

시간과 공간 데이터마이닝을 적용한 gCRM 구조

황정래, 강혜영, 이기준
부산대학교 시공간데이터베이스연구실
{jrhwang, hykang}@quantos.cs.pusan.ac.kr, lik@pusan.ac.kr

Architecture of gCRM using time and spatial data mining

JungRae Hwang, HyeYoung Kang, KiJoune Li
Spatio-Temporal Database Lab. Pusan National University

요약

오늘날의 고객들은 실시간으로 자기가 필요로 하는 유용한 정보를 신속하고 정확하게 시각화하여 보기를 원한다. 그리고, 고객들은 자기가 원하는 시간대에 필요한 프로그램이나 기능이 있기를 원한다. gCRM은 이러한 고객들의 충격을 만족시키고자 많은 노력을 하고 있다. 그러나, 오늘날의 gCRM은 이러한 충격을 만족시키지 못하고 있다. 따라서, 본 논문에서는 오늘날의 gCRM 구조에 시간의 개념과 공간 데이터마이닝 기법을 적용하여 새로운 gCRM 구조를 제시하였다. 그리고, 공간 데이터마이닝 기법을 gCRM에 적용하여 공간 데이터를 좀더 효율적으로 사용하여 적절한 gCRM의 기능에 부합하고자 하였으며, 시간의 개념을 적용시킴으로써 시간대에 따른 효과적인 고객 관리가 가능하게 되었다.

1. 서론

고객관계관리(CRM)나 마케팅과 같은 경영 방식에서도 대용량의 공간 데이터베이스를 사용하는 지리정보시스템(GIS)과 같은 응용 분야를 접목하고 있다. 이러한 실정을 단적으로 보여주는 것이 gCRM이다. 하지만, gCRM에서도 지리 공간 데이터와 GIS 같은 여러 응용 기술을 적절하게 사용하지 않는다면 효율적인 고객관리가 이루어지지 않을 것이다.

공간 데이터마이닝은 공간 데이터베이스로부터 어떤 특정한 패턴이나 규칙 등을 찾아낸다. 즉, 방대한 지리 공간 데이터를 사용하는 gCRM에서도 공간 데이터마이닝이 필요하다는 것이다. 이러한 공간 데이터마이닝을 gCRM에 사용한다면 지리 공간 데이터와 GIS를 효과적으로 이용할 수 있는 결과가 나타날 것이다.

오늘날의 gCRM은 그림 1과 같이 CRM과 GIS의 결합 형태로 이루어져 있다.

즉, gCRM 기능들이 지리 데이터를 제외한 통합데이터베이스로부터 데이터마이닝이나 OLAP(Online Analytical Processing)과 같은 기술을 사용하여 얻은 결과를 GIS를 통하여 시각화하여 보여준다는 것이다. 그림 1과 같은 구조에서는 지리 데이터와 GIS를 비효율적으로 사용하고 있다는 것을 의미한다. 즉, 방대한 지리 데이터와 GIS가 시각화 기능에 머물러 있다는 것이다. 따라서, 본 논문에서는 지리 데이터와 GIS의 효율적인 사용과 더불어 오늘날의 고객 성향에 맞추어 시간과 공간 데이터마이닝을 gCRM에 적용한 새로운 구조를 제시하고자 한다. 현재 본 논문에서 제시한 gCRM 구조를 바탕으로 구현 단계에 있다. 시간과 공간 데이터마이닝이 적용된 gCRM은 오늘날의 gCRM 시장을 좀더 발전시킬 수 있을 것으로 예상된다.

본 논문은 다음과 같이 구성된다. 2장에서는 gCRM과 공간 데이터마이닝에 대하여 간단하게 소개한다. 3장에서는 시간과 공간 데이터마이닝을 적용한 새로운 gCRM 구조를 제시하고, 그것에 대한 예를 바탕으로 살펴보도록 한다. 그리고, 마지막으로 5장에서는 결론 및 향후과제에 대하여 언급한다.

