Multiple Spatial Model Fusion in Heterogeneous Sensor Networks

Feng Jiangfan* and Wei Wei

College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

E-mail: fengjf@cqutp.edu.cn

* Corresponding Author

Abstract

With the growth of location-based services (LBS), location-based data is a crucial role for many sensor-network applications. Although the conversion of data formats in heterogeneous sensor networks has been studied extensively, the integration of spatial data models is not well understood, which is important to understand the semantics of spatial data in sensor networks. To determine the spatial data integration in model level, we have investigated the cognitive model of spatial information, the conceptual model of space phenomena, and the logic of spatial data management mode, leading to the integration of various spatial data sets from the view of model integration. Furthermore, we provide a method for the effective storage and management of massive and multi-source heterogeneous spatial data. Ultimately, the experimental result shows that the method offers a better way of model integration that cannot be achieved with existing schemes.

Keywords: Geographic Information System, Integration, Location Based Services, Sensor Networks

1. Introduction

Spatial data with space, time and thematic attribute are the groundwork of the digital city and smart city. With the development of the Internet of things, spatial data obtained from wireless sensor networks always have the features of being numerous, multi-resource, dynamic, isomorphic and distributed. In order to get the panorama of information in aggregate and analyze several different types of record simultaneously, it must have heterogeneous sensor information integrated effectively to guarantee the perceptual interaction. However, due to the same geographic space phenomenon have differences on the cognitive perspective, industry demand, data acquisition method and data modeling method by different spatial data gatherers, there are multidimensional heterogeneous characteristics of the multiple spatial dataset with different spatial data model, database schema, spatial operators and even data file format. Isolated island of information formed between industries, even different parts of the same industry, repeated data acquisition and data redundancy among diverse parts results in greater waste of resource. Therefore, research on the integration of multi-source and heterogeneous spatial data and providing data sharing service, based on the internet of things, are of great importance to the rational utilization of the existing spatial data resource, reducing repeated investment and supporting the sustainable development of spatial information industry.

Strong ability of collection and storage of information and massive data resource of the Internet of things facilitates the growth of spatial data management technology. Firstly,
powerful object awareness ability of the Internet of things front-end provides the innovative processing model for spatial information application. Secondly, transparent resource accessing characteristic of the Internet of things back-ends offers a unified platform with computing, storage and access for the spatial information application. Although, Internet of things provides an unprecedented opportunity to the spatial data acquisition and management, we must realize clearly the challenges which the entire new framework poses. As an infrastructure, sensors on the internet of things lack effective management, programming, scheduling, collaboration and sharing mechanism, without collaborative incident observing and correlative holistic factors in incident detecting, which only observe incident but not analyze effectively, verify authenticity and predict change tendency for the dynamically changing phenomenon. It is a major challenge facing the spatial information service in Internet of Things environment.

Therefore, how to search the corresponding modeling method and integration algorithms in the spatial data model level, with considering of the technology factors of spatial data integration? Generating the semantic information with medium granularity, flexible organization and appropriate expression, it can process the magnitude data and the complex large-scale anonymous application demands immediately, which has customizable pattern recognition ability. On this basis, scale magnitude multi-source heterogeneous spatial data is collected and managed by the Internet of things platform, according to the space information service demand for smart city. And it supports the multidimensional spatial data integration in data model levels, and supports the fusion-storage of different types of spatial data including the vector data, grinned data, text and real-time multimedia data, with the purpose of improving the current service ability of digital city and providing the technical storage support for smart city system facing the future.

2. Related Work

2.1. Modeling and Integration of Spatial Data

Different adopted extraction views and modeling methods of spatial data have resulted in the difference of concept extraction structure of spatial data models among different industries, which reflect the specific spatial database mode. Typical two-dimensional spatial data model consists of the entity data model that based on points, lines and planes and gridded data model based on continuous blanket (Chen Jun, 1995) [1]; Coverage data model (ESRI, 1981) and Geodatabase use object-oriented data model (ESRI, 2002); Vector gridded integration and object-oriented spatial data model (Gong Jianya, 1992, 1993) [2-3]; About 3d, Chen Jun (1998) [4], Zlatanova (2004) [5] and Coors (2003) performed the study of 3d vector data model separately; Li Qingquan and Li Deren et al. (1997) [6] investigated the data structure and coding algorithm of Octree model systematically; Pilout and Li Qingquan et al. carried on the model research of Tetrahedral Network (Pilout, 1994; LiQingQuan, 1998) [7], and Wu Lixin, Qi Anwen and ChengPengGen et al. conducted the analogical/quasi tri-prism volume research respectively (Wu Lixin, 2002; Qi Anwen, 2002; ChengPengGen, 2005, [8]).

