

단순화 관계를 갖는 네트워크간의 인스턴스 동일성 확인

권용식, 차상균
서울대학교 전기공학부

Instance Identification between a Network and its Simplification

Yong Sik Kwon and Sang Kyun Cha
School of Electrical Engineering, Seoul National University

요 약

지리 정보 서비스에 대한 요구가 증가하면서 서로 다른 공간 데이터베이스를 통합하여 사용하기 위한 연구가 증가하고 있다. 그러나 공간 데이터베이스에 존재하는 이질성으로 인하여 통합 작업은 단순하지 못하다. 본 연구는 네트워크 형태의 정보를 공간 데이터베이스를 통합하기 위한 노력으로, 단순화 관계를 갖는 네트워크간의 인스턴스 동일성 확인에 대한 연구를 기술한다. 네트워크 구조의 공간 데이터베이스는 위치 정보와 위상 정보를 명시적으로 포함하고 있으며, 도로, 수도, 배전 등 많은 응용 분야를 갖고 있다. 본 연구의 결과는 이러한 네트워크 구조의 공간 데이터베이스 통합에서 기본이 되는 인스턴스의 동일성 확인을 용이하게 할 수 있을 것으로 기대된다.

1. 개 요

인터넷의 대중화가 가속화되고 기업 경영에서 정보의 중요성이 부각되면서 기존의 데이터베이스의 정보를 사용자의 요구에 맞추어 제공하기 위한 연구가 다양하게 진행되고 있다. 특히 지리 정보 서비스는 그 대상이 광범위하고 응용 분야도 다양하므로 사용자의 요구에 맞는 결과를 생성하는 것에 대한 수요가 증가하고 있다. 이러한 요구 중에는 복수의 지리 정보 데이터베이스를 통합하여 정보를 제공하는 것이 있다. 지리 정보들은 동일한 2차원상의 평면 위에 나타나는 다양한 정보를 다루게 되는데 하나의 데이터베이스가 모든 정보를 포함할 수 없으므로 필요에 따라 개별적인 데이터베이스를 통합하여 사용할 수 있다면 보다 다양한 용도로 정보의 사용이 가능하며 보다 경제적이다. 예를 들어 배전 설비 공사 계획 시 수도, 통신과 같은 다른 시설물의 공사와 연동 하여 수행한다면 보다 경제적인 것이다. 또 이러한 통합된 정보는 새로운 정보를 생산할 수 있다. 그러나 공간 데이터베이스에는 다양한 이질성이 존재하고 있으며 이들은 통합에 기본이 되는 데이터의 동일성을 확인을 어렵게 한다.

네트워크 구조의 데이터는 위치 정보와 함께 위상 정보를 갖고 있는데 이들 정보를 이용하면 보다 정확하고 일반적인 동일성 확인이 가능하다. 따라서 본 연구에서는 네트워크 형태의 동일성 확인, 특히 단순화 관계를 갖고 있는 두 네트워크간의 동일성 확인문제를 효율적으로 기술할 수 있는 방법을 제시한다

본 논문의 구성은 다음과 같다. 2장에서는 본 연구와 관련된 그 동

안의 연구를 정리하고, 3장에서는 본 연구에서 다루고자 하는 문제를 정의한다. 4장에서는 동일성 확인을 위한 추가적인 정보인 제약 조건을 설명하고, 5장에서는 이를 이용한 통합 방법에 대하여 설명한다. 마지막으로 요약 및 향후 연구 계획으로 결론을 대신한다.

2. 관련 연구

공간 데이터베이스의 통합에 관한 연구는 크게 스키마의 통합에 관한 연구와 인스턴스의 동일성을 확인하는 방법에 대한 연구로 나누어질 수 있다. 일반적인 데이터베이스의 스키마 통합에 관한 연구는 80년대부터 진행되었으며, 공간 데이터베이스의 스키마 통합에 관한 연구는 앞서의 연구들을 확장하는 형태로 90년대 후반에 들어 진행되고 있다[1,6].