2. 관련 연구

대용량 데이터로부터 어떤 특정한 패턴이나 규칙, 관계를 찾아내는 방법으로 데이터마이닝을 들 수 있다. 이러한 데이터마이닝은 데이터베이스 마케팅, 주식/외환시장 예측, 천문관측자료/의료데이터 분석 등 거의 모든 분야에서 분석 및 예측에 활용되고 있는 기법이다. 데이터마이닝에 대한 연구는 꾸준히 진행되어 오고 있다. 여기에 맞추어 오늘날 공간 데이터마이닝에 대한 연구도 진행되고 있다. 공간 데이터마이닝은 데이터마이닝과 마

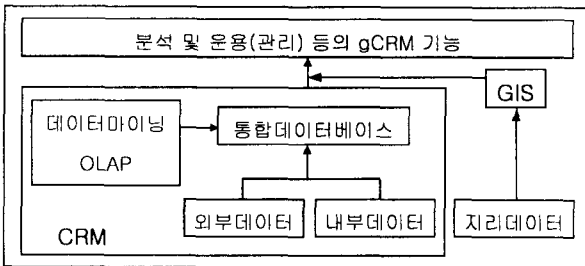


그림 1. 오늘날의 gCRM 구조

찬가지로 공간 데이터베이스로부터 어떤 특정한 패턴이나 규칙을 찾아내는 방법이다. 공간 데이터마이닝으로 잘 알려진 기법으로는 공간 클러스터링, 공간적 분류, 공간적 연관 규칙, 공간적 특성 묘사, 공간 방향 탐사, 일반화 등이 있다. 본 논문에서는 가장 대표적인 몇 가지 공간 데이터마이닝 기법에 대해서만 언급한다.

공간 클러스터링 기법은 개념적인 계층구조와 같은 배경지식을 사용하지 않고도 공간 데이터로부터 직접 관심 있는 구조나 클러스터를 찾는 방법이다. 그리고, 이 기법은 공간 데이터베이스의 객체들을 클러스터화 함으로써 그 클러스터된 객체들로부터 유사성을 찾는 방법이다. 일반적으로 많이 알려진 공간 클러스터링 기법으로는 각 클러스터에 대하여 대표 객체를 결정하는 접근 방식을 가진 PAM(Partitioning Around Medoids)[2], 무작위 검색을 기반으로 하는 CLARANS(Clustering Large Applications based on RANdomized Search)[2] 등이 있다. 공간적 연관 규칙 기법은 하나 또는 그 이상의 객체가 다른 객체들과의 연관 관계를 규칙화 시키는 방법이다. [3] 공간적 분류(Spatial Classification) 기법은 객체의 속성값을 기반으로 한 주어진 클래스의 집합으로부터 하나의 객체를 하나의 클래스에 할당하는 방법이다. [1]

gCRM은 다양한 데이터들을 통합한 데이터베이스로부터 데이터마이닝을 사용하여 여러 규칙이나 패턴을 찾아낸다. 그리고, 찾아낸 결과를 이용하여 고객을 관리하고, 상관분석이나 시장분석 등을 하게 된다. 특히, 지리 공간 데이터와 GIS를 이용하여 그 결과들을 시각화하기도 한다. 하지만, gCRM에서 사용되는 지리 공간 데이터를 시각화하는 기능만으로 사용하기에는 너무 비효율적이다. 공간 데이터마이닝을 사용하여 지리 공간 데이터로부터 어떤 특정한 패턴이나 관심 분야를 찾아내고 분석할 수 있다면 더욱 효율적인 gCRM의 기능을 가지게 될 것이다. gCRM에 대한 자세한 언급은 [6][7]에서 하고 있다. 다음은 대표적인 공간 데이터마이닝 기법이 gCRM에 적용될 수 있는 예를 각각 들어 설명하고 있다.

