A Novel Method for Mining Fuzzy Co-location Patterns

FSDM3691

Jinyu GUO, Lizhen WANG
School of Information Science and Engineering, Yunnan University, Yunnan, China

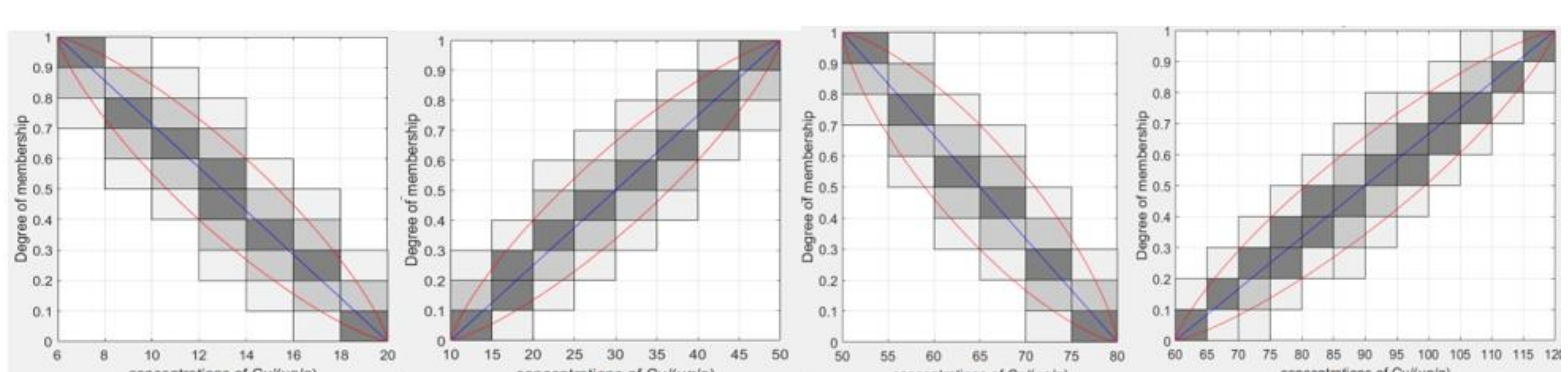
Introduction

We find that spatial data often contains attribute information besides spatial location information, such as the content of heavy metals in topsoil at a certain spatial location, which is usually expressed by specific numerical values. However, people are usually less sensitive to a specific value of attribute information, but when the attribute value changes fundamentally, that is, when it changes from a certain range to a new range, people are sensitive and attach importance to the result. In this paper, a method of mining fuzzy spatial co-location pattern based on type-2 fuzzy membership function and Join-based algorithm is proposed, which is used to mine fuzzy co-location pattern from spatial data sets with attribute information.

Methods

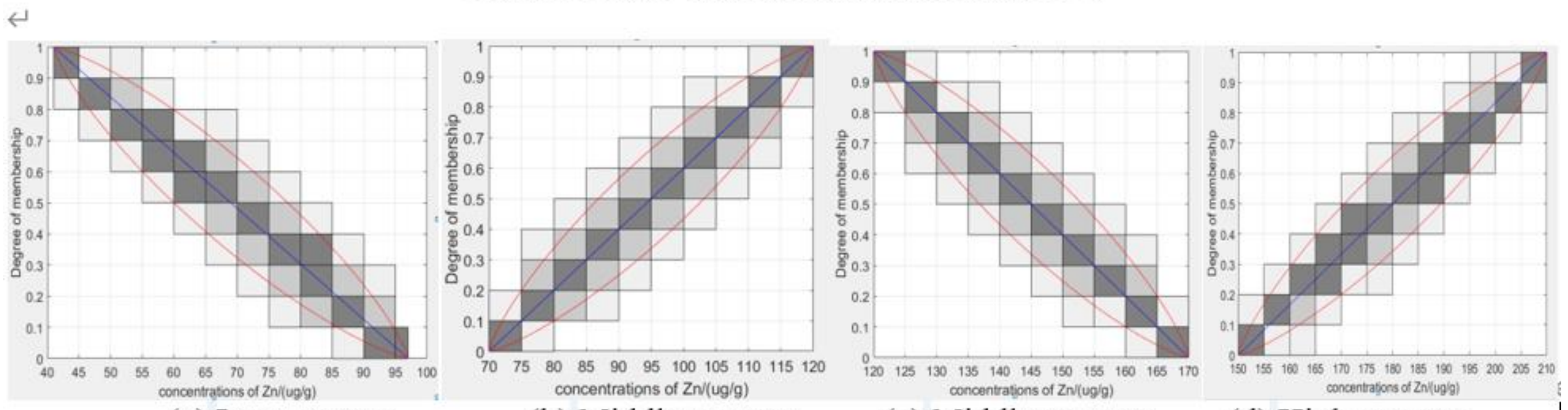
Firstly, we collected interval evaluation values of interval data of attribute information from 1000 experts, and formed granular data. Then, on the basis of the original type-1 membership function, a type-2 fuzzy membership function based on elliptic curve is expanded, and the parameters of the type-2 fuzzy membership function are adjusted by using a gradual method, so that the footprint of uncertainty (FOU) in the function meets the connectivity and the threshold given by the user. After that, we design a fuzzy co-location pattern mining algorithm incorporating type-2 fuzzy membership function into the traditional Join-based algorithm. In which, we define the concepts of fuzzy feature, fuzzy co-location pattern, upper bound participation index, lower bound participation index. In order to improve the efficiency of our method, we also put forward a pruning strategy.

