An Ontology-Based Spatial Clustering Reasoning System

Wei Gu[1] Xin Wang[2]

Department of Geomatics Engineering
University of Calgary
Calgary, AB
T2N1N4
[1]wgu@ucalgary.ca
[2]xcwang@ucalgary.ca

Abstract

Spatial clustering, which groups similar spatial objects into classes, is an important research topic in spatial data mining. Many spatial clustering methods have been developed recently. However, common users may not know how to choose the most suitable spatial clustering method for their applications due to lack of expertise in the area. In this paper, we develop an ontology-based spatial clustering reasoning system. Using the system, the most suitable clustering method can be chosen for the target dataset with the support of the users' goal and a spatial clustering ontology. The system consists of the following parts: a spatial clustering ontology, an ontology reasoner, a web server and a user interface.

Background and Relevance

With the rapid growth of volume of spatial datasets, spatial clustering becomes an important topic in knowledge discovery research. It aims to group similar spatial objects into classes and is useful in exploratory pattern-analysis, grouping, decision-making, and machine-learning(Han et al., 2001). However, most existing clustering algorithms do not consider semantic information during the clustering process. Thus, a spatial clustering user needs to be familiar with the features of spatial clustering methods in order to choose a most suitable spatial clustering method for the dataset. In addition, users need to specify the parameters for the clustering algorithm, such as the number of clusters k for the k-Means method, which is also quite challenging for common users. Thus, providing knowledge support in clustering will be helpful for common users.

An *ontology* is a formal explicit specification of a shared conceptualization(Gruber, 1993). It provides domain knowledge relevant to the conceptualization and axioms(Wang & Hamilton, 2005). An clustering ontology can help us represent the knowledge in the spatial clustering domain.

In this paper, we propose an ontology-based reasoning system for spatial clustering. The purpose of the system is to guide a user to select an appropriate spatial clustering algorithm. The main contributions of the system are summarized below:

● The users' clustering goal is given at the semantic level. With the assist from a friendly user interface, users can easily find and use an appropriate clustering algorithm without knowing details about the algorithm.

● The ontology is represented in Web Ontology Language (OWL)(Owl Web Ontology Language Overview), the standard web ontology language and the system is built closely with the web service platform, so the system can be used to find clusters in web environments.

The Structure of System

Fig. 1 shows the structure of the system.

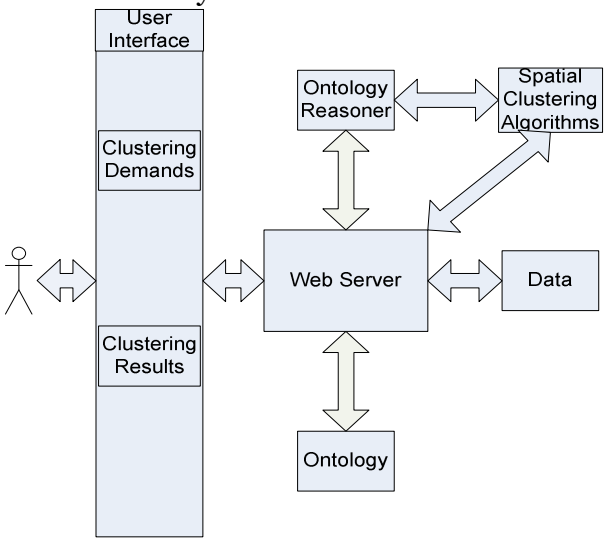

Fig. 1 The Structure of Ontology-Based Spatial Clustering System

The system performs spatial clustering reasoning under the following steps. First, the spatial clustering ontology is generated in a web ontology language and is posted on the Internet. Secondly, the users' goal is translated into queries that can perform reasoning on the ontology by Ontology Reasoner component. Then, the appropriate spatial clustering algorithms and spatial data sets are selected from the reasoning results. Thirdly, the selected clustering algorithm performs clustering on the datasets and clustering results are returned to the user.

A. *Clustering Ontology*

The ontology in the system is a formal representation of a set of concepts within the spatial clustering domain and the relationships between those concepts. The ontology can be used to explicitly represent the meaning of terms in vocabularies and make the information be easily accessed by computers. In the system, the spatial clustering knowledge in the ontology is organized by clustering techniques and clustering characteristics.

[5] provides a hierarchical classification of spatial clustering algorithms in terms of clustering techniques, shown in the Fig. 2. Every spatial clustering algorithm would be stored under a right node in the tree structure shown in the Fig. 2. For

example, K-Means algorithm is a partitioning clustering method based on distance, so it can be stored under the Node3.

Other ways to classify spatial clustering algorithms are based on the following characteristics: *Assignment Way, Attributes Type, Constraint, Dataset, Dataset Size, Dimensionality, Distance Measure, Measure Way, Noise Points Influence, and Search Way*. This approach tries to describe a spatial clustering algorithm from different aspects and extract algorithms' features. The ontology is developed using protégé-OWL. The snapshot is shown in Fig. 3.