Having settled the data format problem, the data integrated with different spatial data set needs to solve the different spatial data model problem and different logical structure of the data model problem. Gong Jianya (1993) [2] proposed a transformation and integration method of the two dimensional vector gridded data; Monika Sester and Karl Heinrich focused on the integration of different data sets and the generation of new dataset from the integration and clustering of spatial object angle, and gave us the corresponding processing component [9]; Catriel Beeri (2005) [10] et al., performed the study of searching and coupling algorithm of the corresponding spatial object among three or even multiple data sets. Gong Jianya (1993
et al., investigated the two-dimensional vector gridded integrated data model and the transformation of vector gridded data. All of these researches are rarely studied the transformation and comprehensive integration of the spatial data model in different perspectives from a total dimension angle.

The research on spatial database semantic consists of classification and coding system, geographic information ontology research and mapping research of different semantic rules. He Jianbang (2004) proposed the classification and coding system of resources and environment information, and suggested that it should be divided into six levels including the category, group, type, primary, secondary and tertiary, and viewed the geographic information ontology as coding tool in different areas. Y. Bishr pointed to the semantic heterogeneity, pattern heterogeneity and grammar heterogeneity are the reason for GIS interoperability barrier, and solving the semantic heterogeneity is the key to tackling the other two problems, adding the semantic expression on the basis of formalization data structure (FDS) proposed by Molenaar. Huang Yuxia and Chen Changsong et al have performed related research from semantic sharing and interoperation angle. Michael Lutz (2003) has put forward a classification framework to solve the semantic interoperability problems with geographic information network service. Villie Morocho (2003) proposed a prototype system of multi-source data semantic integration with federal architecture, and used ratio model to evaluate the semantic similarity degree; Jorge Campos and Max J. Egenhofer (2003) performed a new semantic model for the virtual reality object classification, and described the semantic of object existence, visibility and behavior semantic. I Budak Arpinar (2006) analyzed the development and semantic of earth space information ontology, and presented a method to research the data integration combined the thematic ontology, space ontology and space-time ontology.

Although the current spatial data model solves the related problems of spatial data modeling in different domains, it is not well done in comprehensive modeling, total dimension analysis and transformation and seamless integration of the spatial information. The data representation method based on GML provides an optional scheme for the expression, transmission and sharing of geographic data. However, the simple feature specification of OGC is hard to meet the actual data modeling need of a larger space information system, and GML also has some defects in the efficiency of the spatial data organization and supporting spatial analysis. The spatial data exchange specification and data format conversion tool, such as SEDRIS and FME, provide an effective technical support for the spatial data sharing as a data exchange center at a data exchange level, but they can not perform services directly to the seamless spatial data integration as a spatial data model. The space information system that needs to process all essential factor's data, including ocean, atmosphere, terrain, underground and watershed, is desperate for a more powerful spatial data model to support the efficient modeling and seamless integration of multi-source polymorphous spatial data of vector graphics, gridded images and video.

