AN ITERATIVE PROCESS FOR MATCHING NETWORK DATA SETS WITH DIFFERENT LEVEL OF DETAIL

Yoonsik Bang, Chillo Ga and Kiyun Yu*

Dept. of Civil and Environmental Engineering, Seoul National University, 599 Gwanak-ro, Gwanak-gu, Seoul, Korea – (bangys1004, crow75, kiyun)@snu.ac.kr

Commission IV, WG IV/4

KEY WORDS: Network Matching, Level of Detail (LOD), Iterative Process

ABSTRACT:

Nowadays, car navigation systems are widely used in many countries, which in turn has given rise to the production of many road network data sets. In Korea, there are two principal road networks—the Navi-Network produced by navigation systems companies; and the Traffic-Network maintained by the government for traffic information services. In order for the users of the Navi-Network to have access to the traffic information reported by the Traffic-Network, the two systems must be compatible with each other. However, the procedures to enhance compatibility are not particularly easy to carry out because each system is produced for different purposes and has a different level of detail (LOD). In this paper, an iterative process was proposed to match the nodes and links of the two road network data sets with different LOD. We first found the ‘node matching pairs’ based on their locations and the shapes of the links connected to them. We then found the ‘link matching pairs’ using our findings on the node matching. Next, considering the topological relationships of the nodes as delineated from the previous step, we matched the previously unmatched links and nodes in turn. This step was performed iteratively, and at the end of every stage of iteration, similarity values between the two matched datasets were computed. When we discovered the stage with the best similarity value, the results of that stage were regarded as the most appropriate. Finally, the proposed process was applied to the real road network datasets and the results were analyzed.

1. INTRODUCTION

1.1 Motivation

In GIS, the different ways of describing features in the real world are determined according to the types of features. In case of a road, it can be described as a polygon with a value for width reflecting the shape of the real feature. But when it comes to road maps used to represent conditions or attributes of road objects, the network data structure which is comprised of 1-dimensional line objects and 0-dimensional point objects can be efficiently used to draw the road features. Generally in the network data structure, intersections or some specific points with special attributes are represented by ‘points’ and road sections between two intersections are represented by ‘polylines’. A ‘Node’ and a ‘link’ (or, ‘road’, ‘line’) mean a point and a polyline object, respectively. (Researchers dealing with road network datasets use different terminologies.)

The largest representation of users for road network datasets is from car navigation service providers. They show their users the location of the vehicle by mapping it onto the road network map. They also find and show the most efficient path to travel between two points on the road network. In addition to navigation services, some attributes of the road can be provided through the navigation device. For example, real-time traffic information can be provided. To implement this, the road network datasets for the traffic information and for the navigation service must be matched to each other at the object level. In other words, a correspondence between objects in the two different datasets must be found. Once we get a matching table containing this information, the two different network datasets can be linked to each other and applied to the various fields of GIS.

This paper proposes a method for matching objects in two road network datasets. It was designed to make the best use of topological relationships, and it can be efficiently applied if the two network datasets show large different levels of detail (LOD). An iterative process was developed and applied to this method to achieve the best performance.

1.2 Related Works

Many attempts have been made to match and relate network datasets. They were developed based on different datasets with different characteristics.

Walter and Fritsch (1999) proposed a basic method to integrate two different spatial datasets. Their method was based on the geometrical structure of spatial objects and not the node-link structure. It relied on statistical analysis rather than computational geometry. Most importantly, their conclusion was focused on data conflation using the result of matching.

Xiong (2000) proposed a method that improves matching accuracy by combining bottom-up and top-down subprocesses. The bottom-up process is done first, and the top-down process is subsequently done in the opposite direction. This method helps to correct irregular inconsistencies between the two datasets taking into consideration the topological relationship of the objects in one dataset. Mustiére and Devogele (2008) developed a systematic network matching process and applied it to datasets with different levels of detail.

* Corresponding Author.

Due to advances in techniques for data acquisition, the spatial and geometrical accuracy of data has been achieved at a reasonable level. On the other hand as many specific network datasets have been produced for different purposes, they usually show different levels of detail. For this reason, researchers from the latest studies are focusing more on matching between datasets with different levels of detail.

