Applied Mathematics & Information Sciences
An International Journal

Research of spatial data cache index of electricity GIS

Bo Hu[1,2]

[1]North China Electric Power University, Beijing 102206 P.R. China
[2]Liaoning Electric Power Company Limited, Shenyang, China

Received: 8 Jul. 2012, Revised: 4 Oct. 2012, Accepted: 6 Jan. 2013
Published online: 1 May 2013

Abstract: This paper studies the application of present situation of the GIS platform in the power industry, in view of the existing problems, putting forward a kind of spatial data model of the power facilities for power industry, and putting forward clustering indexing method based on semantic meaning, and accelerating the response speed of the multi-level cache mechanism. And finally the article is presented the possible problems, and the future outlook of the development of electric power GIS platform.

Keywords: Spatial data, Index, Cache, GIS

1 Introduction

Since the modern power grid system has the following characteristics: large volume of data, electrical wiring complexity, dynamic change frequently, real-time information demanding. In order to meet the requirement of safe, reliable, high quality, efficient, economy in power industry, information technology and management tool were used in planning and design, to enhance the asset, operation and monitoring management, improve the power supply reliability and power quality of electric network, realize the digitization and informatization of power grid, to provide high-quality, efficient, safe service to customers. By using GIS core technologies in building power GIS platform, could realize the network data pool, provide system integration interface, and integrate power system in various applications, so as to achieve grid information integration and sharing the target. Spatial data management, query, analysis, and graphics update and display efficiency is the core in power GIS platform, and the platform of the response rate to a large extent depending on the spatial data index and cache mechanism.

Spatial index is one of the necessary means for improving the spatial query and retrieval efficiency, the domestic and foreign scholars in different application present various types of spatial index mechanism. R-tree is one kind of height-balanced tree with object defined technology, which is a natural extension of the B-tree in the k-dimensional space[1], as one of most popular dynamic spatial index structure, since the proposed by Guttman in 1984[2], through constant evolution and development, has formed the different areas of the variant [3]. For example, in order to avoid multi-path queries arising from the overlap of the R-tree node at the same level, Sellis in 1987 designed the R+ tree [4], to improve the retrieval performance; in order to improve the utilization rate of the storage node of R tree structure, and improve the storage utilization of the node, optimize the structure of the R-tree; Kamel, in 1994, proposed the Hilbert R tree [5] using one-dimensional linear ordering of k-dimensional spatial data, and then sort the tree nodes, to obtain area, perimeter to minimize tree nodes; considering the create process of R tree, Brakatsoulas proposed the clustering algorithm k-means cR tree in 2002 [6], which insertion cost is similar to the R tree, and query performance is close to the R * tree .etc..

Power grid GIS platform with industry-specific, the Index caching mechanism of common GIS platform applied to power grid GIS platform should consider the following questions [7].

First, the complexity of the spatial data model of electric power facilities. In addition to the traditional spatial data and attribute data management, also need to manage grid topology class data and devices running class data.

Second, relevance and semantic correlation of the topology of the power facilities. When calculating in the field of power grid, the equipment need repeated retrieval; when checking substation working condition, the need for

* Corresponding author e-mail: dianlihubo@sina.com

the substation into the line inquiry, when power failure analysis, the need for checking all tower for some line.

Third, real-time monitoring of grid operation. Measuring point real-time supervision requires updated the monitoring data in real time, at the same time, according to the operating rules for predicting and alarming; playback the operation status of equipment, meet on the temporal and spatial characteristics of processing equipment.

Fourth, the specificity of power grid GIS layer. Power grid facility layer update frequency is high, geographic layers is relatively stable, must use the high speed display cache strategy, and satisfy the real-time processing requirement of incremental data.

The goal of commonly used spatial index cache mechanism is to improve query efficiency of spatial data. Some of typical queries, such as range queries (search within a given range of space objects), point location queries (range queries are a special case), and nearest neighbor query (search for a given object is the closest space object), etc., direct at the issues raised above, are birth defects. This article is based on traditional spatial index cache, considering the characteristics of the power GIS platform for spatial data, put forward a set of index caching mechanism for the power industry GIS platform substantially increase the response speed of the power of GIS platform as well as spatial data query and analysis efficiency.