2.2. Spatial Data Model of Internet of Things

The process of movement and interaction of instrumented objects in instrumental space has the characteristics of dynamic and issuance. In order to perceive and manage them efficiently, the effective data model is required for managing. Diversiform storage format and lacking metadata or conceptual model of the current common spatiotemporal database will lead to the absence of the sectional semantic information involved in geographic information. The research on Spatial database semantic consists of classification and coding system, geographic information ontology research and mapping research of different semantic rules. He Jianbang (2004) proposed the classification and coding system of resources and
environment information, and suggested that it should be divided into six levels including the category, group, type, primary, secondary and tertiary, and viewed the geographic information ontology as coding tool in different areas. Y. Bush pointed to the semantic heterogeneity, pattern heterogeneity and grammar heterogeneity are the reason for GIS interoperability barrier, and solving the semantic heterogeneity is the key to tackling the other two problems, adding the semantic expression on the basis of formalization data structure (FDS) proposed by Molenaar. Huang Yuxia and Chen Changsong et al have performed related research from semantic sharing and interoperation angle. Michael Lutz (2003) has put forward a classification framework to solve the semantic interoperability problems in geographic information network service. Villie Morocho (2003) proposed a prototype system of multi-source data semantic integration with federal architecture, and used ratio model to evaluate the semantic similarity degree [18]; Jorge Campos and Max J. Egenhofer (2003) performed a new semantic model for the virtual reality object classification, and described the semantic of object existence, visibility and behavior semantic [19]. I Budak Arpinar (2006) analyzed the development and semantic of earth space information ontology [20], and presented a method to research the data integration combined the thematic ontology, space ontology and space-time ontology.

Ultimately, how to bridge the semantic gap is the greatest challenge to the integration of heterogeneous spatial data in the Internet of Things. The intelligent processing method based on data driven has become one of the most promising solved approach. However, there is another following problem. How to deal with magnitude data and uncertain unknown application, how to make the intelligent processing of spatial data have a scalable pattern recognition ability? The machine learning method based on attribute provides a new way. Different with the machine learning method of traditional pattern classification, machine learning method based on the attribute is used in classification and learning to obtain the middle-level attribute and to achieve the semantic recognition based on the mapping between middle-level attribute and high-level semantic concept. Recently, lots of applications based on attribute learning method are emerging out constantly. B. Siddiquie et al. proposed an image retrieval and sorting method based on the multiple attributes query to improve the query accuracy, with considering of the relationship among query attributes and that between query attribute and the attribute appeared in codebook but not in query; J. Liu had obtained a good classification result of human behavior recognition using the artificial calibration attributes and the mapping attributes with data driven method, and solved the zero-shot classification problem. The machine learning method based on the attribute is a new research hotspot that deserves further study in spatial data field, especially in the intelligent processing field of location-based multimedia data.

Although the current spatial data model solves the related problems of spatial data modeling in different domains, it is not well done in comprehensive modeling, total dimension analysis and transformation and seamless integration of the spatial information. The data representation method based on GML provides an optional scheme for the expression, transmission and sharing of geographic data. However, the simple feature specification of OGC is hard to meet the actual data modeling need of large space information system, and GML also has some defects in the efficiency of spatial data organization and supporting spatial analysis [21]. The spatial data exchange specification and data format conversion tool, such as SEDRIS and FME, provide an effective technical support for the spatial data sharing as a data exchange center at data exchange level, but they can not perform services directly to the seamless spatial data integration as a spatial data model.
3. Problem Description and Modeling

The abstract process of real world phenomena into computer data has divided by OpenGIS into nine levels, namely real world $\rightarrow$ conceptual world $\rightarrow$ geospatial world $\rightarrow$ dimensional world $\rightarrow$ project world $\rightarrow$ points world $\rightarrow$ geometry world $\rightarrow$ feature world $\rightarrow$ feature collection world. From the above process analysis, we can see that different emphasis between different subjects result in all aspects of disparity in the process of front to back abstraction. Abstraction differences in the first four levels are the natural language, geography, language, scale and dimension difference respectively, and are mainly the model conceptual semantic difference when performed in the data model. From the dimensional world, later abstraction differences displays in the heterogeneous characteristics of coordinate reference, feature type, relation type and data structure.

Therefore, we define the integration model of multidimensional spatial data as that, it offers multiple rules for the type selection and constraint matching from different aspects including the semantic coding type of spatial data, and allows user to customize the spatial data integration model that fits into the set rules and constraints within bounds of the above type collections.

According to the above definition, summary modeling architecture of spatial data integration application using the customizable spatial data integration model is shown in Figure 1. The customizable spatial data integration model is the core of level system which based on the relation feature type collections, constraint rule knowledge base and spatial feature type collections, and is fused or organized with object-oriented method and open architecture through supporting the modeling platform of customizable spatial data integration.