일반적인 인스턴스의 동일성 확인에 관한 연구는 주로 키 값을 이용하거나 이를 확장한 형태로 이루어졌으며, 공간 데이터베이스의 인스턴스의 동일성을 확인하기 위한 연구는 주로 위치 정보를 비교하여 동일성을 판단하는 방법으로 지도 제작 시 축척이 다른 경우의 동일성을 찾는 연구가 진행되었다[2,3,4,5]. 또 이와 관련하여 서로 다른 두 네트워크간의 단순화 관계를 정의하고 이를 이용하기 위한 연구도 진행되고 있다[7].

네트워크의 위상 정보를 비교하는 연구는 주로 두 네트워크가 동일한 위상정보를 갖고 있는지를 효율적으로 확인하는 알고리즘의 개발에 초점을 맞추어 진행되었다[8].

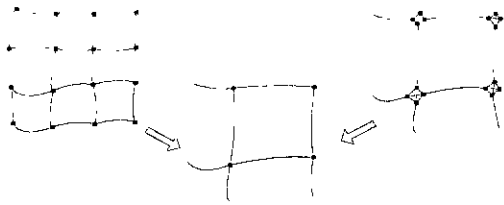


그림 1 네트워크 단순화의 예

결과	유형	예
Node	하나의 link와 link로 연결된 2개의 node	그림 2(a)
	2개의 node	그림 2(b)
Link	양 끝을 공유하는 2개의 link	그림 2(c)

표 1 Amalgamation의 유형

3. 문제 정의

본 연구에서는 단순화 관계의 네트워크를 이루고 있는 서로 다른 두 데이터베이스에서 동일한 객체를 찾는 것을 목표로 한다. 그림 1은 단순화 관계의 예를 나타낸 것이다. 다음에서는 문제를 단순화하기 위해 몇 가지 용어를 정의한다.

3.1. 네트워크

이 연구에서 네트워크 및 그 구성은 다음과 같이 정의된다.
정의 1 Network: 서로 연결되어 있는 node와 link의 집합
정의 2. Node: 위치 정보를 가지고 있으며, 적어도 하나의 link와 연결되어 있다.

정의 3 Link: 두 개의 서로 다른 node와 연결되어 있다.

3.2. 네트워크의 단순화

네트워크의 단순화는 selection과 amalgamation의 두 가지 연산의 조합으로 정의될 수 있다[7].

3.2.1. Selection

Selection은 주어진 network에서 node나 link가 삭제되는 것을 의미한다. 그림 2는 selection의 예로 각 node와 link의 selection

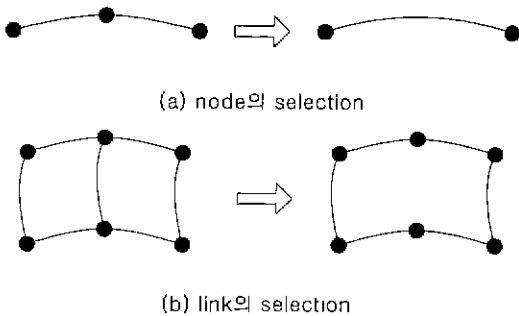


그림 2 Node와 link의 selection

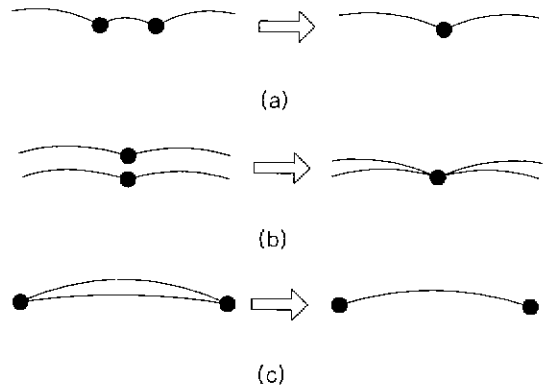


그림 3 Amalgamation의 예

의 결과를 그림으로 나타낸 것이다. Selection은 주로 해상도의 차이에 의해 인스턴스가 단순화 과정에서 생략되는 현상을 나타내며, 임의의 sub-network는 selection의 조합에 의해 삭제될 수 있다.