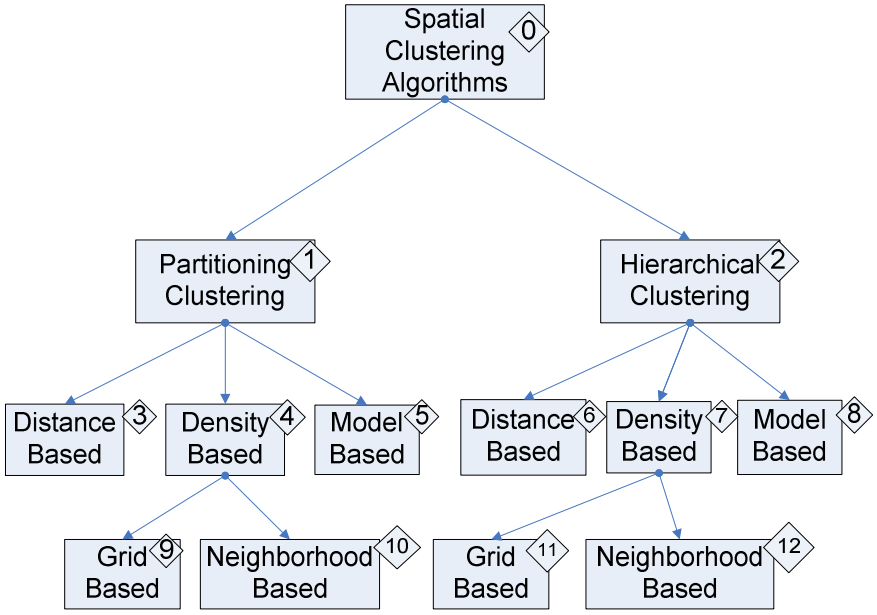

Fig.2 A Hierarchical Classification of Spatial Clustering Algorithms(Berkhin, 2002)

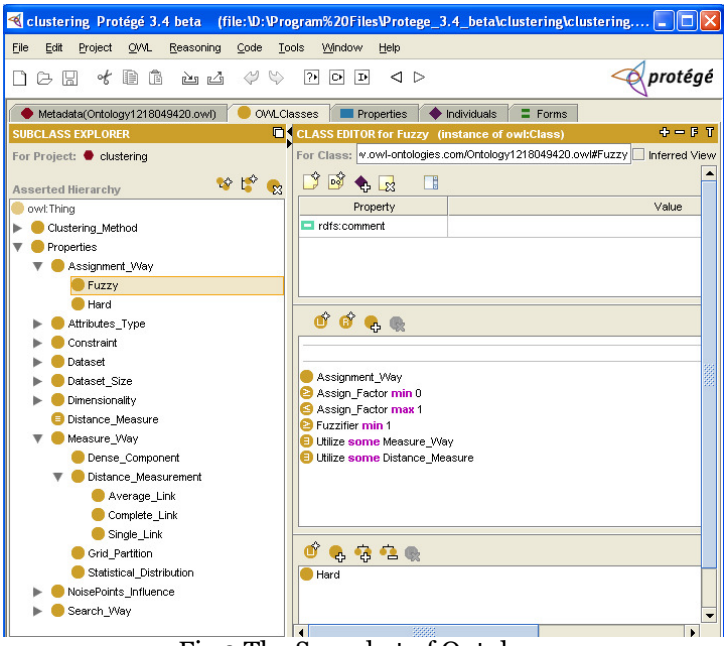

Fig.3 The Snapshot of Ontology

B. *Ontology Reasoner*

The ontology reasoner is used to reason the knowledge in the ontology. The input of the reasoner is the semantics information given by users, and the output is a set of appropriate spatial clustering algorithms. We use Pellet (Pellet website) as the reasoner for the system. Pellet is an open-source Java based OWL-DL reasoner. It provides functionality to validate ontology species, check consistency of ontologies, classify the taxonomy, check entailments and answer SPARQL queries.

C. *User Interface and Web Server*

User interface is used to receive the semantics requirements from users, transfer the information to web server and return clustering results to users. The users' requirements are described by type of attributes, scalability requirement, dimensionality of data, assignment method, outliers handling, defined parameters and constraints. Users can select the characteristics from the interface. The snapshots of user interface are shown in Fig.4.

