Computers and Electronics in Agriculture 190 (2021) 106411

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Multidrug resistance analysis method for pathogens of cow mastitis based on weighted-association rule mining and similarity comparison

Buwen Liang [a], Xinxing Li [a,*], Ziyi Zhang [a], Congming Wu [b], Xin Liu [c], Yongjun Zheng [c,*]

[a] Beijing Advanced Innovation Center for Food Nutrition and Human Health, College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, PR China
[b] Beijing Key Laboratory of Detection Technology for Animal-Derived Food Safety, College of Veterinary Medicine, China Agricultural University, PR China
[c] Beijing Advanced Innovation Center for Food Nutrition and Human Health, College of Engineering, China Agricultural University, Beijing 100083, PR China

ARTICLE INFO

Keywords:
Cow mastitis
Drug resistance
Multidrug resistance
Weighted association rule mining
Similarity analysis

ABSTRACT

Mastitis is one of the most common diseases and causes the greatest economic loss in dairy farming. Antibiotics are the most effective drugs to prevent and treat bacterial infection of mastitis. However, yet the growing problem of drug resistance, especially multidrug resistance (MDR), poses a great threat to disease control. To understand the MDR rules in bacteria of cow mastitis from national level, the main bacteria from cows with mastitis in large-scale farms were isolated and identified in China, and then drug sensitivity tests were conducted to establish a drug resistance data set. Aiming at the problem of numerous and disordered drug resistance data and lack of extensive correlations, a weighted Apriori association rule mining algorithm in conjunction with the bacterial drug resistance prevalence is proposed. We analyzed the associations between different antibiotics of key bacteria, extracted and visualized the key trends of high resistance prevalence and frequent occurrence, and discovered MDR patterns. Finally, a similarity comparison method based on Euclidean measurement was proposed to compare the relative MDR rules of different bacteria from the overall level with support, confidence, and promotion as characteristic parameters. The drug resistance data set showed that *staphylococcus* were the main bacteria isolated from dairy cow mastitis in China. Then based on the association rule algorithm, the important rules between different antibiotics resistance in this dataset were identified. In addition, the MDR patterns of different bacteria were visualized and analyzed by using the chord diagram. The results showed the bacteria are highly resistant to *penicillin*, *gentamicin*, and *ampicillin*, and most other antibiotics were linked with these three antibiotics. Finally, the high correlations and main rules in different bacteria were confirmed by a similarity comparison method. The assessment model and conclusions of this study are potentially valuable for assessing the evolution of MDR patterns, providing a scientific basis for relevant authorities to guide the rational use of antibiotics in the farming industry.

1. Introduction

The high rate and irregular use of antibiotics make the drug resistance of bacteria increasingly serious, which has become one of the most serious problems in the world (Clifford et al., 2018; Davies and Wales, 2019), especially the problem of multidrug resistance (MDR) (Van Boeckel et al., 2019). At present, animal antibiotic consumption accounts for more than half of the total consumption of antibiotics, significantly affecting the prevention and treatment of animal diseases (Van Hecke et al., 2017), and the drug-resistant bacteria produced by farming can spread to the population through the food chain or to the environment through excrement, posing a great threat to human health (Hudson et al., 2017; Zhang et al., 2015). The government has taken strong measures to solve the drug resistance problem (Hvistendahl, 2012; Li et al., 2020), but currently mainly focuses on hospital and meat products such as poultry and pigs, with insufficient attention to dairy products (Xu et al., 2020).

Cow mastitis is one of the most frequent and damaging diseases in dairy cows (Motaung et al., 2017; Rajamanickam et al., 2019), in which pathogenic bacterial infection is the most leading cause. In MDR, pathogenic bacteria are resistant to two or more antibiotics, which is a great threat and seriously affects the effective prevention and treatment of cow mastitis (Cheng and Han, 2020). Understanding the patterns of MDR in specific bacteria can help to guide antibiotic drug management

* Corresponding authors.
E-mail addresses: lxxcau@cau.edu.cn (X. Li), zyj@cau.edu.cn (Y. Zheng).

https://doi.org/10.1016/j.compag.2021.106411
Received 11 February 2021; Received in revised form 16 August 2021; Accepted 17 August 2021

and therapy, and it is also important for the rational use of antibiotics and prevention and control of animal diseases on farms (Ludwig et al., 2013).

The current research on drug resistance and MDR is still inadequate (Sommer et al., 2017). Studies mainly focused on the bioinformatics of genes and structural proteins (Su et al., 2019; Thomas et al., 2015; Zhong et al., 2011) and the mechanism of action of antibiotics (Gao et al., 2019; Yang et al., 2019). Although it is possible to obtain MDR resistance genes and propose new solutions to solve the problem of drug resistance from a microscopic and pharmacology perspective, this approach is time-consuming to implement, and it is difficult to evaluate the overall use of antibiotics to guide dosing on farms.

The main methods for MDR analysis from the macrolevel by data analysis include Markov network, Bayesian network, and association rule mining algorithms (Safdari et al., 2020). The Bayesian network was used to estimate the interaction and mine the MDR pattern of *E. coli* and have identified three-way and four-way interactions between resistances (Ludwig et al., 2013), but it is an oriented graph, which only reflects the probability of bacterial emergence of resistance. Therefore, it is difficult to analyze the resistance association among multiple antibiotics. Furthermore, although the Markov network can analyze the drug resistance data well, it is statistically independent in each matter and not suitable for dealing with large amounts of data.

