AMMCS 2022

CS438

Analysis of food sampling data based on CARMA algorithm

Yongchun JIAO[a,b], Tongqiang JIANG[a,b], Tianqi LIU[a,b], Wei DONG[a,b], Qi YANG[a,b], Qingchuan ZHANG[a,b]
a School of E-business and Logistics, Beijing Technology and Business University, Beijing , China
b National Engineering Research Centre for Agri-Product Quality Traceability, Beijing Technology and Business University, Beijing,China

Introduction

China's national food sampling and inspection work refers to the State Council, provincial, municipal and county food supervision and management departments to carry out quality sampling and inspection work on the production and sale of food products in accordance with the law, which is one of the important technical supervision means for the supervision of various types of food products. In recent years, society has been paying more attention to food safety issues, and government food safety supervisory departments have increased their efforts in food sampling and inspection, thus accumulating a large amount of sampling data. By mining and analyzing food sampling data, meaningful correlations or interrelationships between sets of items hidden behind and between sets of data can be discovered.

In this study, we use the CARMA algorithm in association rules to establish association analysis models for qualified data and unqualified data, and analyze them separately, in an attempt to explore potential factors affecting food safety, identify priority and non-priority testing targets, and provide a reference for risk identification

Material

The data source for this study is the 2019 national food sampling database. The database mainly contains the following information: food id, food name, food category (category/sub-class, sub-subclass), time for sampling inspection,types of food packaging, testing site (province/city/district/county), sampling inspection sample No., pollutant category, detection value of pollutants, testing site, qualified item, etc.

Methods

CARMA (Controlled Auto-Regressive Integrated Moving Average) algorithm used in this paper was first proposed in 1999 by Professor Christian Hidber of Berkeley University, which improves the traditional association rule mining algorithm.Compared with other static association rule algorithms, CARMA algorithm has its obvious advantages:

(1) The algorithm has a higher execution efficiency than other static association rule algorithms, and usually performs at most two traversals to achieve the data set;

(2) The algorithm can handle data in Tabular format, and can also transform data in Transactional format;

(3) The algorithm can set the support degree for the antecedent and postecedent of the rule separately, and CARMA allows rules with multiple postecedents;

(4) The algorithm has a lower memory consumption.

The algorithm flow is shown in Figure1:

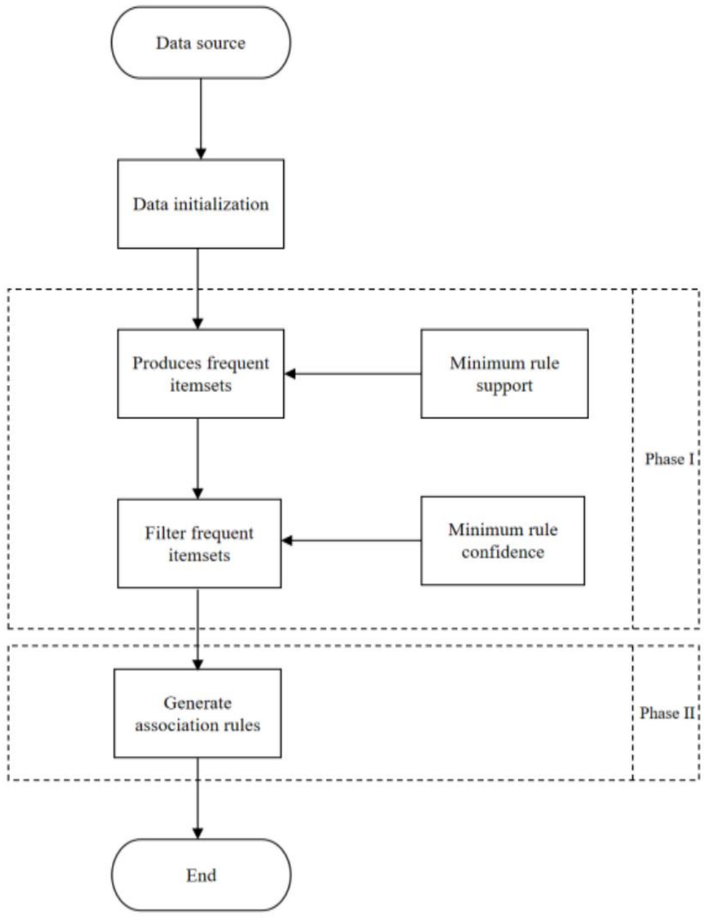

Figure 1 .CARMA algorithm flow chart

Results

Table .Association rules in condiments

Consequent	Antecedent	Support degree /%	Confidence /%
Relative risk level = low	Contaminants = Acesulfame and Aspartame and Sudan Red I and Benzoic acid and its sodium salt	98.12	100
Relative risk level = low	Province=Jiangsuand Packaging=glass bottle and Contaminant=dehydroacetic acid and its sodium salt and Contaminant=sodium saccharin	52.23	96.33
Contaminants = monosodium glutamateand Food packaging = plastic bags	Sampling site = supermarkets and Contaminants = disodium presenting nucleotides	19.8	83.54

Rule 1: Among the qualified semi-solid compound seasoning records, 98.12% of the tested samples contained the contaminants acesulfame, aspartame, Sudan red Ⅰ, benzoic acid and its sodium salt with low relative risk; 100% of the tested samples contained both contaminants acesulfame, aspartame, Sudan red Ⅰ, benzoic acid and its sodium salt with low relative risk level. It means that in the sampled semi-solid compound seasoning food, if the contaminants acesulfame, aspartame, Sudan red Ⅰ, benzoic acid and its sodium salt are contained at the same time, then the relative risk level of these four contaminants are low

Conclusion

By using the CARMA algorithm to classify the data of the major category of condiments in the national food sampling data during 2019 by food subcategories, association rules were mined for data in terms of province, sampling site, contaminant name, and relative risk level of contaminants in different categories of data, respectively, and eight effective strong association rules were obtained. Through further interpretation of the association rules, it can be found that: when several contaminants are present in certain condiments at the same time, the relative risk levels of the contaminants in such condiments are all low; the same sample may also have several contaminants failing at the same time, such as when contaminant A fails, contaminant B has a higher possibility of failing. Thus, it can be applied to identify the food category and the relative risk level of contamination, and reasonably determine the sampling items of the same type of food. For the higher risk of contamination of areas and food should focus on supervision, for those areas and food in good quality condition can extend the sampling cycle, reduce the frequency of sampling, reduce the detection of non-essential items.