공간 클러스터링 기법은 gCRM에서 영업지점별 고객분석, 지역별 고객분석 그리고 지역별 시장분석 등 유사한 부분들을 그룹화 시켜 사용자들이 원하는 결과를 얻어내는데 유용한 방법이다. 그리고, 각 영업지점의 거리 변화에 따라 고객의 점유율을 분석하는 기능에도 유용하게 적용될 수 있다. 물론, 다양한 공간 클러스터링 기법이 존재하기 때문에 더 많은 기능들에 적용될 수 있을 것이다. 공간적 분류 기법은 gCRM에서 고객의 성향이나 등급을 감안해서 지역적으로 분류한 후 차별화 된 고객을 유지하고, 그에 따른 마케팅 정책을 펴고자 하는 경우에 적용될 수 있을 것이다. 공간적 연관 규칙 기법은 '많은 회사들이 모여있는 곳이나 상업 지역에는 하나 이상의 은행이 존재할 가능성이 높다'거나 '대규모 아파트 단지를 구성한 곳에는 스포츠 센터가 근교에 있을 가능성이 높다' 등의 관계가 성립할 수 있다. 이렇게 공간적 연관 규칙 기법은 gCRM의 공간적으로 연관되어 있는 객체를 조사하거나 그런 관계를 분석하는 등의 기능에 적용되어질 수 있다.

이러한 적용 가능한 예들을 조사하여 gCRM에 공간 데이터마이닝을 적용할 수 있는 부분을 정리해 보았다.

표1. gCRM과 공간 데이터마이닝의 정리

공간 데이터마이닝 기법	gCRM의 기능
공간 클러스터링[2]	목표배정 및 성과관리 고객분석 및 시장분석
공간특성 묘사[4]	시장분석
공간적 분류[1]	상관분석 고객분석
공간방향 탐사[4]	영업 마케팅 분석
일반화 [5]	고객관리
공간적 연관[3]	점포구조조정 영업 마케팅 분석

3. 시간과 공간 데이터마이닝을 적용한 gCRM

gCRM은 다양한 일반 데이터와 방대한 지리 공간 데이터를 통합한 통합데이터베이스로부터 유용한 정보를 찾아내어 분석함으로써, 고객 관리와 이익 창출이라는 두 가지의 목적을 이루게 된다. 그리고, 오늘날의 고객은 자기가 원하는 시간대에 자기가 필요로 하는 것이 있기를 원한다. 따라서, 본 논문에서는 이러한 배경을 바탕으로 gCRM에 공간 데이터마이닝과 시간을 적용하여 고객의 만족도를 높이고, 영업의 이익을 창출에 도움이 될 것이다.

3.1 시간과 공간 데이터마이닝을 적용한 사례

다음은 gCRM의 기능에 시간과 공간 데이터마이닝이 적용 가능한 사례를 들어 살펴본다.

사례 1

대규모 아파트 단지가 구성되어 있는 주거 지역 근교에는 스포츠 센터가 있을 가능성이 높다. gCRM에서는 이러한 내용을 가지고 더욱 자세한 분석이 가능할 것이다. 예를 들면, 여성이 가장 스포츠 센터를 많이 찾는 시간대와 남성이 가장 스포츠 센터를 많이 찾는 시간대를 조사하여 여성과 남성을 고려한 스포츠 과목을 적절하게 배치하는데 도움을 줄 수 있을 것이다. 또한 여성과 남성이 아닌 연령별로 스포츠 센터를 찾는 시간대를 조사하여 마케팅 정책을 펴는 데도 매우 유용하게 적용될 수 있을 것이다.