Association rule mining algorithms can effectively analyze sparse bacterial drug resistance data to obtain the strength of the relationship between different MDR (Agrawal and Srikant, 1994). At present, it has been used to mine significant relationships in public health and medical datasets (Lamma et al., 2003; Ma et al., 2003). Besides, in terms of drug resistance in animal-derived pathogenic bacteria, the MDR patterns of *E. coli* to 15 antibiotics in chickens from 2004 to 2012 based on association rule mining algorithm were analyzed, which is the first analysis of MDR patterns in the *E. coli* of chicken origin by using association rule mining algorithms (Cazer et al., 2019).

However, the prevalence of antibiotic resistance is one of the most important standards for evaluating the risk of resistance (Collignon et al., 2018), while traditional association rule algorithms pay more attention to the frequency of antibiotics, ignoring the differences between the prevalence of drug resistance. Therefore, some invalid rules with high support and low prevalence rate have been more easily discovered, but it is difficult to dig out the significant rules with high prevalence and low support at the same time, resulting in great uncertainty and inaccuracy in the MDR patterns. To solve this problem, some studies discarded the threshold support and confidence and used the fuzzy algorithm to discover more rules (Bansal et al., 2017). However, the results are miscellaneous and relatively poor, and it is difficult to accurately locate the required association rules.

In order to dig out effective and scientific rules, we believe that a reasonable weighted threshold value must be established for rule identification. Therefore, in this study, the drug resistance prevalence of bacteria was considered as the weight values, then a weighted Apriori association rule algorithm was proposed to determine the multidirectional relationship between drug resistance of different antibiotics caused and explore the MDR patterns of bacteria in cow mastitis.

Finally, to compare the relative MDR rules and confirm the high correlation between of different bacteria, a method is proposed for comparing MDR based on the Euclidean metric similarity algorithm using *Staphylococcus aureus* (*S. aureus)* and *coagulase negative staphylococci* (*CNS*) as the study subjects from the overall level with support, confidence, and lift as characteristic parameters. This study aims to guide the rational use of antibiotics in the farming industry and provide a scientific basis for relevant departments to deal with drug resistance.

2. Materials and methods

2.1. Data collection

Pathogenic bacteria were isolated and identified from the samples of cows with mastitis in large-scale farms, and then drug sensitivity tests were conducted. A total of 13,425 pieces of drug resistance data was collected. The data content includes the ID, site, time, quantity, minimum inhibitory concentration (MIC), and drug resistance of the bacteria, providing a complete data basis for the MDR analysis from the perspective of data-driven.

After statistical analysis, the bacteria of mastitis in dairy cows were mainly *Staphylococcus*, accounting for 52.95%, with 28.03% of *S. aureus*, 24.92% of *CNS*. The results indicated that *Staphylococcus* may be the main isolated bacteria in cow mastitis, among which *S. aureus* can directly cause disease, while *CNS* is an environmental conditional pathogen (Nagasawa et al., 2019). Therefore, *S. aureus* and *CNS* were taken as examples to analyze their MDR patterns.

In this study, 11 drugs commonly used in the treatment of cow mastitis were selected and tested for drug sensitivity using the MIC method (Michael et al., 2020). In view of the lack of a relatively complete standard in China at present, VET01-S2, S22, and M45-A2 standards recommended by Clinical and Laboratory Standards Institute in the USA are used. The drug resistance prevalence is shown in Table 1.

2.2. Methods

Association rule mining is an unsupervised machine learning algorithm for analyzing the relationships in large datasets, which can effectively identify and discover the most strongly associated patterns in a dataset (Ceglar and Roddick, 2006; Kumbhare and Chobe, 2014). It is well suited for analyzing MDR to antibiotics because it can analyze the strength of relationships and determine the rules between the resistance of each bacterium, especially the drug resistance data of the structure encoded by one-hot.

2.2.1. Related definitions of association rule mining

Definition 1. The drug resistance data are translated to binary system, which is arranged as transactions and items with one transaction in each row and one item in each column.

Definition 2. The bacteria isolated from a cow are considered as a transaction. Database $D = \{T_1, T_2, T_3, ..., T_m\}$ is composed of a series of transactions with a unique ID.

Definition 3. The drug resistance is considered as an item, in which 1 represents drug resistance to the drug, and 0 represents no resistance to the drug. Suppose $I = \{i_1, i_2, i_3, ..., i_k\}$ is the set of all items. If $X \subseteq I$, $Y \subseteq I$, then X and Y are called as the item sets, and if the count of items is k, then they are called as k-itemset.

Definition 4: An association rule can be defined as a logical implication of $X -> Y$, where $X \subseteq I$, $Y \subseteq I$, and $X \bigcap Y = \varnothing$.

Definition 5: Support. Support of the association rule X→Y is the proportion of transactions contains X ∪ Y in the database D.

$$Support(X \rightarrow Y) = P(X \cup Y) \tag{1}$$

Definition 5: Confidence. Confidence of association rule X → Y is the proportion of transactions containing X ∪ Y to transactions containing X in the database D.

$$Confidence(X \rightarrow Y) = \frac{support(X \cup Y)}{suppport(X)} \tag{2}$$

Definition 6: Lift. The lift represents the proportion of probability of containing Y conditional X, to the probability of containing only Y. It reflects the correlation between X and Y in the association rule. A lift greater than 1 shows a higher positive correlation, and a lift < 1 and lower shows negative correlation.