2 GIS spatial data model of power grid

The spatial data model is the core of GIS spatial data management, spatial data model design quality, directly related to spatial data storage and read-write efficiency, and affect the performance of the entire GIS system. Power GIS spatial data management, is not only the equipment management of attribute class data, space equipment and geospatial geographic distribution of class data, but also the management of topological space topological structure of power grid data and electrical physical space equipment operating data. Meanwhile integrating into the rules of the power industry and electricity operating parameters. Therefore, the spatial data model as the core of traditional point, line, surface, management and operation will not be able to meet the power GIS spatial data management requirements.

Power facility of full life-cycle management. Object in the grid infrastructure repository is a digital description of the reality grid infrastructure software space, therefore, the power facilities model must define the various stages of the whole life cycle of the facilities, each state, such as planning, design, construction, in service and disposal, demolition and so on, must be process-oriented managed.

Model complexity of the power facilities. The power GIS spatial data model not only manage spatial data and attribute data in traditional GIS, but also need to manage grid topology class data as well as electrical equipment running class information.

The topological correlation of the power facilities. Power GIS grid need connectivity analysis, retrospective power point, power flow calculation, short circuit calculation in power calculations, also need storage topology detail information of the power facilities.

The semantics of the power facilities. For example: when graphics rendering, a range of equipment need repeated retrieval; when checking the substation working conditions, the need for the substation outlet, into line inquiry; when blackout analysis, the need for checking all tower of some line correspond.

Power real-time operation of the facility. Measuring point data need update the monitoring data in real time, and alarm according to the rules of operation; playback equipment operation, to deal with spatial and temporal characteristics of the equipment.

In view of the power GIS spatial data model, analysing the power GIS construction process to the facility data conversion, feature extraction, and the urgent demand of network topology and network knowledge base, put forward power GIS spatial data model of the system concept, as shown in Figure 2.1.

The information, based on the equipment ledger management, and the management geometric information of point, line, surface, and based on the electrically connected topological information, and based on the running equipment electrical physical information, to power facilities are equally important, designing the specificity components for each type of information are need. At the same time in order to reduce the running maintenance of power facilities, management model, the complexity of management, will need various spatial domain knowledge object, as with all kinds of component object peer components. Considering the above factors, proposed one kind of based on the knowledge base of the “four-dimensional, four-part” power facility spatial data model.

Domain knowledge objects, and geospatial collection of Component Object, the object of the properties of components of the device space, topological space topology component object and run component objects of the power physical space objects “four part” together to form a “four-dimensional” (the three space dimension and time dimension) space power facilities. it will need to describe abstraction of Power facilities as an independent logical characteristic component, loosely coupled, scalable space entity object[8], this space entity will manage the different components of the model of the same space, according to component characteristics, frequency of visits to design appropriate data structures and storage mechanisms, improving the response performance of spatial data, and be able to adapt to the diversity of spatial data, and establish links from the common characteristics, affiliation, connection, topological relations, spatial and temporal relations, which can meet the diverse needs of the traditional spatial

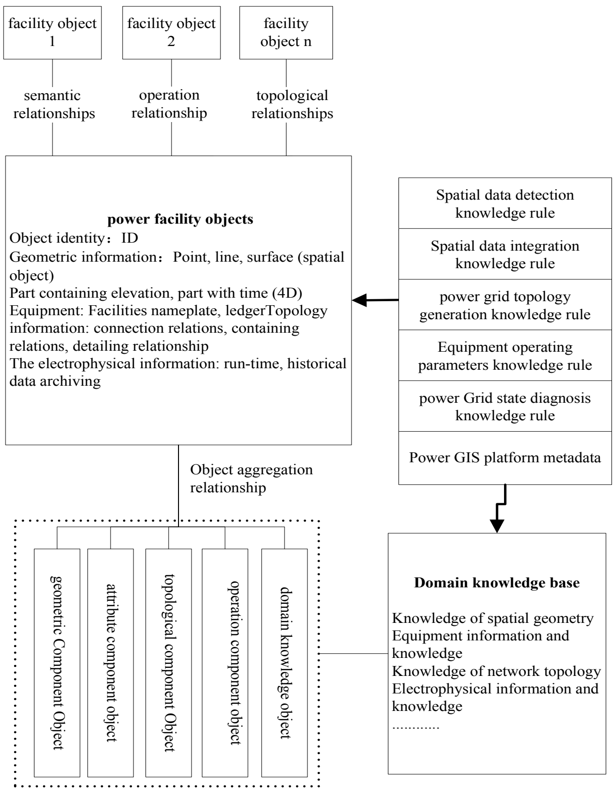

Figure 2.1: Spatial data model of power facilities

data manipulation, redundant storage of the traditional spatial data model, topology difficult maintenance, and temporal characteristics of expression, semantic independence, etc.

