high importance, they cannot be discarded as redundant rules. The above three examples indicate that when processing continuous data

association rule mining for forest fire risk assessment, the raw Apriori algorithm’s neglect of extreme scenario rules is widespread, and ICAA alleviates this neglect to a certain extent.

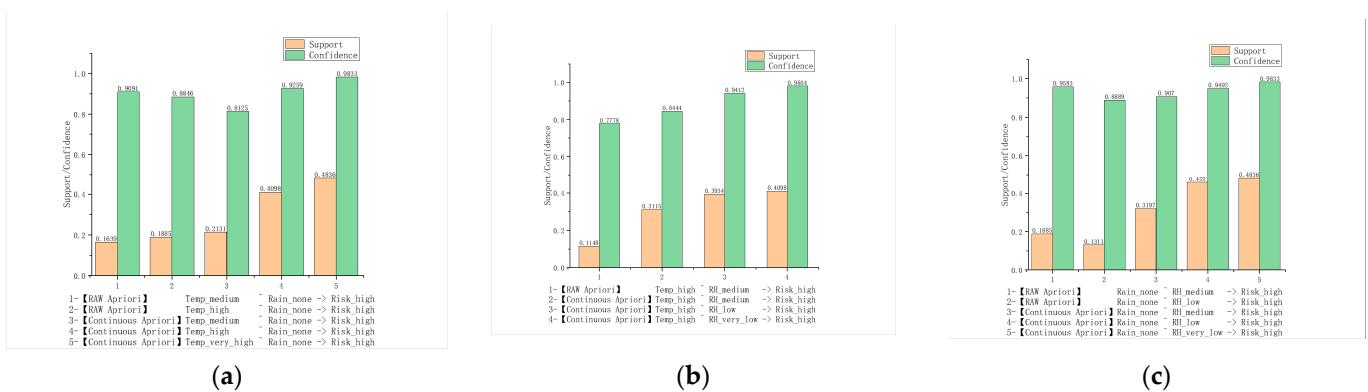


Figure 7. Comparison of raw Apriori algorithm results with ICAA results: “Rain_none and Temp” (a), “Temp_high and RH” (b), and “Rain_none and RH” (c).

The comparison of ICAA, raw Apriori, and three other Apriori algorithm improvements in Table 9 shows that the neglect to rules in extreme scenarios also exists to varying degrees in other improved algorithms. The ICAA has made great progress in increasing attention to rules under extreme scenarios. This proves the superiority of the ICAA, considering that the application area of the algorithm is forest fire rule mining, and the rules under extreme scenarios are very important.

Table 9. Comparison with other Apriori algorithm improvements.

Method	Application Area	Improvement Ideas	Advantage	Attention to Rules in Extreme Scenarios
raw Apriori	Association rules mining for discrete data.	-	Can handle association rule mining of discrete data well.	When continuous data is input, rules under extreme scenarios are ignored.
ICAA	Association rules mining for forest fire data.	Introducing prior knowledge to classify continuous data for data directional discretization.	Discretization with direction strengthens the emphasis on extreme scenarios, and avoids ignoring important but low-frequency rules.	By introducing prior knowledge, directional discretization methods effectively increase the importance of rules in extreme scenarios.
[19]	Association rules mining for urban water supply.	Cluster continuous data using K-means to determine the threshold for data discretization.	Compared to the raw Apriori algorithm, it can support continuous data input.	Considering that extreme scenarios are often not recognized as a separate cluster in clustering, rules may be ignored.
[20]	Association rules mining for power equipment fault diagnosis.	Using distribution probability of the equipment condition for data discretization.	The division reflects statistical significance and can effectively reflect the status of equipment.	Probability based methods can effectively distinguish data under normal operating conditions, but for extreme low-frequency scenarios, rules may be ignored.
[21]	Association rules mining for big data.	Introducing fuzzy technology and constructing symmetric membership functions for fuzzy mining of continuous data.	Able to mine fuzzy association rules under the conditions of big data.	The enhancement of symmetric fuzziness for data in extreme scenarios is limited. Meanwhile, to combat the exponential increase in search space caused by fuzzy sets, this method restricts the search depth, which may further lead to ignored rules.

5. Conclusions

This study provides an improved association rule mining algorithm, ICAA, which enhances the adaptability of the Apriori algorithm to continuous data by introducing prior knowledge to classify the input data. In addition, the rules obtained from mining are combined with ontology to implement semantic reasoning, reducing the knowledge requirements for users in forest fire risk assessment.

Based on ICAA and the designed ontology, an ontology and data individuals were constructed using the ontology management software protege on a forest fire dataset in the Bejaia region of northeastern Algeria. The proposed ICAA was used to mine the rules of the dataset, and the mined rules were combined with the constructed ontology in the form of SWRL specifications, achieving accurate automated reasoning.