Table 1
Drug resistance prevalence of 11 antibiotics.

Antibiotics	abbreviations	Class	resistance prevalence of *S. aureus*	resistance prevalence of *CNS*
cefalexin	CEF	β-Lactam	0.1016	0.12
trimethoprim/sulfamethoxazole	TRI	sulfonamides	0.1250	0.1878
amoxicillin/Clavulanic acid	AMO	β-Lactam	0.2109	0.0954
lincomycin	LIN	lincosamide	0.2500	0.1692
enrofloxacin/ciprofloxacin	ENR	fluoroquinolones	0.2812	0.1477
sulfasomizole	SUL	sulfonamides	0.3281	0.2554
erythromycin	ERY	macrolides	0.3828	0.2215
gentamicin	GEN	aminoglycosides	0.4219	0.1262
ampicillin	AMP	β-Lactam	0.4140	0.2246
rifaximin	RIF	ansamycin	0.5703	0.7538
penicillin	PEN	β-Lactam	0.7031	0.6308

$$Lift(\mathrm{X}\rightarrow\mathrm{Y}) = \frac{Confidence(\mathrm{X}\rightarrow\mathrm{Y})}{suppport(Y)} \tag{3}$$

2.2.2. Apriori method

Apriori association rule mining is one of the most influential algorithms for mining frequent itemsets of Boolean association rules (Agrawal and Srikant, 1994). The core idea of the algorithm is to mine the *k*-order frequent itemsets by conducting an iterative search of the target transaction database until the highest-order frequent itemset is found, and finally, the frequent itemsets are mined to obtain association rules to achieve the final goal of mining the association relationship between the target data set.

The process of mining association rules can be converted into the process of finding strong association rules that satisfy the minimum support threshold and minimum confidence threshold: 1) Generating all frequent itemsets, i.e., finding itemsets with support higher than or equal to the minimum support threshold; 2) Generating strong association rules, i.e., finding association rules with frequent itemsets higher than or equal to the minimum confidence threshold (Hipp et al., 2000; Zhang and Zhang, 2002). However, this algorithm reflects the dependence and association between transactions through support, easily generating a large number of unnecessary candidate items. Furthermore, the drug resistance prevalence, the most important index to reflect the risk level of bacterial drug resistance, is not reflected in the Apriori association rule mining algorithm.

2.2.3. Weighted Apriori algorithm

Compared with other datasets, the MDR dataset has less data volume, but more deep associations between drugs. These bacteria are mainly 2 or 3-way resistance bacteria, or even super resistance bacteria with more than 3 drugs. Each MDR rule has a low support but high confidence and prevalence level. Thereby, the weights of rules for antibiotics with very high resistance and low support should be increased, while the weights of rules for antibiotics with low resistance and high support should be decreased, so that more stable, reliable, and realistic resistance rules can be obtained.

However, the Apriori algorithm mainly considers support and neglect other metrics. To dig out more useful rules, the most direct and effective way is to introduce weighted metrics to evaluate the level of support, increase the weight of key items, improving the importance of certain transactions (Sun and Bai, 2008; Zhou et al., 2017). Therefore, the Apriori association rule algorithm was improved by combining the actual situation of bacterial resistance to different antibiotics as weighted values. And the following two hypotheses are proposed: (1) A high-quality transaction should contain many high-quality items. (2) A high-quality item should be present in many high-quality transactions.

These hypotheses apply to drug resistance means a high-quality MDR rule should include drugs with higher resistance prevalence, and drugs with higher resistance prevalence should be included in many MDR rules. Based on the above hypotheses, the main strategy of weighted Apriori is to obtain all frequent itemsets in the target transaction dataset according to the predefined minimum support and drug resistance weights and then quickly obtain association rules according to the weighted frequent itemsets. The detailed mining steps are as follows and the Weighted Apriori algorithm is shown in Table 2.

S1: The resistance prevalence of each drug is standardized, and a resistance weight with a mean of 1 is established and expressed as values. In Equation (4), x is the resistance prevalence of original antibiotic; $\overline{x}$ is the mean value; and σ is the variance.

$$\text{values} = \frac{x-\overline{x}}{\sigma} \tag{4}$$

S2: Traverse dateset D and build the dictionary {c: values}; c is the item (drug resistance); values mean the weight value (drug resistance prevalence); each item has a corresponding weight value and finally generates the candidate item C_1.

S3: Filter the weighted frequent itemset L_k according to the value of threshold w-support.

$$w-support(X) = \frac{sum(x)}{N} * values \tag{5}$$

S4: L_{k-1} ($k \geq 2$) is used to generate the itemset of *k*-order candidate items C_k using a self join. C_k is a superset of L_k, and after compressing C_k by subset testing, the dataset can be scanned to determine the count of each candidate in C_k and thus L_k.

S5: According to any subset of a frequent itemset is a frequent itemset, the candidate item C_k of order k can be pruned. Suppose C_{k-1} is any (k-1)-order subset of C_k. If $C_{k-1} \notin L_{k-1}$, then $C_k \notin L_k$, and then the candidate item must not be frequent.

S6: Repeat steps S2 and S3 until no higher-order frequent itemsets can be mined. Then the association rules satisfying the requirements are calculated in the resulting all-frequent itemsets, and the mining process ends.

Table 2
Weighted Apriori algorithm.