![Figure 1. Summary Modeling Architecture of the Customizable Spatial Data Integration](image)

4. Framework Design

Following the hierarchical principle, the integration and service framework of multidimensional spatial data is divided into a three layer architecture, which consists of object sensing layer, information integration layer, application service layer, shown in Figure 2. Object sensing layer achieves the perception and data acquisition of spatial object; Information integration layer provides the processing of restructuring, cleaning and fusion for
uncertain information acquired by network, and integrates for the relative accurate conclusion; Application service layer supplies the service that transforms information into content.

![Diagram of Multi-Sensor Spatial Data Fusion](image)

**Figure 2. The Framework to Multi-Sensor Spatial Data Fusion**

### 4.1. Object Sensing Layer

With the sensors mounted on the instrumental objects, Internet of Things dynamically acquires the information in real time, and provides services on the basis of the interaction between individuals and groups. Change rule of moving objects that loading equipment, the interaction among individuals and the perception of the Internet of Things environment using perception meaning provide the basis for the application of spatial internet of things information. The paper has decrypted it from data level and model level.

#### 4.1.1. Spatial Data Fusion from Data Level

The key of data level spatial data fusion is real-time processing and intelligent analysis of large spatial datasets. However, people always try to find the one-step processing and analysis method, which means that the concept classification with high-level semantics can be obtained from the entered heterogeneous multi-dimensional spatial data using intelligent processing method directly, and the final transmissible information can be formed by the entered sensor data directly. But the manner that tight coupling of a particular application lack in the poor adaptability. The reality has proven that the one-step processing manner is difficult to adapt the complex application demand and the dynamic application environment, especially for the massive sensor data of the internet of things. How to process and analyze the data immediately and following manipulate it, which cannot meet the complex demands of large-scale concurrent users. In most application places, the aim of information perception is to obtain some event information or semantic information and not all of the sensor data. Therefore, there is no need to transmit all sensor data to the sink node in most cases, but the information observer interested. It concludes that, adopting the advanced learning and spatial data expression based on the middle-layer attribute, the scalable spatial data stream can be achieved through simple processing rapidly, which also adapt to the demands of cloud computing, terminal device and complex applications.
To adapt to the heterogeneous network, multiple terminal device and complex spatial data application in the internet of things, the composition of middle-layer attribute should be confirmed. Middle-layer attribute is consisted of various descriptive information, which could be different granularity. And the massive redundancy exists in the description information and between different descriptions. Our study has tried to remove the data redundancy, and to find an appropriate structure to organize these description information, with the purpose of simplification and structuralization of the stream information. The middle-layer attribute is consist of various description information, which could be different granularity, such as video stream, which could be the whole sequences level, frames level, macroblocks level, and even pixel level. And the massive redundancy exists in the description information and between different descriptions. In view of the complexity and high parallel characteristics of spatial information, it design a MapReduce method that fitting the extraction of middle-layer attributes using the usual parallel programming models MapReduce.

Map method: each analysis task is distributed to every codestream parsing server. The output result is the optimal model and correlation information of each subset.

Reduce method: the output information from all of the codestream parsing server is entered to the information predicting model, while the redundancy between different code streams is reduced. For the correlation information organized by themes, all output correlation information is entered to the correlation information quantization module, while the information amount within the same pattern and the internal redundancy are all reduced.

In addition, MapReduce that taking batch processing does not meet the real-time processing demands of data stream in internet of things. We attempt to apply the method proposed in document [22], and the work process is shown as follows.

1) Cache the intermediate results. The intermediate result generated by historical data is preprocessed, and cached to the local storage of each task node distributively.

2) Local phased process. Map phase groups the data stream with Hash function, partitions and filters the data that related to the local intermediate result, which is asynchronously delivered to the Reduce phase and compute with the intermediate result.

3) Data synchronism. The result computed in local is stored synchronously in a distributed environment. Among them, each node maintains the cache of the local intermediate result and phased pipeline, while control node is responsible for the reliability and scalability of tasks and the cycle management.