사례 1에서 아파트 단지 근교에 새로운 여가 센터를 신설한다고 가정하자. 먼저, '아파트 단지 근교에 스포츠 센터나 문화 센터 같은 여가 활동을 즐길 수 있는 공간이 있을 가능성이 높다' 라는 가정은 공간 데이터마이닝 기법 중 공간적 연관 규칙을 적용하면 쉽게 찾을 수 있다. 이것을 바탕으로 새로이 신설된 여가 센터 지역을 찾아야 할 것이다. 그 지역을 찾기 위해서는 주위에 기존의 스포츠 센터나 문화 센터를 조사하여, 그것이 영향을 미치는 지역과 고객을 분석한다. 이러한 분석은 공간 데이터마이닝 기법 중 공간 클러스터링 기법을 적용할 수 있다. 즉, 그 영향력이 미치는 지역의 점유율과 그 지역의 고객의 고객을 분석하는데 공간 클러스터링 기법이 적합하게 적용된다. 그리고, 그 지역의 고객들을 대상으로 시간대별로 또는 연령대별로 그 곳을 찾는 유형을 조사하여야 할 것이다. 이것은 고객들의 성향에 맞는 프로그램이나 기능을 추가하는 등의 고객의 만족도에 맞는 마케팅을

펼치는 데 유리하게 작용할 것이다.

사례 2

일반적으로 토요일이나 일요일 또는 명절 며칠 전에는 백화점 같은 곳에는 사람들이 많이 찾는다는 것을 알 수 있다. 따라서, 백화점을 가장 많이 찾는 요일과 시간대를 조사하고 백화점내의 가장 사람들이 물리는 장소를 조사하는 것이 가능할 것이다. 그리고, 백화점에서 소매치기와 같은 범죄가 가장 많이 나타나는 시간대와 장소를 조사하면, 조사한 두 경우가 서로 연관성이 있다는 것을 알 수 있다. 따라서, 고객이 백화점을 가장 많이 찾는 요일 및 시간대, 장소 그리고 범죄율을 분석하여, 백화점에서는 소매치기와 같은 범죄를 미리 예방함으로써 고객을 관리하는데 더욱 유용하게 적용하여 사용이 가능할 것이다.

사례 2에서는 시간의 개념을 중요시하고 있다. 고객이 많이 찾는 시간대를 조사하고, 범죄율이 높은 시간대를 조사하여 두 경우의 연관 관계를 분석한다. 즉, 고객이 많이 찾는 시간대에 가장 범죄율이 높다는 것을 연관 규칙으로 알 수 있을 것이다. 사례 2와 같은 경우는 백화점 이외에도 적용이 가능하다.

사례 3

연휴가 있는 경우나 명절이 있는 경우 아파트나 주택들이 밀집한 지역에 강도나 도둑 같은 범죄가 많이 일어날 가능성이 높다. 이러한 경우 역시 범죄가 발생하는 지역과 시간대를 분석함으로써 범죄 예방이 가능할 것이다.

사례 3 역시 공간적 연관 규칙 기법을 적용하여 ‘연휴가 있는 경우나 명절이 있는 경우 아파트나 주택들이 밀집한 지역에 강도나 도둑 같은 범죄가 많이 일어날 가능성이 높다’ 라는 것을 알 수 있다. 이것을 바탕으로 범죄가 미치는 지역을 공간 클러스터링 기법을 적용하여 분석하고, 그 시간대를 추가적으로 분석함으로써 효율적인 범죄 예방이 가능하다.

위의 세 사례는 고객 성향에 맞추어 고객을 관리하면서 효율적인 마케팅 정책을 펼 수 있다. 즉, gCRM에 시간과 공간 데이터마이닝을 적용함으로써 시장 및 상관분석이나 고객 관리 기능이 더욱 효율적으로 이루어진다는 것을 보여주고 있다. 물론, 이외의 여러 사례들이 있을 수 있다.