Algorithm: Weighted Apriori

Input: C_k: Candidate itemset of size k
L_k: Frequent itemset of size k
$w-support_i$: drug resistance values of this item
Steps:
1. $L_1 \rightarrow \{Frequentitems | w-support\}$
2. $for(k=1; L_k \neq \varnothing; k++)$
3. $if \frac{sum(x)}{N} * w-support_i > threshold$
4. $C_{k+1} \leftarrow candidate(L_k)$
5. for each transaction t
6. $Q \leftarrow \{c | c \in C_{k+1} \bigwedge c \subseteq t\}$
7. $count[c] \leftarrow count[c]+1, \forall c \in Q$
8. end for
9. $L_{k+1} \leftarrow \{c | c \in C_{k+1} \bigwedge count[c]/N \geq \sigma$
10. end for
11. return $\bigcup_k L_{k+1}$

2.2.4. Similarity comparison

Correlation analysis mostly focuses on two or more variables to measure the degree of correlation between these variables. However, there are significant limitations, making it hard to compare a single individual in detail. The Euclidean metric similarity comparison algorithm obtains the similarity by determining the distance between vectors: The closer the distance, the more similar it is, reflecting the absolute difference in individual characteristics. The association rules for MDR belong to individuals, and each association rule represents an MDR rule. Therefore, for the main bacteria (*S. aureus*, *CNS*), based on the method, we need to compare the similarities and differences for each rule and assess the MDR between these bacteria.

The Euclidean metric defines the differences between two points in an *m*-dimensional space. Combined with MDR rules, a multidimensional space was created with the support, confidence, and lift of two bacterial resistance rules as the axes. Then, the degree of similarity can be obtained using the formula (6).

$$d = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{6}$$

Due to the distance between two points is always non-negative, and the maximum value is positive infinity. To set the value range of similarity between [0,1], Formula (7) was used to normalize it. Euclid's formula for calculating the similarity between two items can be expressed as follows:

$$w_{AB} = \frac{1}{1+d} \tag{7}$$

3. Results

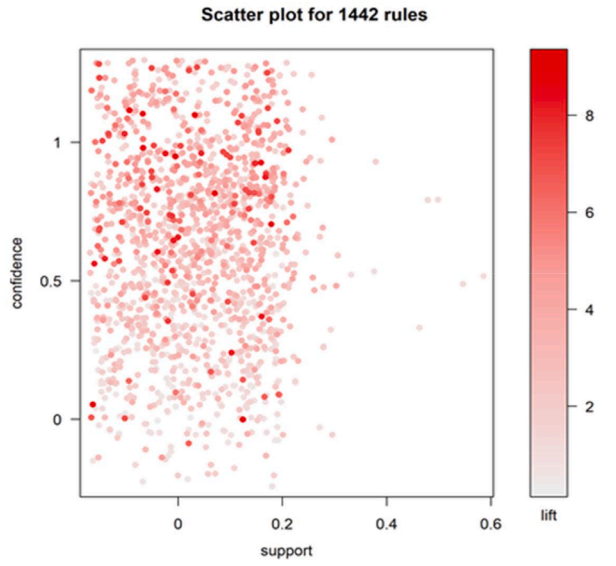
3.1. Association rule results

3.1.1. Overall results and visualization

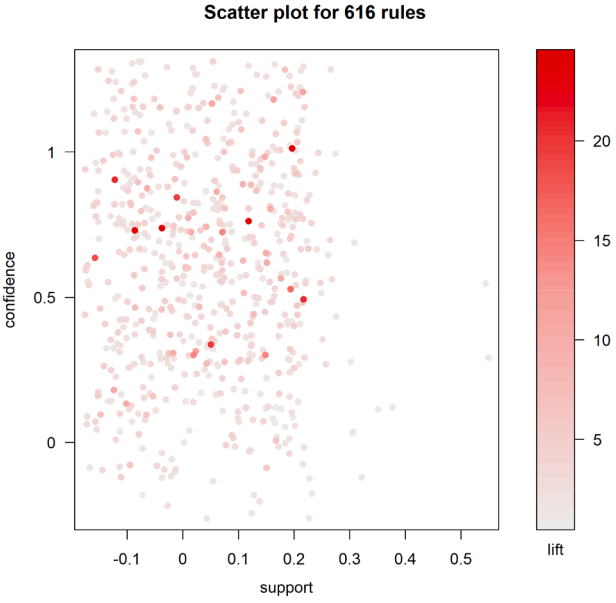
In this section, *S. aureus*, *CNS* are used as examples for their MDR analysis. All possible interdependencies are considered for weighted frequent itemset mining and generating association rules based on the proposed algorithm and MDR dataset.

The generation rules are shown in Fig. 1. A total of 1442 rules were generated for *S. aureus*, 616 rules for *CNS*, indicating that MDR in *S. aureus* is more serious and the drugs are mainly used for *S. aureus*, less for *CNS*. This also shows that the drug resistance of *CNS* to these 11 drugs is not as high as that of *S. aureus*. The generated rules were drawn as s scatter diagram, with X axis for Support, Y axis for Confidence, and color for Lift.