4.1.2. Spatial Data Fusion from Model Level: The modeling of the geometric figure element and relation element is the foundation and the key of the customizable data model. During the analysis of every spatial database model, the preserve degree of original geometric data and topology relationship of each spatial dataset depends on the granularity and the span of the corresponding element types provided by the low-level geometric figure element and relation element of the customizable data modeling. The differences of geometric figure element and topology structure types of every spatial data model, including the simplex, replication, manifold and non-manifold, is directly proportional to the span of the element type. The boundless element type set is bound to increase the complexity of top spatial data operation algorithm infinitely. Therefore, only extraction of the key geometric spatial and topology element modeling set and providing the extended modeling set interface with controllable complexity, that can solve the contradictions between the span of element type and the complexity of top operation algorithm.

In addition, semantic mapping and matching rule of spatial database in different industries is one of the import parts of the seamless spatial data integration. The spatial data integration
model must be integrated the particular semantic matching method and rule between different databases of industry or department, and matched and integrated with the spatial data model. Owing to the absence of the accepted code classification and mapping standard, the semantic matching rule and result are fuzzy and uncertainty. To void the structural influence between integrated data model and the uncertainty search result of current semantic mapping and classification, semantic mapping modeling is only moderately extracted and integrated using a particular method to ensure the effective semantic integration and to meet the possible future changes of semantic classification method and mapping rule.

4.2. Information Integration Layer

In order to improve the accuracy of the information integration, it is necessary to fuse the multi-source information. In terms of network, fusion is the information interaction based on heterogeneous network elements. Set out from the incompleteness and the redundancy of the information gained by internet of things, it adopts multi-stage and multi-level fusion strategy for the relevant information that existing in multiple heterogeneous network elements. In view of the time sequence of environmental perception in the internet of things, the organic fusion is carried out for the information integration at different levels based on the multi-source characteristic expression collection using the sequence analysis and Bayesian network fusion method.

From the data fusion, the corresponding basic graphic elements are top-down analyzed from different dimensions for the spatial entity. The real spatial entity is abstracted to 0, 1, 2, 3 dimension space from various industry perspectives. Based on the simplex and complex theory of manifold and non-manifold topology of regular entity, it defines and analyzes the abstract unit granularity of graphic element. Compared with the basic geometric object model of OGC, 13 kinds of geometric graphic elements of SDTS standard and China CNSDTF standard, it extracts the core element type collection. Meanwhile, using the finite state machine of knowledge inference and regular expression technique, the various classifications and mapping transforming the rules of semantic are deduced and integrated into the spatial data model restriction based on the integration method that consistent with other constraint rules.

In basis of this, using the latent intelligent processing that an intelligent processing method of massive heterogeneous data, multiple middle-layer attributes of heterogeneous spatial data is attained in advance which viewed as the middle-layer semantic description of spatial data. And the high-level semantic applied in the specific application can be obtained fleetly with the mapping between middle-layer attribute space and high-level semantic space. The intelligent processing method that latent learn in advance can adapt to the dynamic, diverse and varying application demands of massive spatial data with the scalable recognition ability.

4.3. Application Service Layer

In the dynamic acquisition and evolution research of the service demands of the Internet of Things, the dynamic environment that service depends on and evolution characteristics are clearly defined and modeling, combining with the service demand acquisition method that facing object ontology and skin in multi-view Internet of Things. On that basis, the control function and reflection mechanism of context awareness environmental information are introduced to adapt to the dynamic changes of service demands, and update the metadata to make service evolution dynamic.

In the research of dynamic service construction and self-evolution adapted mechanism, the Internet of Things service is classified and organized under the new concept of service clustering. Then the introduced concept of the abstract service domain can organize the
network information service in different service clusters into a service resources view with the user functional and non-functional requirement. Based on the problem domain search and knowledge reasoning, the dynamic discovery, selection, clustering and integration of service is achieved with the bionic intelligent computation method. In the study of dynamic safeguard method of service quality, machine learning method is used to analyze and mine the dependence or causal relationship between context information dynamically, and to generate the topological structure of context information and the measurement value of nodes. Finally, the context-aware dynamic safeguard mechanism and method of service quality is established for the guaranteed service quality of Internet of Things.

Specifically, the service demand expression consists of user demand expression, application demand expression and the demand expression of multi-user multi-application concurrent environment. Using the metadata description mechanism in the CDMI standard of cloud data management interface, the description interface is established in user demand expression, considering the description of incremental service quality, such as access bandwidth and response time. The demand expression of multi-users, multi-application concurrent environment is complex, and the storage demand is not a sample linear accumulation of multiple user demands which needs further study. Based on the application feature extraction, that is the research of interaction and mutual influence among multi-users and multi-applications, and the research of expression modeling and method of storage service demand in a complex environment for the demand description accuracy.