The results show that the ICAA outperforms the raw Apriori algorithm, the number of generated rules increased by 191.67%, increasing the attention to extreme scenario rules for continuous data in association rules mining. The output rules can be integrated with the constructed ontology to achieve automated semantic reasoning to support forest fire risk assessment.

The work presented can be used to enhance the forest fire risk assessment and contribute to the generation and sharing of forest-fire-related knowledge, alleviate the problem of insufficient knowledge in forest fire risk assessment, and be extended to forest fire risk management in other regions or other forest fire management services such as forest fire spread management and forest fire point identification.

Author Contributions: Conceptualization, Y.D.; Data curation, C.X.; Investigation, Y.D.; Methodology, Y.D.; Software, Y.D.; Supervision, Z.L.; Validation, C.X.; Writing—original draft, Y.D.; Writing—review and editing, Z.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Key Research and Development Program of China, grant number 2021YFC3000302.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ju, W.Y.; Xing, Z.X.; Wu, J.; Kang, Q.C. Evaluation of forest fire risk based on multicriteria decision analysis techniques for Changzhou, China. *Int. J. Disaster Risk Reduct.* **2023**, *98*, 104082. [[CrossRef](#)]
2. Naderpour, M.; Rizeei, H.M.; Ramezani, F. Forest Fire Risk Prediction: A Spatial Deep Neural Network-Based Framework. *Remote Sens.* **2021**, *13*, 2513. [[CrossRef](#)]
3. Wen, H.R.; Guo, Q.Z.; Zeng, Y.H.; Wu, Z.P.; Sun, Z.H. Study on forest fire risk in Conghua district of Guangzhou city based on multi-source data. *Nat. Hazards* **2022**, *114*, 3163–3183. [[CrossRef](#)]
4. Brys, C.; Navas-Delgado, I.; Aldana-Montes, J.F. Wildfire risk weighting and behaviour prediction using open geospatial data and ontologies. *J. Inf. Sci.* **2023**. [[CrossRef](#)]
5. Rubi, J.N.S.; de Carvalho, P.H.P.; Gondim, P.R.L. Forestry 4.0 and Industry 4.0: Use case on wildfire behavior predictions. *Comput. Electr. Eng.* **2022**, *102*, 108200. [[CrossRef](#)]
6. Gruber, T.R.; Olsen, G.R.; Runkel, J. The configuration design ontologies and the VT elevator domain theory. *Int. J. Hum. Comput. Stud.* **1996**, *44*, 569–598. [[CrossRef](#)]
7. Chandra, R.; Agarwal, S.; Singh, N. Semantic sensor network ontology based decision support system for forest fire management. *Ecol. Inform.* **2022**, *72*, 101821. [[CrossRef](#)]
8. Ge, X.T.; Yang, Y.; Peng, L.; Chen, L.J.; Li, W.C.; Zhang, W.Y.; Chen, J.H. Spatio-Temporal Knowledge Graph Based Forest Fire Prediction with Multi Source Heterogeneous Data. *Remote Sens.* **2022**, *14*, 3496. [[CrossRef](#)]
9. Masa, P.; Kintzios, S.; Vasileiou, Z.; Meditskos, G.; Vrochidis, S.; Kompatsiaris, I. A Semantic Framework for Decision Making in Forest Fire Emergencies. *Appl. Sci.* **2023**, *13*, 9065. [[CrossRef](#)]
10. Liu, J.C.; Tang, F.L.; Zhu, Y.M.; Yu, J.D.; Chen, L.; Gao, M. INFER: Distilling knowledge from human-generated rules with for STINs. *Inf. Sci.* **2023**, *645*, 119219. [[CrossRef](#)]
11. Zhao, X.F.; Huang, L.L.; Sun, Z.; Fan, X.T.; Zhang, M. Design Optimization of Building Exit Locations Based on Building Information Model and Ontology. *Sustainability* **2023**, *15*, 12922. [[CrossRef](#)]