Graphics / Images



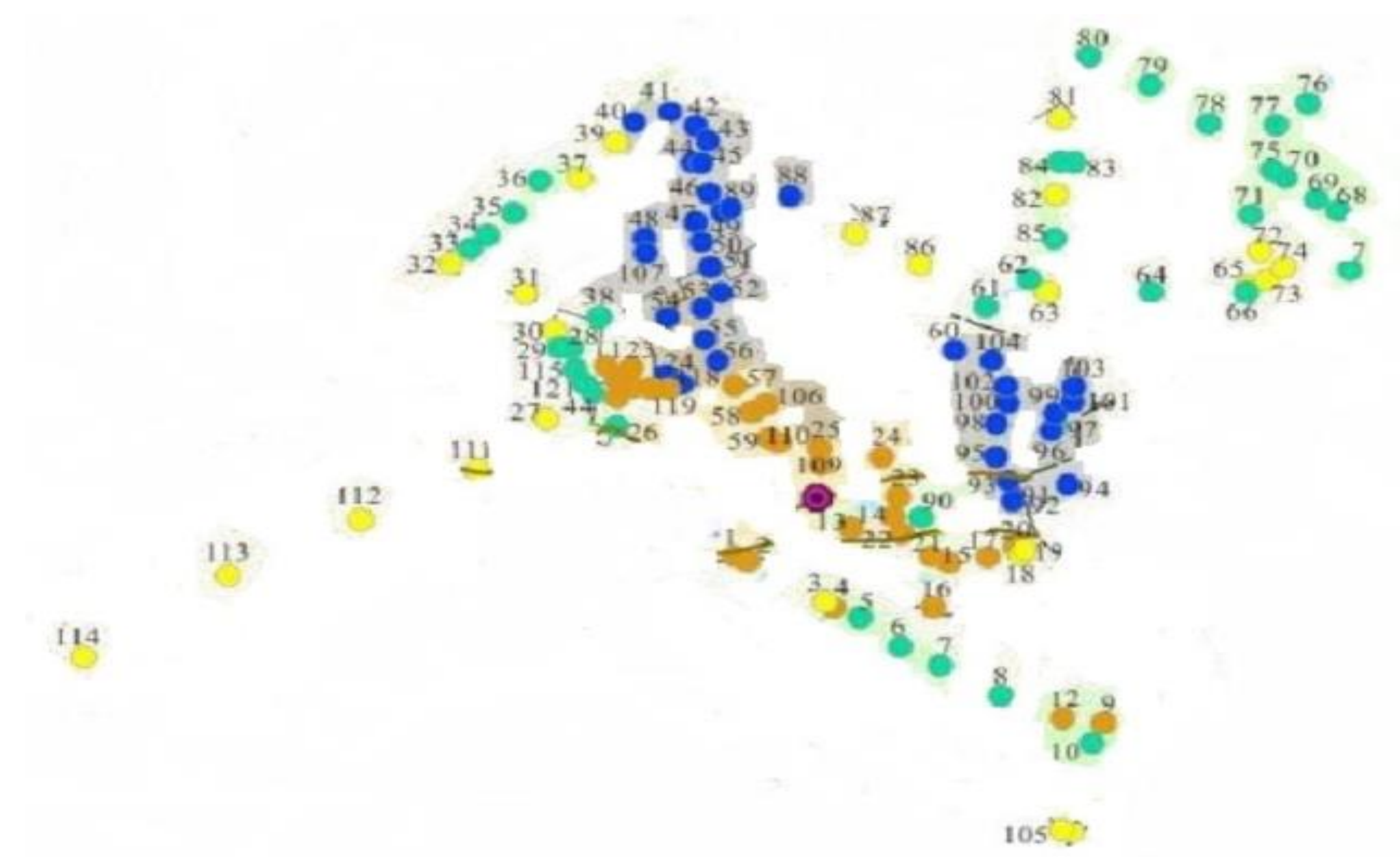
(a) Low content (b) Middle content (c) Middle content (d) High content

