Knowledge-Based Systems 000 (2017) 1–10

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

# Modelling and application of fuzzy adaptive minimum spanning tree in tourism agglomeration area division

Wei Gao[a], Qiuluan Zhang[a], Zi Lu[a], Dianshuang Wu[b,*], Xiaohui Du[a]

[a] School of Tourism, Hebei Normal University, Shijiazhuang, PR China
[b] Decision Systems & e-Service Intelligence Lab, Centre for Artificial Intelligence, Faculty of Engineering and Information Technology, University of Technology Sydney, Australia

ARTICLE INFO

*Article history:*
Received 7 March 2017
Revised 1 May 2017
Accepted 5 June 2017
Available online xxx

*Keywords:*
Adaptive minimum spanning tree
Tourism source management
Decision making
Clustering
Fuzzy systems

ABSTRACT

Tourism agglomeration area division plays an increasingly important role in government's policy making on planning and development of tourism industry nowadays. With the development of ICT technologies, tourism “big data”, such as geographic data, tourist attractions evaluation data, tourist service data and related traffic information, becomes available to access, which provides great opportunities to develop intelligent tourism decision support systems (TDSS) for related government policy making. To effectively divide tourism agglomeration areas to support tourism-planning decision, by use of the “big data” resources, a fuzzy adaptive minimum spanning tree (F-AMST) model, which integrates adaptive minimum spanning tree (AMST) method and fuzzy level evaluation method, is proposed in this study. The F-AMST model consists of three parts: F-MST generation, F-MST splitting, and clustering solution evaluation and adjustment. The proposed model is then applied to cluster 142 A-level scenic spots in mountain areas of Hebei province, China, and the optimal tourism clustering solution with seven tourism agglomeration areas is finally obtained. The result is then analysed from the following aspects: the spatial agglomeration degree of each tourism agglomeration area is analysed by use of the fuzzy dense degree based on F-AMST model; the spatial distribution type of scenic spot nodes within each agglomeration area is analysed by using the nearest neighbour index R with consideration of spatial distance factors; the scenic spot node level system of each agglomeration area is analysed by use of the node level perfection index Z considering the scenic spot node level factor; the influences of the two factors, the spatial distance and the scenic spot node level, to the agglomeration degree are then analysed by use of the correlation between R and fuzzy dense degree and that between Z and fuzzy dense degree respectively. The findings are carefully described in this paper and the results can directly support government's decision making in tourism resources planning and construction of tourism agglomeration areas so as to improve the regional tourism competitiveness.



## 1. Introduction

Nowadays, the regional characteristics of tourism resources are explained more and more by use of network structures, data clustering and agglomeration theory. The tourism agglomeration area division is becoming increasingly important in the government's policy making on planning and development of tourist attractions. The construction of tourism agglomeration areas can effectively increase the polymerization development of regional tourism resources, expand the regional information channels, optimize the allocation of human and material resources, and improve the overall competitiveness of the region [1], which becomes an important aspect in government decision making. The rapid development of information technologies brings “big data” in many fields. Particularly, in tourism industry, much useful data, such as geographic data, tourist attractions and their level evaluation information, tourist services data and related traffic information, provides great opportunities to develop intelligent tourism decision support systems (TDSS) for government policy making. Some intelligent big data analytical methods have been used in tourism agglomeration area division [4–8]. The mainly used methods include K-means clustering, spatial autocorrelation and fuzzy clustering, among which K-means clustering was applied earlier and more widely. The graph-based clustering methods construct an appropriate graph and then divide the graph into a number of effective partitions to express the data intuitively and effectively [2]. Most graph-based clustering methods use K-Nearest Neighbour (KNN)

* Corresponding author.
*E-mail address:* dianshuang.wu@uts.edu.au (D. Wu).

http://dx.doi.org/10.1016/j.knosys.2017.06.007
0950-7051/

Please cite this article as: W. Gao et al., Modelling and application of fuzzy adaptive minimum spanning tree in tourism agglomeration area division, Knowledge-Based Systems (2017), http://dx.doi.org/10.1016/j.knosys.2017.06.007

algorithm to construct the graph. For example, in CHAMELEON clustering algorithm, Karypis et al. [3] used KNN graph to represent a data set, and merged the partition graphs through the interoperability and proximity of two clusters, which formed a number of homogeneous clusters.

In the traditional applications of K-means, there are drawbacks of subjectivity in the selection of cluster number and the initial cluster centres and the inability to identify different shapes of clusters. For example, Huang [4] subjectively specified the cluster number and the initial cluster centres when conducting the time and space clustering of tourist numbers of eco-tourism areas, and pointed out the drawbacks of the research results in their paper. Fuzzy clustering also lacks of objectivity in the choice of the threshold and related parameters. For example, when using fuzzy clustering method to analyse the tourism market, D' Urso P et al. [5] determined the membership functions according to their personal experiences, which affects the clustering results. Moreover, the above methods cannot reflect nodes' own attributes and lack of testing the effectiveness of different clustering solutions. To handle the above drawbacks, this paper proposed a fuzzy adaptive minimum spanning tree (F-AMST) model. F-AMST does not need the subjective selection of the coefficients, such as cluster number, initial cluster centres and thresholds; it combines node level weight factors, which can reflect the scenic spot nodes' own attributes, i.e., nodes' subjectivity; it also has the clustering validity evaluation process, which can test various clustering solutions and select the best one.