12. Lv, M.X.; Cao, X.D.; Wu, T.X.; Li, Y.H. A Civil Aviation Customer Service Ontology and Its Applications. *Data Intell.* **2023**, *5*, 1063–1081. [[CrossRef](#)]
13. Alrahan, B.A.; Farouk, M. ASCF: Optimization of the Apriori Algorithm Using Spark-Based Cuckoo Filter Structure. *Int. J. Intell. Syst.* **2024**, *2024*, 8781318. [[CrossRef](#)]
14. Bao, F.G.; Mao, L.H.; Zhu, Y.L.; Xiao, C.C.; Xu, C.H. An Improved Evaluation Methodology for Mining Association Rules. *Axioms* **2022**, *11*, 17. [[CrossRef](#)]
15. Zaki, M.J. Scalable algorithms for association mining. *IEEE Trans. Knowl. Data Eng.* **2000**, *12*, 372–390. [[CrossRef](#)]
16. Sitanggang, I.S.; Rakhmianti, M.; Khotimah, H. Association Patterns of Hotspot Sequence with Socio-economic Aspects in Peatland in Sumatra. In Proceedings of the International Conference on Computer, Control, Informatics and its Applications (IC3INA), Tangerang, Indonesia, 3–5 October 2016; pp. 175–178.
17. Boulehmi, M.; Ayadi, A. Extraction of Spatiotemporal Association Rules for Forest Fires Prediction. In Proceedings of the 6th International Conference of Reliable Information and Communication Technology (IRICT), Electr Network, Virtual, 22–23 December 2021; pp. 209–218.
18. Cleland, Z.W.; Dao, K.A.; Dao, T.H.D. Detecting changes in spatial characteristics of Colorado human-caused wildfires using APRIORI-based frequent itemset mining. *Comput. Environ. Urban Syst.* **2023**, *101*, 101941. [[CrossRef](#)]
19. Liu, X.; Sang, X.F.; Chang, J.X.; Zheng, Y.; Han, Y.P. The water supply association analysis method in Shenzhen based on kmeans clustering discretization and apriori algorithm. *PLoS ONE* **2021**, *16*, e0255684. [[CrossRef](#)] [[PubMed](#)]
20. Liu, H.; Wang, Y.Y.; Yang, Y.; Liao, R.J.; Geng, Y.J.; Zhou, L.W. A Failure Probability Calculation Method for Power Equipment Based on Multi-Characteristic Parameters. *Energies* **2017**, *10*, 704. [[CrossRef](#)]
21. Alcalá-Fdez, J.; Alcalá, R.; Herrera, F. A Fuzzy Association Rule-Based Classification Model for High-Dimensional Problems with Genetic Rule Selection and Lateral Tuning. *IEEE Trans. Fuzzy Syst.* **2011**, *19*, 857–872. [[CrossRef](#)]
22. Bentekhici, N.; Bellal, S.A.; Zegrar, A. Contribution of remote sensing and GIS to mapping the fire risk of Mediterranean forest case of the forest massif of Tlemcen (North-West Algeria). *Nat. Hazards* **2020**, *104*, 811–831. [[CrossRef](#)]
23. Aini, A.; Curt, T.; Bekdouche, F. Modelling fire hazard in the southern Mediterranean fire rim (Bejaia region, northern Algeria). *Environ. Monit. Assess.* **2019**, *191*, 747. [[CrossRef](#)] [[PubMed](#)]
24. Meddour-Sahar, O. Wildfires in Algeria: Problems and challenges. *Iforest-Biogeosciences For.* **2015**, *8*, 818–826. [[CrossRef](#)]
25. Abid, F. Algerian Forest Fires Dataset. 2019. Available online: <https://archive.ics.uci.edu/dataset/547/algerian+forest+fires+dataset> (accessed on 9 January 2024).
26. Yan, S.R.; Pirooznia, S.; Heidari, A.; Navimipour, N.J.; Unal, M. Implementation of a Product-Recommender System in an IoT-Based Smart Shopping Using Fuzzy Logic and Apriori Algorithm. *IEEE Trans. Eng. Manag.* **2022**, *71*, 4940–4954. [[CrossRef](#)]
27. Xu, Z.; Huo, H.X.; Pang, S.H. Identification of Environmental Pollutants in Construction Site Monitoring Using Association Rule Mining and Ontology-Based Reasoning. *Buildings* **2022**, *12*, 2111. [[CrossRef](#)]
28. Hawke, S.; Horridge, M.; Parsia, B.; Schneider, M. *OWL 2 Web Ontology Language Conformance*, 2nd ed.; 2012. Available online: <https://www.w3.org/2012/pdf/REC-owl2-conformance-20121211.pdf> (accessed on 20 January 2024).
29. Bassiliades, N. A Tool for Transforming Semantic Web Rule Language to SPARQL Inferencing Notation. *Int. J. Semant. Web Inf. Syst.* **2020**, *16*, 87–115. [[CrossRef](#)]
30. Yumin, D.; Ziyang, L.; Xuesong, L.; Xiaohui, L. Using ontology and rules to retrieve the semantics of disaster remote sensing data. *J. Syst. Eng. Electron.* **2024**, 1–8. [[CrossRef](#)]
31. *GB/T 36743-2018*; Forest Fire Danger Weather Ratings. State Administration for Market Regulation of China: Beijing, China, 2018.
32. Bharambe, U.; Durbha, S.S. Adaptive Pareto-based approach for geo-ontology matching. *Comput. Geosci.* **2018**, *119*, 92–108. [[CrossRef](#)]
33. Compton, M.; Barnaghi, P.; Bermudez, L.; García-Castro, R.; Corcho, O.; Cox, S.; Graybeal, J.; Hauswirth, M.; Henson, C.; Herzog, A.; et al. The SSN ontology of the W3C semantic sensor network incubator group. *J. Web Semant.* **2012**, *17*, 25–32. [[CrossRef](#)]
34. Musen, M.A.; Protege, T. The Protege Project: A Look Back and a Look Forward. *AI Matters* **2015**, *1*, 4–12. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.