On the basis of domain knowledge base, Building up corresponding knowledge rules, can assist in management of spatial data model of power facilities, to ensure the accuracy of the spatial data, and achieve enhanced based on the function of the spatial data, and improve the response speed of spatial data. The rule of test knowledge and integration knowledge in spatial data can determine the validity of spatial data from geometry, and fusion space multi-source data; the rule of grid topology to generate knowledge can help to generate the topology of the grid, one-time wiring diagram and provide a basis for grid power tracking, fault locating and other applications to; the rule of equipment operating parameter of knowledge and grid status diagnostic knowledge, track the status of the operation, and ensure the operation of the grid normally.

Building up the electric power GIS platform metadata management mechanism, that will be unified storage management for all space model metadata, platform management model, the role of rights management, to facilitate the establishment of a long-term synchronous updating and spatial data management system.

Through the perfect metadata management system, realizing a large number of concurrent user control, mash up the vector data and raster data, rasterization of vector data management, support the massive vector data, while supporting the client Spatial SQL query, convenient to the realization of query positioning function.

3 Power GIS spatial indexing mechanism

3.1 Analysis of spatial indexing methods

The four fork tree index and the R tree index in current mainstream GIS platform, in particular for electric power industry GIS application cannot function well in complex power facilities. In a complex power facilities spatial data model, spatial multidimensional just is a respect that we concerned. In addition, the topological structure of power grid, the grid operation information, and the association with network equipment, need attention spatial information at the same time.

Therefore, When creating the index of the power facilities spatial data model, should calculate semantic distance, unified index, Only in this way can we ability in power GIS platform to achieve power facilities and efficient data access.

According to the characteristics of R-tree dynamic insertion algorithm, its node splitting algorithm performance directly affects the performance of the R-tree and query efficiency. R-tree index requires that the objects are in close proximity in space as much as possible in the same node, in order to meet the requirements of complex spatial data in multidimensional R-tree, many experts and scholars put forward the concept of clustering R-tree optimize the node split on the basis of the traditional R-tree.

Expanded on the basis of the concept of the general clustering algorithm and R-tree, logically related to the object (The possibility of these objects is the same query access) stored in the same node of the R-tree can greatly reduce the overlap region of the R-tree and blank space, as same as greatly improve the efficiency of the node splitting and query. It is verified that the cluster R-tree algorithm to build R-tree and R-tree, and significantly lower than the R * - tree, and search efficiency significantly higher than the R-tree and R * - tree [9].

Cluster analysis is divided the data set into clusters so that cluster internal data as far as possible similar to each other and between data of clusters as much as possible different, different clustering methods using different similarity measures and technology. Traditional clustering methods in the processing of data, does not consider the semantic relationships between data, thus leading to the clustering is not satisfied, and the deadlock may often happen. Especially for multi-dimensional spatial data, traditional clustering methods often ignore the relationship between the various dimensions, the similarity calculation between the data is not accurate enough, the clustering results were not reasonable, which could not provide a good clustering mechanism.

Reference[10] is a research based on the semantic distance of the text clustering algorithm, selecting semantic ontology, using semantic distance calculating the similarity between documents, transform the similarity of document to the semantic distance between words, and the semantic distance between the meaning of the original. In the aspect of class center, to introduce the concept of similar weight, and candidate the fittest classes. Document Clustering results based on semantic distance shows which is from the more semantically subdivided theme to provide better navigation for users to collect text information.

In this paper, based on the research of cluster analysis and facilities model semantic similarity, proposed an improved R-tree spatial index for the electricity industry, that is semantic clustering R-tree (ScR-Tree).

3.2 R-tree index based on semantic clustering

R-tree index requires the objects which are in close proximity in the space as much as possible in the same node, that reflects the characteristics of spatial data clustering. Clustering is an effective way to improve spatial data query processing performance. Using Semantic-based clustering algorithm to achieve the R-tree node split, the criteria of Splitting is various types of semantically related objects as possible to be in the same node, which greatly improve the efficiency of query access.

ScR-Tree, in the node splitting process, establish contact through the semantic correlation between nodes, each cluster choose a super node which is responsible for calculating the class of all node semantic similarity and semantic distance. In this way, it could be effectively generate the R-Tree indexes, and achieve high node hit rate and low cost to the query process.