Figure 5. type-2 membership function of Cu



(a) Low content (b) Middle content (c) Middle content (d) High content

Figure 6. type-2 membership function of Zn



Prevalent fuzzy co-location pattern	Prevalence tendency degree	Lower participation index and Upper participation index	Participation index in traditional method
{A.Cu(L), C.Cu(L)}	0.6125	0.0709,0.6621	0.2312
{A.Cu(L), D.Cu(L)}	0.2328	0.036,0.3801	0.1169
{A.Cu(M), C.Cu(M)}	0.2636	0.1567,0.3513	0.2384
{D.Cu(M), D.Zn(H)}	1	0.4781,0.9135	0.6599
{D.Cu(L), D.Zn(M)}	0.9096	0.2414,0.8896	0.6681
{A.Zn(L), A.Cu(M), C.Zn(L)}	0.2636	0.1567,0.3513	0.2384
{A.Zn(L), C.Zn(L), D.Cu(M)}	0.2167	0.1749,0.3346	0.2492
{B.Cu(L), B.Zn(L), C.Zn(L)}	0.8586	0.2597,0.5448	0.4342
{B.Cu(M),B.Zn(L),C.Cu(M),C.Zn(M)}	0.7039	0.1804,0.5844	0.3358
{B.Cu(M),B.Zn(L),C.Cu(M),C.Zn(L)}	1	0.3933,0.738	0.5726

Conclusions

Traditional spatial co-location pattern mining methods can not directly mine fuzzy co-location patterns. Therefore, in this paper, we propose a fuzzy co-location pattern mining algorithm based on type-2 fuzzy sets and join-based algorithm to mine prevalent fuzzy co-location patterns from spatial instances with attribute information. Since the fuzzy attributes of spatial data are uncertain, we propose the concepts of upper and lower bound participation ratios of fuzzy features, upper and lower bound participation indexes of fuzzy co-location patterns, to measure the prevalence degree of fuzzy patterns. We also propose a pruning strategy, which effectively prunes the absolute non-prevalent fuzzy co-location patterns in the process of mining prevalent fuzzy co-location patterns.

The method of mining fuzzy co-location pattern based on type-2 fuzzy membership function has great practical significance. For example, it can help us find out whether the industrial areas will cause heavy metal pollution to the surrounding residential areas, and then find out whether the pollutants in the industrial area are related to the epidemic of the surrounding residents; In the distribution of urban facilities, we can see whether the per capita consumption level of the commercial areas is related to the housing prices of the surrounding residential areas, and so on. Our method also has some limitations: generating table-instances of candidate fuzzy patterns require a lot of connection operations, so the efficiency is not high; It does not consider how to mine fuzzy co-location patterns when different features have different attributes.

In the future work, we consider that there are several directions for further researches:

(1) Propose new fuzzy co-location pattern mining algorithms to improve the efficiency of the fuzzy co-location pattern mining method.

(2) Use more real data sets to verify the effectiveness of our proposed methods, for example, a data set of tumor diseases in a certain area and pollutant emissions from surrounding factories.