2. ROAD NETWORK DATA

2.1 Node-Link Data Structure

Road network data are comprised of 'nodes' representing intersections or some special points and 'links' representing road sections between two intersections. Node is the point object which has some geometric attributes (x-y coordinates) and some semantic attributes (e.g. NAME, ID, TURN_INFO, LINK_IDs etc.). Link is the polyline object which has some geometric attributes (X-Y coordinates) and some semantic attributes (e.g. NAME, ID, CLASS, ST_NODE_ID, ED_NODE_ID etc.) The road network data used in this paper were established following the form of a shapefile comprised of 'nodes' and 'links'. A data structure of a road network is shown in Figure 1.

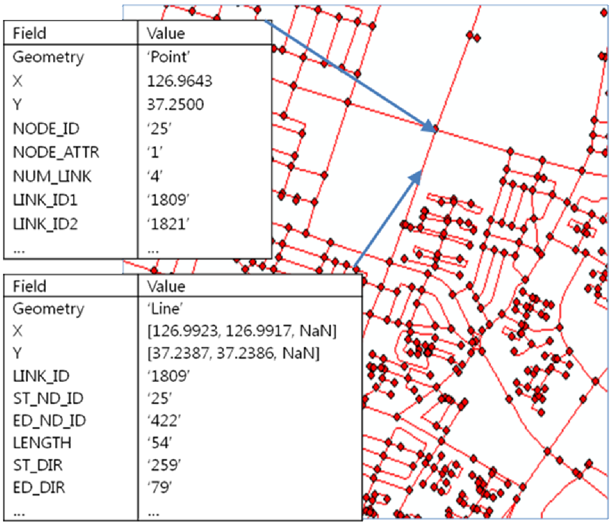


Figure 1. Node-link data structure of a road network dataset

2.2 Level of Detail

Road network datasets are produced in various forms according to their purpose. The most remarkable characteristic identifying the datasets is the 'level of detail (LOD)'. A high-level dataset even contains small roads such as narrow alleys. A low-level dataset, on the other hand, contains only relatively wide or important roads. The producer of the datasets determines the appropriate level of detail according to its purpose.

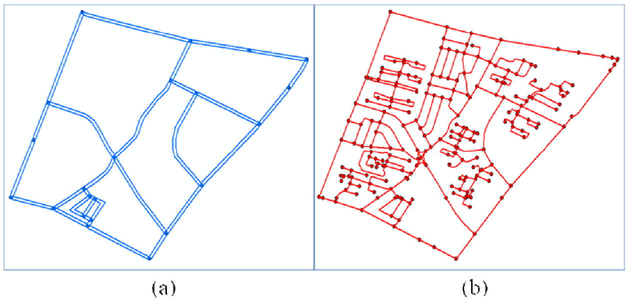


Figure 2. Two sample datasets with different levels of detail (datasets with (a) a low LOD and (b) a high LOD)

In Korea, there are two principal road networks: the Navi-Network produced by navigation system companies, and the Traffic-Network maintained by the government for traffic information services. In order for users of the Navi-Network to have access to the traffic information reported by the Traffic-Network, the two systems must be compatible with each other. However, the procedures to enhance compatibility are not particularly easy to carry out because each system is produced for different purposes and has a different level of detail (LOD). In this paper setting the Traffic-Network as the reference data (Net1) and the Navi-Network as the target data (Net2), the matching process was applied.

3. NETWORK MATCHING PROCESS

3.1 Basic Matching Before the Iterative Process

Generally, a matching algorithm starts by finding matching pairs of nodes from two network datasets. First using the locations of the nodes and the shape information of the links connected to them, node matching pairs are found. Link matching pairs are found via node matching. The initial parameters of these processes can be determined by a simple statistical analysis. A set of processes is termed 'Basic Matching'. It consists of three subprocesses shown in Table 1.