According to the electric power facilities spatial data model of “four parts” feature, divide each power facilities (“nodes”) into four parts, the semantics of the node is described by the semantics of each category and the corresponding categories of weight, For example: Feature = (c1, w1, c2, w2, c3, w3, c4, w4), in which the Feature means each facility, ci ($1 \leq$ i ≤ 4, i $\in$ n) means the keyword of each category in the electric power facilities spatial data model. wi (1≤i≤4, i∈n) means the weight of each category in the electric power facilities spatial data model. Semantic distance between facility X and facility Y, show as following (1), the distance in each category were calculated, then product the weight, that expressed the semantic

$$D(X,Y) = \sum\nolimits_{1\leq i\leq 4, i\in n} (C_{X,i} * C_{Y,i}) \times W_i \qquad (1)$$

Specific indexing process is as follows, selecting k cluster centers with a typical point method, using semantic distance, clustering the facilities, update the cluster centers according to the results of the clustering, repeated until the clustering results unchanged. Such indexing method, the cluster core algorithm is no longer to be simple calculation the distance between the spatial objects , but make the model of power facilities as a whole, and calculate the semantic distance, which could more enhanced index semantic correlation between the power facilities.

Semantic based clustering of R-Tree, on the basis of the power GIS facility model presented above, using the knowledge base, for a general clustering algorithm, expanding the aspect of semantic, that can be used for splitting R-tree node which is integrated various types of the dynamic data, and the query efficiency is greatly improved. Based on semantic clustering R-Tree spatial data model for electric power facilities, is applied huge amounts of data and efficient spatial data indexing methods.

4 Power GIS caching mechanism

Power GIS model is very complex, with large volumes of data and high-quality real-time requirements,at the same, with the frequent operation of history playback and real-time tracking, thus multi-level spatial data cache can effectively reduce the pressure on the server and network load, realize the efficient real-time access to spatial data. It is essential for improving system efficiency and reduce the map response time[11].

In order to quickly and efficiently handle massive spatial data stored in spatial database, In the overall power Web GIS sharing platform, multi-level cache, is divided into the database-side, server-side and client-side. Dynamic cache of platform model shown in Figure 4.1. Creating spatial index and the tile cache for the database-side, using attribute queries instead spatial query to improve query efficiency; server-side through management of vector cache server and map tile server, Using high-speed dynamic display cache (DDC) cache dynamic layers of vector data, user data and layer version, caching a background map in the form of tile data, and using distributed deployment strategy to reduce network traffic, and accelerate the access speed, substantial increase concurrent users ability and the cache server’s efficient and stable; client-side through the Intelligent Agent Technology, asynchronous transfer technology to communicate the server-side cache, Cache set different access levels by frequency of visits, realizing the sensitive interaction and rich operational experience, to improve the response efficiency[12].

The high-speed dynamic display of server-side caching and background map tile cache, not only to meet the requirement of the data real-time change for electric power facilities, as well as the vector graphics rendering, meanwhile able to quickly generate a map without having to query the back-end database, meeting in the

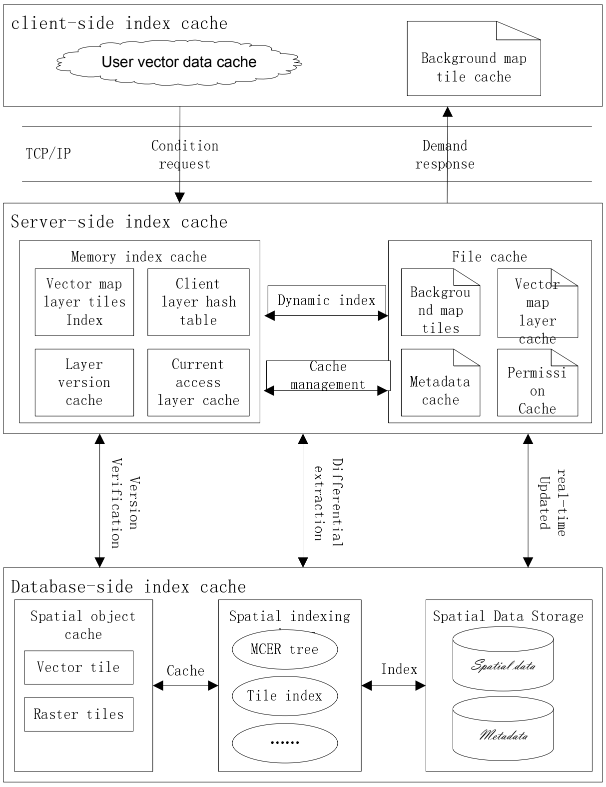

Figure 4.1 power GIS dynamic cache model

management of huge amounts of data while the system has a high operating efficiency. Display cache can be flexibly placed anywhere in the network, considering the system performance, storage space and network traffic. Demand for real-time updating, using the DELTA mechanism, incremental updating dynamic information, through the combination of DDC + DELTA, could better support for huge amount of data and accelerate speed and correctness of the data access.