3.2.2. Amalgamation

Amalgamation은 몇 개의 node 또는 link가 하나의 node나 link로 변화하는 것을 의미한다. 표 1은 amalgamation의 기본적인 연산 3가지를 나타낸 것이고 그림 3은 이에 대한 예를 보인 것이다

Amalgamation은 주로 해상도의 차이에 의해 여러 개의 인스턴스들이 하나의 인스턴스로 단순화되어 나타나는 현상을 나타낸다. 임의의 amalgamation은 앞서의 3가지 기본 amalgamation의 조합으로 정의된다. 그림4는 위의 유형의 조합으로 amalgamation을 분해한 것을 나타내며, 음영 부분은 기본 amalgamation을 나타낸다.

4. 제약 조건

네트워크의 대응 정보는 제약조건의 형태로 표현되게 된다. 이들 제약 조건은 크게 스키마 수준에서의 제약 조건과 인스턴스 수준에서의 제약조건으로 나누어질 수 있는데 스키마 수준에서의 제약 조건은 주로 node와 link의 클래스간의 대응 관계를 표시하고, 규칙의 형태로 나타내며, 인스턴스 수준에서의 제약 조건은 실제 인스턴스의 대응 관계를 명시적으로 표시하게 된다. 이외에도 위치 정보를 활용하기 위한 휴리스틱 정보가 제약 조건에 포함될 수 있다.

4.1. 스키마 수준에서의 제약 조건

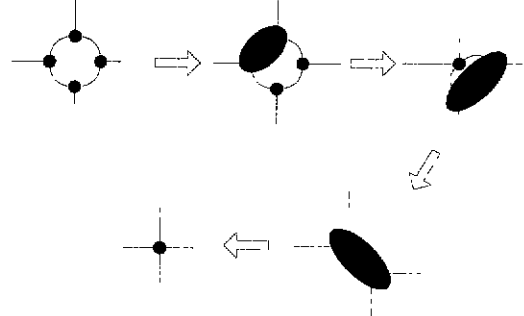


그림 4 Simplification 유형

대응 관계	의미
equal	두 element가 1:1로 대응
part_of	하나의 element는 다른 하나의 부분
no_correspondence	대응되는 element가 존재하지 않음

표 2 인스턴스의 대응 관계

스키마 수준에서의 제약 조건은 주로 인스턴스들이 포함된 클래스 간의 대응 관계를 나타내는 것이다. 가장 기본적인 대응 관계는 클래스의 인스턴스들이 대응될 수 있다는 것을 표시하는 것이고 여기에 추가적으로 단순화에 대한 관계가 추가될 수 있다.

4.2. 인스턴스 수준에서의 제약 조건

인스턴스의 동일성 정보는 다음과 같이 표현될 수 있다. 우선 각 인스턴스의 대응관계는 표 2와 같이 나타낼 수 있다. 대응 관계 중 part_of는 amalgamation관계를 나타내며 no_correspondence는 selection관계를 나타낸다.

인스턴스에 관한 제약 조건에는 이 밖에도 위상관계, 상대적 위치 관계 등이 있을 수 있다. 그리고 이러한 인스턴스 수준에서의 제약 조건은 스키마 수준에서의 제약 조건에 위배되지 않는 것이어야 한다.

4.3. 휴리스틱 정보

실용적인 동일성 확인을 위해서는 앞에서 서술한 스키마와 인스턴스 수준에서의 제약 조건 이외에 보조적인 지식이 필요하다. 위상 관계는 그 정의가 명확한 반면 위치 정보는 포함하고 있지 않기 때문에 난일한 대응 관계가 존재하지 않을 수 있다. 그림 5는 이를 나타낸 것으로 좌측 네트워크의 음영 부분은 우측의 네트워크와 매칭된 부분을 나타내며 위상 정보만을 만족하는 여러 개의 해가 존재할 수 있음을 알 수 있다. 따라서 앞에서 주로 사용한 위상 정보 이외에도 위치 정보를 이용한 동일성 확인 휴리스틱이 보조적으로 사용되어야 실용적인 작업 수행이 가능하다.