Based on the service demand expression, the cost model that involves the space and time cost is established combined with the data layout and organization in a large-scale network storage system, with considering of the data characteristic, user demand and system environment. According to the user service demand, multiple target optimization model is set up including the performance, energy consumption and safety target. It guides the self optimization of storage system and controls and manages the data selectively, such as automatic data transfer and multi-copy management, etc.

5. Case Study

5.1. Data Preparation

The 2-dimensional video image, 2-dimensional vectorgraph, terrain surface TIN data, 3-dimensional body metadata of underground rock and 3D landscape model data that in the same region is taken as experimental sample dataset. Among them, the format of video data is GIF, while 2-dimensional vectorgraph is DWG and terrain surface TIN data is TIN, shown as Figure 3.

![Figure 3. Data preparation](image-url)
5.2. Technical Process

According to the above definition, the summary modeling architecture of spatial data integration application using the customizable spatial data integration model is shown in Figure 3. The customizable spatial data integration model is the core of level system which based on the relation feature type collections, constraint rule knowledge base and spatial feature type collections, and is fused or organized with object-oriented method and open architecture through supporting the modeling platform of customizable spatial data integration.

Figure 4. Technical process of the customizable spatial data integration

In the process of data integration, central server or proxy server that deployed the semantic parsing component conducts the further query optimization and decomposition based on the matching results after the semantic mapping completed. And intelligent agent operates the different data sources according to the scheduled process, and integrates the sub-query results.

5.3. Experimental Results

5.3.1. Storage Cost: Under the condition of different scale, average storage cost of two ways is compared respectively, and the result is shown in figure 5.
5.3.2. Multi-dimensional Spatial Query: On convenient comparison, the experiment is based on sample dataset with the comparison of expression and integration of all dimensions spatial data in R3. The results of spatial query are shown in Figure 6. Based on studies with alternative, choice, parallel, level, and other methods for multi-dimensional spatial data index organization, reverse transcription polymerase chain reaction from dissected neural tissues, the framework can support expression and data integration of the existing R3 space in all dimensions.

![Figure 6. Compare results of multi-dimensional query](image)

5.3.3. Integrity of Integration: As we know, spatial data sets corresponding to the same reality entities differ in dimensions, project, coordinate reference systems, spatial location, and coordinate precision. In addition, combining spatial data integrity, consistency constraints, as well as between the different data elements alternative, parallel, level of detail, organizational methods, addressing data instance layer position matching precision matching scale matching problems. Figure 7 shows simulation of soil erosion by reason of rainfall from...
various sensor data and basic geographic information. The research has eliminated the differences and conflicts from spatial data sets.

6. Conclusion

This paper presented a framework for seamlessly adapting distributed applications to spatial data integration. The framework presents two main advantages over the existing solution: As a spatial data model platform, it provides the primitives and relationships that constitute the various spatial data model element, and the organization of these elements. In addition, it proposed semantic unrelated / related spatial data integration modeling techniques, such semantic study results do not reflect the uncertainty of spatial data integration model.

As a future work, we will propose the enhancement of multi-dimensional spatial model, and the fusion of more types of multimedia data to deal with the underlying information. In addition, large amounts of spatial data should be conducted to set various parameters, such as thresholds and fusion weights.

Acknowledgements

The work is supported by the National Nature Science Foundation of China (41101432, 41201378), the Natural Science Foundation Project of Chongqing (cstc2012jjA40014), scientific and Technological Research Program of Chongqing Municipal Education Commission (KJ130532).

References


Authors

JiangFan Feng

He received his B.S. degree from Southwest Agricultural University, and his Ph.D. degree from Nanjing Normal University, in 2002 and 2007. He works as associate professor of Chongqing University of Posts and Telecommunications. His main research area includes spatial information integration and multimedia geographical information system.
Wei Wei

He received the B.S. degree in computer science and technology from Southwest University, in 2011. He is currently working towards his M.S. in the Chongqing University of Posts and Telecommunications. His current research interests include search engine and location based service.