Subprocess	Description	parameters
Node pre-matching	To find the relation between the node sets, first find the matching pairs using the distance between them is within a given critical value.	Critical values of distance between nodes
Node matching by ADG	For unmatched modes, by comparing the shapes of angle distribution graphs of nodes, some more node matching pairs are added the node matching table.	Critical values of ADG difference
Link matching by connectivity searching	When the start node and the end node were matched, using the connectivity of two nodes, link matching pairs are found.	Critical values of Hausdorff distance between segments

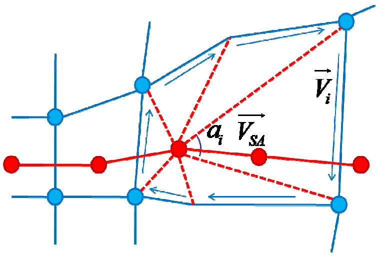
Table 1. Subprocesses of 'Basic matching' and their description

The goal of basic matching is to obtain a matching table which is needed to perform the iterative process of the next section. So the strictest parameters have to be applied to extract only the obvious matching pairs. We get an initial matching table through basic matching. Some of the remaining blanks of the table will be filled in as the iterative process progresses.

3.2 Subprocesses for the Iteration Stage

An additional matching process is needed to fill in the blanks of the matching table. It is comprised of the following subprocesses.

3.2.1 Similarity Measurement Using Average Influence Vector: To determine whether an iteration is repeated and has its parameters adjusted appropriately, it is necessary to evaluate how the two matched datasets are geometrically similar to each other. In this paper, a method of measuring similarity using 'average influence vector' was developed and applied. The average influence vector is a measurement representing the geometrical influence of objects in the network dataset at a point on a plane which contains the dataset. To measure the similarity, an evaluation of how a Net2' link vector is similar to the average influence vector at every point on Net2' must be done. First, the difference between the link vector and the average influence vector is calculated at every point on Net2'. The weighted mean of these results is the result of the similarity assessment. (Bang et al., 2009)



$$\vec{C}_A = \frac{1}{2\pi}\sum_{i=1}^{n} a_i \vec{V}_i(l,\theta)$$

$$S = \sum_{segments} \frac{1}{l_A}\left(\vec{C}_A - \vec{V}_{SA}\right)$$

Figure 3. Similarity measurement between two matched datasets

3.2.2 Link Matching by Buffer Clipping: After the similarity assessment, for Net1 links which were not matched at the previous step, buffers are generated from them. Then a completely connected subset of Net2 links can be extracted by the buffer. IDs of extracted Net2 links are stored in the link matching table. Size of the buffer is increased by a certain rate as the iteration progresses.

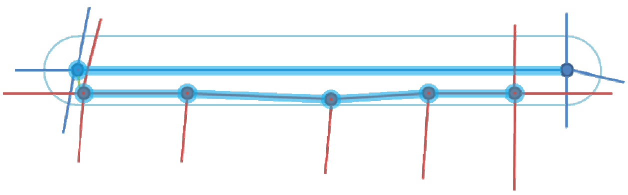


Figure 4. Link matching by buffer clipping (blue: reference [Net1], red: target [Net2])

3.2.3 Node Matching by Connectivity Searching: The start and the end nodes of the subsets matched at step of section 3.2.2 are stored in the node matching table. There's no need to apply the special algorithm in this step.

3.3 The Iterative Process

As mentioned earlier, an additional matching process is needed to fill in the blanks of the matching table. An iterative process was developed to solve this problem. In every stage of the iteration, two subprocesses of section 4.2 (link matching by buffer clipping and node matching by connectivity searching) are performed and the similarity of two matched datasets is calculated. A result table of the step with the best performance becomes the final matching result. To measure the performance, a similarity assessment method is used for the two matched network datasets. A flowchart of the entire matching process is shown in Figure 5.