There are still some meaningless rules in the set of rules generated by the minimum support and confidence threshold. To display their main rules more visually and to obtain a clear understanding of the MDR pattern, a visualization method was proposed based on the chord diagram. It is a visualization method that shows the relationships between the data of the nodes are arranged along the circumference and radial direction of a circle, and the nodes are linked by weighted arcs (Acan, 2017). The association rules of MDR were decomposed into antibiotic nodes and connections. For example, the rule (*ampicillin*, *enrofloxacin*=> *cefalexin*) was decomposed into nodes (*ampicillin*, *enrofloxacin*, *cefalexin*) and directed connections (*ampicillin*=> *cefalexin*, *enrofloxacin*=> *cefalexin*), while redundant connections were removed. Each connection was weighted according to the number of occurrences in the optimal rule set. To visualize the association rule mining results, the nodes were distributed around the circumference of a circle, and the nodes were connected by arcs to show the interrelationship. The size and width of arcs indicate the weights of connections, as shown in As shown in Fig. 2, the basic resistance rules of *S. aureus* and *CNS* are pretty similar. The main association rules are related to *penicillin*, *gentamicin*, and *ampicillin*, and most other antibiotics were linked to these three antibiotics and formed double, triple, or even quadruple resistance. This is because the drug resistance prevalence of *penicillin* and *rifaximin* is higher than others. Therefore, these three drugs are now less suitable for the widespread use, and other drugs should be used instead.



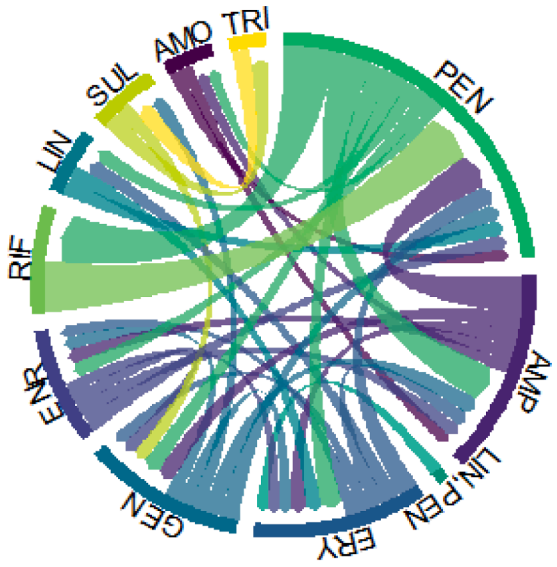
(a) *S. aureus*



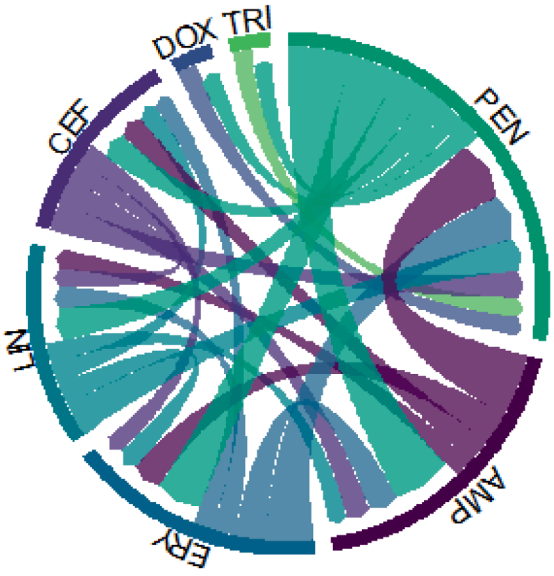
(b) CNS

Fig. 1. Results of association rule.

At the same time, we found that MDR in *S. aureus* is more complex, and all the drugs were strongly associated. *S. aureus* was resistant to *sulfasomizole*; probably they were also resistant to *trimethoprim*, forming double drug-resistant bacteria. Besides, if the bacteria were resistant to *ampicillin*, they would be resistant to *penicillin* and *enrofloxacin*, forming triple drug-resistant bacteria, and other drugs would also form different forms of MDR. For *CNS*, the MDR rules were relatively simple. In addition to the main rules related to *penicillin*, *gentamicin*, and *ampicillin*, *erythromycin* had a great correlation with *ampicillin* and *cefalexin*.



(a) *S. aureus*



(b) *CNS*

Fig. 2. Chord diagrams of association rules.

3.1.2. Results of MDR patterns

S. aureus is the main bacteria isolated by cow mastitis, and it also has more complex MDR rules. *CNS* is an environmental conditional bacterial, which may have probability of causing disease, but it is not the main pathogen. Therefore, in this section, *S. aureus* was selected to further analyze. Rules with support higher than 0.1 and lift higher than 1 were selected, as shown in Table 3.

Combining the results shown in Table 3 with the drug resistance prevalence in Table 1, it is observed the main patterns of MDR are double and triple-resistance patterns, and most of the antibiotics with a high frequency and association strength are mostly associated with each other, suggesting that the MDR patterns are very stable. This phenomenon is closely related to the widespread use of drugs in farms.

Table 3
MDR rules of *S. aureus*.