5.네트워크 간의 인스턴스 동일성 확인

앞에서와 같이 정의된 네트워크에서의 동일성 확인은 두 네트워크의 단순화를 제약조건에 위배되지 않게 amalgamation과 selection을 이용하여 나타내는 문제와 동일하다. 이 경우 selection이 적용된 인스턴스는 대응되는 것이 없는 것이며, amalgamation이 적용된 인스턴스는 n:1의 대응 관계가 성립하게 된다. 그리고 나머지는 1:1 대응 관계를 나타낸다.

인스턴스 확인시 selection과 amalgamation을 적용하게 되는데 이때 다음과 같은 규칙을 적용하여야 한다.

우선 node의 selection을 수행할 경우 node에 연결된 link가 2개 일 경우에만 가능 하며 node를 삭제함과 동시에 두 link를 연결하여 하나의 link로 처리한다.

Link의 selection을 수행할 경우 link가 삭제되면 양끝에 연결된 node는 link의 개수가 하나씩 줄어든다. 이때 link의 개수가 0인 node는 역시 selection되어야 한다. Link의 selection으로 전체 네트워크가 2개의 네트워크로 분리될 수 있다. 이 경우 2개중 하나의 네트워크는 대응관계가 없는 것이므로 모두 selection되어야 한다.

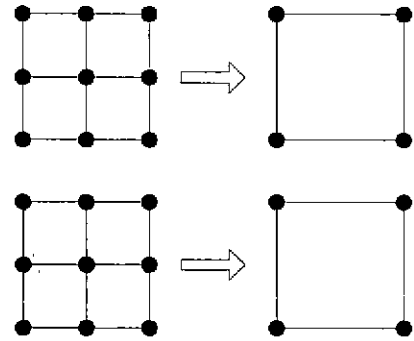


그림 5 위상 정보만을 사용한 단순화 관계의 예

6. 요약 및 향후 연구 계획

네트워크는 GIS분야에서 빈번하게 나타나는 형태로 위치 정보와 위상 정보가 명시적으로 표현되어 유용하게 사용될 수 있다. 본 연구에서는 이러한 네트워크들이 단순화 관계를 가질 경우에 인스턴스의 동일성 확인을 위해 필요한 정보와 이의 표현 방법을 정리하였다. 이러한 결과는 통합에 필요한 정보를 일반적으로 표현할 수 있으며, 이를 이용하면 네트워크와 단순화된 네트워크간의 인스턴스의 동일성 확인을 보다 빠르고 정확하게 할 수 있을 것으로 예상된다.

참고 문헌

- [1] S. Spaccapietra, C. Parent, and Y. Dupont. Model Independent Assertions for Integration of Heterogeneous Schemas *The VLDB Journal*, 1(1):81-126, July 1992.
- [2] Y. R. Wang and S. E. Madnick. The Inter-Database Instance Identification Problem in Integrating Autonomous Systems *In Proc. of the 5th Intl Conf. on Data Engineering. IEEE*, pages 46-55, Los Angeles, California, February 1989
- [3] R. Laurin. Spatial multi-database topological continuity and indexing: a step toward seamless GIS data interoperability. *Intl Journal of Geographic Information Science*, 12(4):373-402, 1998.
- [4] J. D. Nystuen, A. E. Frank, and L. Frank. Assessing Topological Similarity of Spatial Networks. *In Proc. of International Conference on Interoperating Geographic Information Systems*. Pages 58-60, Santa Barbara, California, December, 1997.
- [5] F. Harvey and F. Vaughn. Geometric match processing: applying multiple tolerances *In the 7th International Symposium on Spatial Data Handling*, 1996.
- [6] S. Spaccapietra, C. Parent, and T. Devogele. Analysis of Discrepancies in Spatial Data Representations. *In Proc. of the International Symposium on Cooperative Database Systems for Advanced Applications*, Kyoto, Japan, December 1996.
- [7] J. G. Stell and M. F. Worboys. Generalizing graphs using amalgamation and selection *In the 6th International Symposium on Spatial Databases*, pages 19-32, 1999.
- [8] N. Tryfona and M. J. Egenhofer. Multi-Resolution Spatial Databases' Consistency Among Networks. *In the 6th International Workshop on Foundations of Models and Languages for Data and Objects*, pages 119-132, 1996.