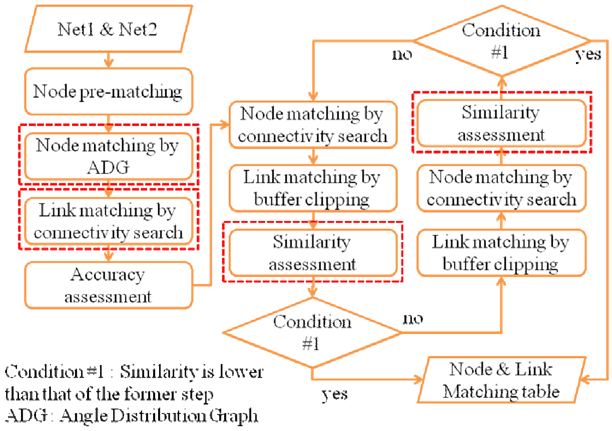


Figure 5. A flowchart for the iterative matching process

For two road network datasets with different level of detail, namely Net1(reference, low LOD) and Net2(target, high LOD), the basic matching table is obtained using the basic matching process. Similarity is then measured for pairs of objects extracted in this process. Next, using 'link matching by buffer clipping' and 'node matching by connectivity searching', the additional matching pairs are found and saved to the matching table. And the similarity is measured again. If the similarity gets better than that of the former stage, the additional matching processes are repeated once more. If not, it means that the matching result of the former stage is the best solution of the matching problem.

4. EXPERIMENTS

4.1 Test Data

For the two road network datasets mentioned above, Traffic-Network and Navi-Network, matching was done by the proposed methodology. ITS (Intelligent Transportation System) Standard node-link dataset which is utilized in Korea was used as the Traffic-Network, and SK Entrack dataset which is a base map in a navigation system was used as the Navi-Network. Some pieces of road map of Suwon-si, Kyunggi-do were extracted for use.

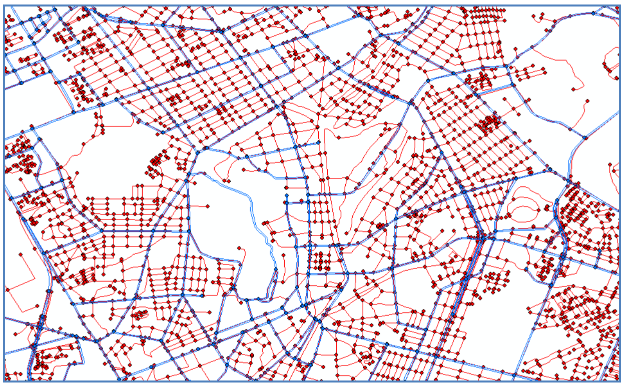


Figure 6. Used dataset (blue: ITS data [Net1], red: SK Entrack data [Net2])

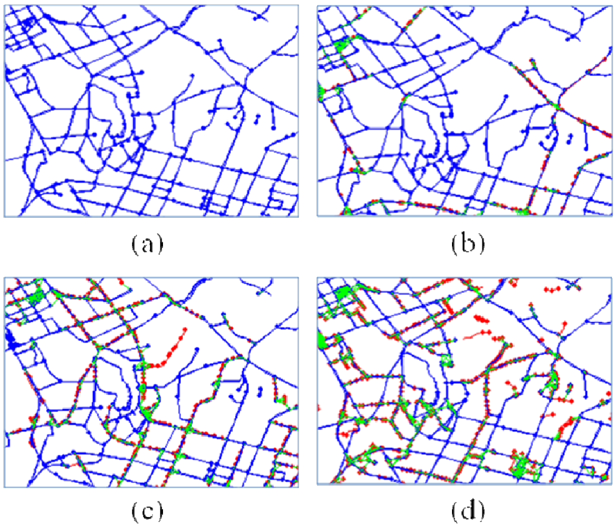
4.2 Performance of the Matching Process

The iterative process proposed in this paper was applied to these network datasets. 'Link dissimilarity' of matched links and 'average distance' of matched node pairs after every stage of iteration. Then the weighted sum of these measures was calculated by the ratio between the total numbers of nodes and links as the weight. This value means the total dissimilarity of two matched network datasets.

stage	# of matched nodes	# of matched links	link dis-similarity (m)	average distance of matched node pairs (node dis-similarity) (m)	weighted sum of dis-similarity (m)
basic	721	2298	21.2871	0.744951	21.54214943
1st iteration	744	2301	14.9225	4.139279	15.06177935
2nd iteration	749	2303	12.199	4.392181	12.25911811
3rd iteration	768	2308	9.3071	4.675256	9.982356234
4th iteration	769	2309	9.1834	4.781891	9.965291438
5th iteration	769	2309	7.9233	4.92883	9.233666861
6th iteration	769	2310	7.37714	4.684585	9.061725429
7th iteration	770	2310	7.37714	4.891072	9.268212425