Rules	Support	Confidence	Lift
trimethoprim <=> sulfasomizole	0.154	0.979	3.646
lincomycin <=> erythromycin	0.131	0.765	3.506
gentamicin, penicillin => ampicillin	0.144	0.956	2.997
erythromycin, penicillin => ampicillin	0.134	0.741	2.324
erythromycin <=> ampicillin	0.148	0.647	2.030
ampicillin, penicillin => gentamicin	0.144	0.462	2.026
ampicillin, penicillin <=> erythromycin	0.134	0.430	1.972
ampicillin, rifaximin => penicillin	0.151	1.000	1.637
ampicillin, erythromycin => penicillin	0.134	1.000	1.637
ampicillin <=> penicillin	0.312	0.979	1.603
ampicillin, erythromycin => penicillin	0.144	0.977	1.600
erythromycin<=> penicillin	0.181	0.831	1.360
lincomycin <=> penicillin	0.131	0.765	1.252
rifaximin<=> penicillin	0.430	0.719	1.177
penicillin, rifaximin <=> ampicillin	0.151	0.352	1.103

As shown in Table 3, *trimethoprim* and *sulfamethoxazole* had the strongest correlation in all rules with a lift of up to 3.64, showing that when the bacteria were resistant to *trimethoprim*, there is a huge probability that they are also resistant to *sulfamethoxazole*. Specifically speaking, *trimethoprim* and *sulfamethoxazole* both belong to *sulfonamides* and most often used together to have a synergistic bactericidal effect by doubly blocking the folic acid metabolites of bacteria. One drug can inhibit the bacterial diacid dehydrogenase, and the other drug can inhibit the bacterial tetrahydrofolate dehydrogenase. They have good efficacy against animal diseases, and they also have a low resistance prevalence.

From the rule (*erythromycin*, *penicillin* => *ampicillin*) and the rule (*gentamicin*, *penicillin* => *ampicillin*), if the bacteria are resistant to *gentamicin* and *penicillin*, they are highly likely to be resistant to *ampicillin*, forming triple drug-resistant bacteria. Besides, Table 1 clearly shows that *S. aureus* has the highest resistance to these three drugs. Therefore, these drugs are not suitable for simultaneous use on farms. The (*rifaximin* => *penicillin*) rule has the highest support of 0.43, i.e., and this is the most popular pattern of double drug resistance. Both rules (*lincomycin, penicillin* => *erythromycin*) and (*lincomycin* => *erythromycin*) have a lift of 3.5, and when the bacteria were resistant to *lincomycin* and *penicillin*, the probability of them being resistant to *erythromycin* is also extremely high, and form triple drug resistance. The *penicillin* resistance prevalence is now as high as 0.70, which is a dangerous drug and not suitable for treating bovine mastitis. *Lincomycin*, which belongs to the *lincosamides* class and *erythromycin*, which belongs to the *macrolide* class, will produce antagonistic effects when used simultaneously, resulting in the decrease or disappearance of the drug effect. Therefore, in most cases, they should not be used in pairs. Finally, according to multidrug resistance rules, general drug recommendations for farm are as follows: 1) The drugs in high resistance are not suggested to use, especially *Rifaximin* and *Penicillin*; 2) if a drug that appears in the antecedent item in an association rule, it is best not to use a drug that appears in the consequents simultaneously.

3.2. Analysis of similarity results

To establish associations and compare differences between *S. aureus* and *CNS*, and assess the MDR, we analyzed the results of similarity. The results are shown in Table 4.

In general, the MDR patterns of *S. aureus* and *CNS* are pretty similar. The similarities of main rules are more than 0.7, and the highest value is up to 0.812. This is because while treating mastitis in cows, although the drugs mainly target *S. aureus*, they also have an effect on *CNS*. The antibiotics that appear most frequently in the rules are mainly *β-lactams*, and other antibiotics appear relatively infrequently. The similarity of (*ampicillin* => *penicillin*), (*erythromycin* => *penicillin*) in the *S. aureus* MDR rule stands at the highest with 0.812 and 0.793, indicating that

Table 4
The similarities of main association rules of *S. aureus* and *CNS*.

Rules	similarity
ampicillin <=> penicillin	0.812
erythromycin <=> penicillin	0.793
rifaximin <=> penicillin	0.775
ampicillin, rifaximin => penicillin	0.774
ampicillin, erythromycin => penicillin	0.774
erythromycin <=> ampicillin	0.757
erythromycin, penicillin => ampicillin	0.736
ampicillin, penicillin => gentamicin	0.726
gentamicin, penicillin => ampicillin	0.718
trimethoprim <=> sulfasomizole	0.708
ampicillin, penicillin <=> erythromycin	0.708
penicillin, rifaximin <=> ampicillin	0.69
ampicillin, erythromycin => penicillin	0.686
lincomycin <=> penicillin	0.613
lincomycin <=> erythromycin	0.459

both *S. aureus* and *CNS* have high resistance to *penicillin*, *gentamicin*, and *ampicillin*, and form MDR bacteria. Although these drugs have significantly contributed to the treatment of bovine mastitis, they are now unsuitable for use due to increasing resistance.

Except for the common features, a comparative analysis of MDR patterns revealed that different bacteria behaved differently in the association rules. Even for the same genus, the patterns of MDR are not the same. (*lincomycin <=> penicillin*) and (*lincomycin <=> erythromycin)* have higher support in *S. aureus* and lower support in *CNS*. Although *lincomycin* is frequently used, *CNS* does not often show MDR. Therefore, we need to differentiate and refine the drug use for bacteria to prevent drug resistance from becoming more severe.

4. Conclusions

In this paper, a weighted association rule mining algorithm was proposed to determine the MDR patterns of bacteria isolated from cow mastitis. Then a method was proposed for comparing the similarity of MDR rules in different bacteria. The results showed that high drug resistance to *penicillin*, *gentamicin*, and *ampicillin*, and most other antibiotics were linked with these three antibiotics. Besides, although the main rules are pretty similar in pathogenic bacteria, the MDR rules in *S. aureus* are more complex than those in *CNS* and should be treated with caution to differentiate the medication for bacteria to prevent further aggravation of MDR.