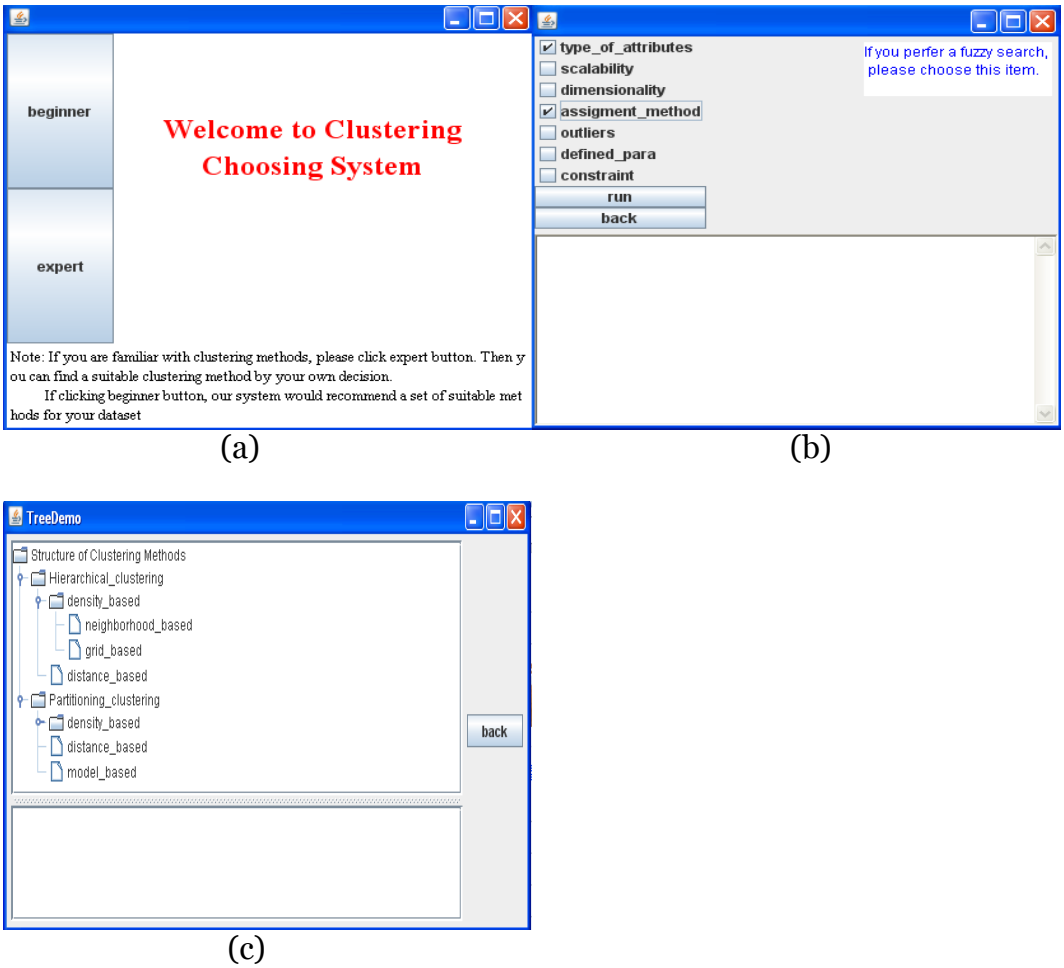

Fig. 4 The Snapshot of Use Interface (a) Main Page of the System, (b) Information Input Interface in Checkbox, (c) Information Input Interface in TreeView.

The web server built on Apache Tomcat (The Apache Software Foundation website) (which is an implementation of the Java Servlet and JavaServer Pages technologies) is responsible for searching an appropriate ontology on the Internet and sending the ontology to the ontology reasoner and return the clustering results to the user interface. Currently, only one spatial clustering ontology written by us is on the Internet for testing.

Conclusion

The major contribution of our work is to develop an ontology-based spatial clustering reasoning system. By reasoning the spatial clustering ontology, the system can automatically translate the semantics information into queries, then let user easily choose the most suitable spatial clustering algorithm without mastering the knowledge in spatial clustering domain.

References

Berkhin, P.(2002). *Survey of Clustering Data Mining Techniques*. Retrieved from Accrue Software Inc: http://www.ee.ucr.edu/~barth/EE242/clustering_survey.pdf

Gruber, T.R. (1993). A Translation Approach to Portable Ontologies Specifications. Knowledge Acquisition, 5(2).

Han, J., Kamber, M., & Tung A.K.H.(2001). *Spatial Clustering Methods in Data Ming: A Survey,* Geographic Data Mining and Knowledge Discovery(pp. 1-29). Taylor and Francis.

Owl Web Ontology Language Overview. Retrieved October 03, 2008 from http://www.w3.org/TR/owl-features/

Pellet website. Retrieved October 03, 2008 from http://pellet.owldl.com/

The Apache Software Foundation website. Retrieved October 03, 2008 from http://tomcat.apache.org/

Wang, X., & Hamilton, H.J.(2005). Towards An Ontology-Based Spatial Clustering Framework. In *Proceedings of 18th Conference of the Canadian Society for Computational Studies of Intelligence*(pp.205-216).