Table 2. Result of the iterative matching process

As you can see in table 2, the similarity between matched datasets was increased (dissimilarity was decreased) as the iteration progresses. In detail, the link dissimilarity was decreased steadily. And the node dissimilarity was converged to about 4.7m. We can see the tendency for matching of two nodes located further away as the iteration progresses. The similarity of the 7th iteration went down compared with that of the 6th iteration. Therefore, the matching result of the 6th stage of iteration can be considered the most reliable result.



5. CONCLUSIONS

In this paper, an iterative matching process was proposed to match two network datasets. For an appropriate determination of whether an iteration step must be repeated, the similarity between the two datasets was measured. The average influence vector was used for measuring similarity. The process was done iteratively and the similarity was measured after every step. At the step when the highest similarity value is obtained, the matching table of that iteration step becomes the final result.

The proposed algorithm was applied to a matching process between the two road networks mentioned above and the overall similarity was measured. The process was then performed iteratively using the similarity values. The degree of similarity was acceptable and was improved by the steps of the process.

Using the methodology of this paper, the integration of network data is simplified. Specifically, traffic information and a number of related attributes can be represented on a navigation map automatically. The full automation of the algorithm and more accurate results need to be done to reflect many more examples of actual roads.

References from Journals:

Mustiére, S. and Devogele, T., 2008. Matching Networks with Different Levels of Detail, *GeoInformatica, 12*(4), pp. 435-453.

Walter, V. and Fritsch, D., 1999. Matching spatial data sets: a statistical approach, *International Journal of Geographical Information Science, 13*(5):445–473.

Xiong, D., 2000. A three-stage Computational Approach to Network Matching, *Transportation Research Part C 8*, 71–89.

Xiong, D. and Sperling, J., 2004. Semiautomated matching for network database integration. *ISPRS Journal of Photogrammetry & Remote Sensing 59*, pp. 35– 46.

Zhang, M. & Meng, L., 2007. An iterative road-matching approach for the integration of postal data, *Computers, Environment and Urban Systems, 31*(5), pp. 597-615.

Zhang, M. and L. Meng, 2008. Delimited Stroke Oriented Algorithm: Working principle and implementation for the matching of road networks. *Geographic Information Sciences, Hong Kong, 14*(1), pp. 44-53.

References from Other Literature:
Bang, Y., Lee, J., Kim, J. and Yu, K., 2009. An Iterative Algorithm for Matching Two Road Network Data Sets, in *Proceedings of ASPRS-MAPPS*, San Antonio, Texas, USA.

Lüscher, P., Burghardt, D., and Weibel, R., 2007. Matching road data of scales with an order of magnitude difference, in *Proceedings of XXIII International Cartographic Conference*, Moscow, Russia.

Volz, S., 2006. An iterative approach for matching multiple representations of street data, in *Proceedings of the ISPRS workshop on Multiple Representation and Interoperability of Spatial Data*, pp. 101–110.

Zhang, M. and Meng, L., 2006. Implementation of a generic road-matching approach for the integration of postal data. In *Proceedings of the 1st ICA Workshop on Geospatial Analysis and Modeling*, Vienna, Austria, pp. 141-154.

Zhang, M., Meng, L. and Qian, H., 2007. A structure-oriented matching approach for the integration of different road networks. In *Proceedings of the XXIII International Cartographic Conference (ICC), Incremental Updating and Versioning of Spatial Data*, Moscow, Russia.

6. ACKNOWLEDGEMENTS

This research was supported by a grant(07KLSGC04) from Cutting-edge Urban Development - Korean Land Spatialization Research Project funded by Ministry of Construction & Transportation of Korean government.

In addition, the authors wish to acknowledge the financial support by SNU SIR BK21 research Program funded by Ministry of Education, Science and Technology.