Besides, this method proposed in this paper is able to deal with drug resistance data and find more information from data driven perspective rather than biological methods, which provided a systematic solution and reference for highly complex and evolving MDR patterns.

In the future, we will collect more drug resistance data from different regions and years. Then combined with pharmacology and biological experiments, we will future analyze the time trend and difference in MDR rules of different years and obtain the overall change of resistance year by year, and dig out more useful information and provide a scientific basis for the changes of MDR and prevent the drug resistance from getting worse.

CRediT authorship contribution statement

Buwen Liang: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing. **Xinxing Li:** Conceptualization, Validation, Formal analysis, Resources, Writing – review & editing, Visualization. **Ziyi Zhang:** Software, Formal analysis. **Congming Wu:** Validation, Investigation, Resources, Data curation, Project administration, Funding acquisition. **Xin Liu:** . **Yongjun Zheng:** Conceptualization, Validation, Investigation, Resources, Data curation, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was supported by the China Agriculture Research System of MOF and MARA (grant No. CARS-36) and National Key Research and Development Project of China (grant No. 2016YFD0501304).

Reference

Acan, H., 2017. On a uniformly random chord diagram and its intersection graph. Discrete Math. 340, 1967–1985. org/10.1016/j.disc.2016.11.004.
Agrawal, R., Srikant, R., 1994. Fast Algorithms for Mining Association Rules. In: Proc of 20th International Conference on Very Large Data Bases, p. {VLDB'94}..
Bansal, M., Grover, D., Sharma, D., 2017. Sensitivity Association Rule Mining using Weight based Fuzzy Logic. Glob. J. Enterp. Inf. Syst. org/10.18311/gjeis/2017/15480.
Cazer, C.L., Al-Mamun, M.A., Kaniyamattam, K., Love, W.J., Booth, J.G., Lanzas, C., Gröhn, Y.T., 2019. Shared Multidrug Resistance Patterns in Chicken-Associated Escherichia coli Identified by Association Rule Mining. Front. Microbiol. 10, 1–15. org/10.3389/fmicb.2019.00687.
Ceglar, A., Roddick, J.F., 2006. Association mining. ACM Comput. Surv. org/10.1145/1132956/1132958.
Cheng, W.N., Han, S.G., 2020. Bovine mastitis: risk factors, therapeutic strategies, and alternative treatments — A review. Asian-Australasian. J. Anim. Sci. 33 (11), 1699–1713.
Clifford, K., Desai, D., Prazeres da Costa, C., Meyer, H., Klohe, K., Winkler, A., Rahman, T., Islam, T., Zaman, M.H., 2018. Antimicrobial resistance in livestock and poor quality veterinary medicines. Bull. World Health Organ. 96 (9), 662–664.
Collignon, P., Beggs, J.J., Walsh, T.R., Gandra, S., Laxminarayan, R., 2018. Anthropological and socioeconomic factors contributing to global antimicrobial resistance: a univariate and multivariable analysis. Lancet Planet. Heal. 2 (9), e398–e405.
Davies, R., Wales, A., 2019. Antimicrobial Resistance on Farms: A Review Including Biosecurity and the Potential Role of Disinfectants in Resistance Selection. Compr. Rev. Food Sci. Food Saf. 18 (3), 753–774.
Gao, X., Fan, C., Zhang, Z., Li, S., Xu, C., Zhao, Y., Han, L., Zhang, D., Liu, M., 2019. Enterococcal isolates from bovine subclinical and clinical mastitis: Antimicrobial resistance and integron-gene cassette distribution. Microb. Pathog. 129, 82–87. org/10.1016/j.micpath.2019.01.031.
Hipp, J., Güntzer, U., Nakhaeizadeh, G., 2000. Algorithms for association rule mining — a general survey and comparison. ACM SIGKDD Explor. Newsl. org/10.1145/360402.360421.
Hudson, J.A., Frewer, L.J., Jones, G., Brereton, P.A., Whittingham, M.J., Stewart, G., 2017. The agri-food chain and antimicrobial resistance: A review. Trends Food Sci. Technol. 69, 131–147. org/10.1016/j.tifs.2017.09.007.
Hvistendahl, M., 2012. China takes aim at rampant antibiotic resistance. Science (80-.). 336, 795. org/10.1126/science.336.6083.795.
Kumbhare, T.A., Chobe, S.V., 2014. An Overview of Association Rule Mining Algorithms. Int. J. Comput. Sci. Inf, Technol.
Lamma, E., Riguzzi, F., Storari, S., Mello, P., Nanetti, A., 2003. Discovering validation rules from microbiological data. New Gener. Comput. 21, 123–133. org/10.1007/BF03037630.
Li, X., Liang, B., Xu, D., Wu, C., Li, J., Zheng, Y., 2020. Antimicrobial Resistance Risk Assessment Models and Database System for Animal-Derived Pathogens. Antibiotics 9 (11), 829. https://doi.org/10.3390/antibiotics9110829.
Ludwig, A., Berthiaume, P., Boerlin, P., Gow, S., Léger, D., Lewis, F.I., 2013. Identifying associations in Escherichia coli antimicrobial resistance patterns using additive Bayesian networks. Prev. Vet. Med. 110, 64–75. org/10.1016/j.prevetmed.2013.02.005.
Ma, L., Tsui, F.C., Hogan, W.R., Wagner, M.M., Ma, H., 2003. A framework for infection control surveillance using association rules. AMIA Annu. Symp, Proc.
Michael, A., Kelman, T., Pitesky, M., 2020. Overview of Quantitative Methodologies to Understand Antimicrobial Resistance via Minimum Inhibitory Concentration. Animals 10 (8), 1405. https://doi.org/10.3390/ani10081405.
Motaung, T.E., Petrovski, K.R., Petzer, I.M., Thekisoe, O., Tsilo, T.J., 2017. Importance of bovine mastitis in Africa. Anim. Heal. Res. Rev. 18, 58–69. org/10.1017/S1466252317000032.
Nagasawa, Y., Kiku, Y., Sugawara, K., Hirose, A., Kai, C., Kitano, N., Takahashi, T., Nochi, T., Aso, H., Sawada, S.I., Akiyoshi, K., Hayashi, T., 2019. Staphylococcus aureus-specific IgA antibody in milk suppresses the multiplication of S. aureus in infected bovine udder. BMC Vet. Res. 15, 1–14. org/10.1186/s12917-019-2025-3.
Rajamanickam, K., Yang, J., Sakharkar, M.K., 2019. Phytochemicals as alternatives to antibiotics against major pathogens involved in bovine respiratory disease (BRD) and bovine mastitis (BM). Bioinformation 15 (1), 32–35.
Safdari, R., GhaziSaeedi, M., Masoumi-Asl, H., Rezaei-Hachesu, P., Mirnia, K., Samad-Soltani, T., 2020. Knowledge discovery and visualization in antimicrobial resistance surveillance systems: a scoping review. Artif. Intell. Rev. 53 (1), 369–406.