3.2 시간과 공간 데이터마이닝을 적용한 gCRM 구조

앞에서 이전의 gCRM 구조가 좀 효율적이지 못하다는 것을 살펴봐왔다. 특히, 지리 공간 데이터의 사용이 텍스트로 나온 결과를 시각화하는 기능에 멈추어 있다는 것을 알았다. 따라서, 본 논문에서는 그림 2와 같은 새로운 형태의 gCRM 구조를 제안한다. 그림 2는 이전의 gCRM 구조에 공간 데이터마이닝과 시간이 추가되어 통합 데이터베이스로부터 얻고자 하는 gCRM의 기능을 더욱 효율적으로 사용하고자 제안한 구조이다. 지리 공간 데이터가 이전에는 통합데이터베이스로부터 나온 결과를 시각화하는 기능에 머물렀다. 하지만, 그림 2에서 공간 데이터마이닝이 적용됨으로써 일반 데이터와는 달리 지리 공간 데이터를 더욱 효과적으로 사용할 수 있다는 것을 보여준다. 특히, 시간이 추가 된 것은 오늘날의 고객이 실시간으로 보다 신속하고

정확한 정보를 보고자 하는 성향에 맞춘 것이라 할 수 있으며, 시간대에 따른 고객 관리가 더욱 효율적으로 이루어질 수 있을 것이다.

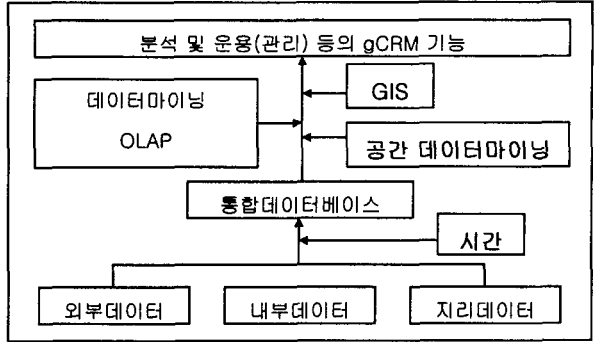


그림2. 시간과 공간 데이터마이닝을 적용한 gCRM 구조

4. 결론

본 논문에서는 gCRM에 공간 데이터마이닝 기법과 시간의 개념을 적용하여 새로운 gCRM 구조를 제안하였다. 따라서, 실시간으로 시각화된 정보를 얻고자 하는 오늘날의 고객 성향에 맞추어, gCRM의 기능이 보다 효과적이고 보다 신속하고 정확한 분석 및 관리가 가능하게 될 것이다. 특히, 시간의 개념을 gCRM에 적용시킴으로써 시간대에 따른 효과적인 고객 관리가 가능하게 될 것이다. 본 논문에서 제안한 공간 데이터마이닝 기법과 시간을 gCRM에 적용한 것을 바탕으로 현재 구현 단계에 있다. 이러한 구현은 현재의 gCRM을 한층 더 발전시킬 수 있는 계기가 될 것으로 예상된다.

5. 참고 문헌

- [1] Ester M., Frommelt A., Kriegel H.-P., Sander J. " Spatial Data Mining: Database Primitives, Algorithms and Efficient DBMS Support", *In Data Mining and Knowledge Discovery*, 4, pp. 193-216, 2000
- [2] Ng R. and Han J. " Efficient and effective clustering methods for spatial data mining", *Proc. 20th Int. Conf. VLDB*, pp. 144-155, 1994
- [3] Ester M., Kriegel H.-P., Sander J. " Spatial Data Mining: A Database Approach", *Proc. 5th Int. Symp. on Large Spatial Databases*, pp. 47-66, 1997
- [4] Ester M., Frommelt A., Kriegel H.-P., Sander J. " Algorithms for Characterization and Trend Detection in Spatial Databases", *Proc. 4th Int. Conf. on Knowledge Discovery and Data Mining*, pp. 44-50, 1998
- [5] Koperski K., Adhikary J., Han J. " Spatial Data Mining: Progress and Challenges", *Proc. SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, TR 96-08, University of British Columbia, Vancouver, Canada, 1996
- [6] 황정래, 어기준, " gCRM과 공간데이터마이닝", 2002 공동통계학회 논문집, p.38~44
- [7] 한국지리정보, *gCRM: GIS와 고객관리 CRM의 만남*, pp.58~63, 2001.10