To comprehensively describe the application of F-AMST model, this paper conducts a case study on the A-level scenic spot nodes in mountain areas of Hebei province, China. The mountain areas of Hebei province are 122.9 thousand $km^2$, accounting for 65.48% of the total land area of the province, which consist of the Taihang Mountain through the western part of Hebei, Yanshan Mountain across the northern part of Hebei, and the northern Hebei mountain areas. Taihang Mountain area is related to Baoding, Shijiazhuang, Xingtai and Handan 4 cities; Yanshan Mountain area is related to Qinhuangdao, Tangshan, Chengde and Zhangjiakou 4 cities; northern Hebei mountain area is related to Chengde and Zhangjiakou 2 cities. The mountain areas of Hebei are rich in natural and cultural tourism resources, while the regional expressway and railway lines together constitute the traffic network system, which provides good basis conditions for the region's tourism development. However, there are many problems in the tourism development in these areas, such as similar scenic spots, lack of leading scenic spots, and the scattered spatial distribution of spots. The lack of spatial organization of scenic spots in these areas has greatly hindered the development of the tourism industry. It is urgent to improve the regional tourism space structure under the background of tourism supply side reform and global tourism, and implement the change from the development of single scenic spot to regional tourism agglomeration area [6–8].

This study offers both theoretical and practical contributions to spatial differentiation partition clustering: 1) the F-AMST model, which is based on the MST and improves the AMST model by adding fuzzy level weight coefficients, is proposed to further optimize the clustering method; 2) the proposed F-AMST model is applied to the tourism agglomeration area division, which can help realize the effective integration of tourism resources and strengthen the holistic and systematic development of regional tourism. The outcomes will contribute to decision support in big data environments in tourism industry.

The rest of the paper is organized as follows. The related works are described in Section 2. Section 3 presents the F-AMST model. In Section 4, the proposed F-AMST model is applied in tourism agglomeration area division of mountain areas in Hebei. The obtained tourism agglomeration areas are then analysed in Section 5. A set of conclusions is presented in Section 6.

## 2. Related work

MST as an important tool for data mining, was first proposed by Boruvka [9], which applied the connections between nodes into the power networks in order to obtain economic optimization. The following studies focused on finding the optimal MST algorithm, among which Prim algorithm and Kruskal algorithm are widely used. MST has the advantages of simple structure, intuitionistic efficiency, strict clustering, less influenced by subjective factors and independent of cluster geometry, which overcomes the limitation of traditional clustering algorithms [10]. However, the traditional MST also has the drawbacks such as complicated computation, more affected by noisy nodes and the inability to detect multiple clusters. To handle these drawbacks and combine the characteristics of various research fields, researchers have made a variety of improvements to the traditional MST. For example, Zhong C [11] constructed approximate MST by use of divide and conquer method, and proposed fast MST (FMST) based on K-means. Pirim H [12] gave a new objective function based on the tightness and separation index (TSI), and proposed a new minimum spanning tree based heuristic B-MST, which aims to perform internal test of the clusters. Jothi R [13] combined the traditional MST and spectral clustering algorithm, and proposed a novel clustering algorithm using Eigenanalysis on Minimum Spanning Tree based neighborhood graph (E-MST), which aims to improve the similarity between nodes within a cluster by use of multiple rounds of MST and conduct accurate analysis on the DNA microarray. Additionally, to solve the limitations of traditional clustering that only circular clusters can be detected, Assuncao [14] proposed static MST (SMST) and dynamic MST (DMST) to detect the clusters of any shapes. To deal with the limitation that only one clustering solution can be returned, Zhou R [15] proposed adaptive MST (AMST) by evaluating the effectiveness of the clustering results, which returns a variety of clustering solutions to determine the optimal clustering.

Based on previous studies, MST has been widely used in clustering analysis [16,17], image segmentation [18,19], density estimation [20], diversity assessment [21] and line optimization [22–24], which are related to biology [25–27], physics [28–30] and many other disciplines. In recent years, there have been regional tourism traffic optimization and line organization research results coming out in tourism science. For example, Bao et al. [31] used Kruskal algorithm to obtain the optimal spanning tree that connects the main tourist destination cities, and optimised the regional traffic through the shortest path principle. Bai et al. [32] built the tourism traffic optimization model by use of MST, and obtained the most economical connection path between two traffic nodes.

To summarise the above applications, it can be seen that as MST is suitable for node partitioning, it has great advantages in partition clustering analysis. The optimal solution can be obtained by multi-cluster verification, which solves the shortcomings of traditional clustering methods which are affected greatly by subjective selection of the clustering number, initial clustering centres, threshold and other related parameters. At present, the MST model has not been applied in the tourism agglomeration area division. In this paper, a fuzzy AMST (F-AMST) model, which is based on AMST model and deals with the lack of self-attribute index of AMST, is proposed and applied into the tourism agglomeration area division of the mountain areas in Hebei province. The F-AMST model combines the fuzzy level evaluation method, assigns fuzzy weights to scenic spot nodes' levels, and gives the objective function to judge the validity of multiple clusters. When applying the F-AMST model, the Euclidean distance between two scenic nodes is calculated by