Sommer, M.O.A., Munck, C., Toft-Kehler, R.V., Andersson, D.I., 2017. Prediction of antibiotic resistance: Time for a new preclinical paradigm? Nat. Rev. Microbiol. 15, 689–696. org/10.1038/nrmicro.2017.75.

Su, M., Satola, S.W., Read, T.D., McAdam, A.J., 2019. Genome-based prediction of bacterial antibiotic resistance. J. Clin. Microbiol. 57 (3) https://doi.org/10.1128/JCM.01405-18.

Sun, K., Bai, F., 2008. Mining weighted association rules without preassigned weights. IEEE Trans. Knowl. Data Eng. org/10.1109/TKDE.2007.190723.

Thomas, V., De Jong, A., Moyaert, H., Simjee, S., El Garch, F., Morrissey, I., Marion, H., Vallé, M., 2015. Antimicrobial susceptibility monitoring of mastitis pathogens isolated from acute cases of clinical mastitis in dairy cows across Europe: VetPath results. Int. J. Antimicrob. Agents 46, 13–20. org/10.1016/j.ijantimicag.2015.03.013.

Van Boeckel, T.P., Pires, J., Silvester, R., Zhao, C., Song, J., Criscuolo, N.G., Gilbert, M., Bonhoeffer, S., Laxminarayan, R., 2019. Global trends in antimicrobial resistance in animals in low- And middle-income countries. Science (80-.). 365. org/10.1126/science.aaw1944.

Van Hecke, O., Wang, K., Lee, J.J., Roberts, N.W., Butler, C.C., 2017. Implications of Antibiotic Resistance for Patients' Recovery From Common Infections in the Community: A Systematic Review and Meta-analysis. Clin. Infect. Dis. 65, 371–382. org/10.1093/cid/cix233.

Xu, J., Sangthong, R., McNeil, E., Tang, R., Chongsuvivatwong, V., 2020. Antibiotic use in chicken farms in northwestern China. Antimicrob. Resist. Infect. Control 9, 1–9. org/10.1186/s13756-019-0672-6.

Yang, J.H., Wright, S.N., Hamblin, M., McCloskey, D., Alcantar, M.A., Schrübbers, L., Lopatkin, A.J., Satish, S., Nili, A., Palsson, B.O., Walker, G.C., Collins, J.J., 2019. A White-Box Machine Learning Approach for Revealing Antibiotic Mechanisms of Action. Cell 177, 1649-1661.e9. org/10.1016/j.cell.2019.04.016.

Zhang, C., Zhang, S., 2002. Association Rule Mining - Models and Algorithms. Lecture Notes in Artificial. Intelligence.

Zhang, Q.Q., Ying, G.G., Pan, C.G., Liu, Y.S., Zhao, J.L., 2015. Comprehensive evaluation of antibiotics emission and fate in the river basins of China: Source analysis, multimedia modeling, and linkage to bacterial resistance. Environ. Sci. Technol. 49, 6772–6782. org/10.1021/acs.est.5b00729.

Zhong, L., Ma, C.Y., Zhang, H., Yang, L.J., Wan, H.L., Xie, Q.Q., Li, L.L., Yang, S.Y., 2011. A prediction model of substrates and non-substrates of breast cancer resistance protein (BCRP) developed by GA-CG-SVM method. Comput. Biol. Med. 41, 1006–1013. org/10.1016/j.compbiomed.2011.08.009.

Zhou, X., Mayer-Hamblett, N., Khan, U., Kosorok, M.R., 2017. Residual Weighted Learning for Estimating Individualized Treatment Rules. J. Am. Stat. Assoc. org/ 112 (517), 169–187. https://doi.org/10.1080/01621459.2015.1093947.