The client cache based on the fact: during a period of time, the user view and retrieval of the power facilities is concentrated in certain map layer. According to the access frequency of the layer, the data cache should be in the different levels of cache. When the user retrieving data, we should check started from the fastest cache retrieval, if not retrieved, then, to the next cache retrieval. The client's local cache content, including all or part of the DDC, the visited background map tiles, as well as the various types of space facility version of the model data, and user's permissions, to achieve real-time requirements through the server-side DELTA mechanism when facilities update occurs.

The multi-agent technology is applied to the WebGIS platform based electric power client, provide the basic map shows, browse function, spatial analysis functions, as well as resource query positioning function, and is able to manage the local cache map data and user data, each client hosting a client's Agent, as the link between the various client, using mobile Agent to complete the collaborative mapping functionality between the clients and cached data sharing.

Power GIS platform using multi-level caching mechanism is able to meet the efficient data access and real-time requirements for the various products, including desktop GIS, WebGIS, mobile GIS, power professional applications .etc., which could significantly improve the hit rate of the data, and effectively speed up the data read speed.

5 Conclusion

This paper analyzes the application of the limitations of the GIS platform in electricity, puts forward for the power industry power GIS space facility data model,based on the spatial data model of "four-dimensional, four-part", introduces a kind of special spatial data indexing method-Based on semantic clustering R-Tree index, and finally proposed dynamic caching mechanism for spatial data. The following problems might exist in a specific application process:

The promotion of information technology of the Electric power system is already at very high level, how to co-ordinate all aspects of application, and

power facility model extended to data center management is a big problem.

Based on the "four part" model of electric power facilities, the ontology proposed, and the establishment of knowledge base, which both need a lot of work, at the same time keyword selection, as well as the weights of the measure are required for the accumulation of experience, these are met will be optimal.

The application of multi-level caching mechanism, needs consider network conditions, as well as the constraints of the client, and the strict data security policy.

Index caching mechanism for the power GIS platform can well solve the existing problem such as the slow response, and the user experience not well .etc., but the problems may arise need to be resolved in the specific application.

Acknowledgement

Project supported by the National Natural Science Foundation of China (Gean No:61273029).
The authors are grateful to the anonymous referee for a careful checking of the details and for helpful comments that improved this paper.

References

[1] Qu Ming-Cheng, Wu Xiang-Hu and Yang Xiao-Zong. Applied Mathematics and Information Sciences, **5(2)**: 113-121.

[2] Zhao Xibin, He Anping etc. Applied Mathematics and Information Sciences, **6(2)**: 345-356.
[3] V. Gaede and O. Gunther, ACM Computing Surveys, **30(2)**: 170-231.
[4] Guttman. A. In Proceedings of the ACM SIGMOD Conference on Management of Data, 47-57.
[5] Zhang Mingbo, Lu Feng and Shen Pai Wei, Journal of Computers, **28(3)**: 289-300.
[6] T.K. Sellis, N. Roussopoulos and C. Faloutsos, Proceedings of the 13th VLDB, Brighton, England, 507-518.
[7] I. Kamel and C. Faloutsos. Proceedings of the 20th VLDB, Santiago, Chile, 500-509.
[8] Wu Kehe, Wang Xiaohui and Bai Xuewei. FCC2010, EI, ISTP.
[9] Li Wei, Gong Shuai and Wang Xiaohui, Grc2010, EI.
[10] Li Qinshuang and Zhang Chunsheng,Journal of Surveying and Mapping, **18(3)**: 201-203.
[11] Lin li, Xiamen University.
[12] Zhu Guiming and Jin Shiyao. International Symposium on Intelligent Ubiquitous Computing and Education, 92-96.
[13] Zuo Tianhui, Yu Jinkang and Gao Jin, Hydropower energy science, **02**.
[14] Wang Zhenyuan, Gong Weihua and Zhou Yingbiao. Computer Science, **32(12)**: (2005).

BO HU received his master degree in Computer Science from North China Electric Power University in 2006, since then, he continued studying his Ph.d of Electric Power System and Automation in North China Electric Power University. He is now a senior engineer in Liaoning Electric Power Company Limited. He is interested in researching the areas of automatic of electric power system, information systems, and data mining.