

시간 및 공간 마이닝 기술을 이용한 다이렉트 마케팅 시스템 프로토타입 개발¹⁾

이헌규*, 최용훈*, 나동길*, 정훈*, 박종홍*
 *한국전자통신연구원 우정물류기술연구부,
 e-mail:hg_lee@etri.re.kr

A Development of Direct Marketing System Prototype Using Temporal & Spatial Mining Techniques

Heon Gyu Lee*, Yong Hoon Choi*, Dong-Gil Na*, Hoon Jung*,
 Jong Heung Park*

*Postal&Logistics Technology Dept, Electronics and Telecommunications
 Research Institute

요 약

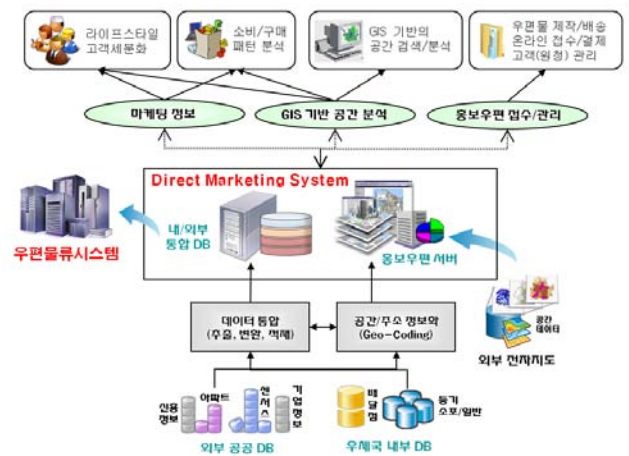
본 논문에서는 시간 및 공간 마이닝 기술을 적용한 다이렉트 마케팅 지원 시스템 프로토타입을 개발하였다. 개발한 프로토타입은 서울시를 대상으로 약 인구 500명 크기의 블록단위 e-Commerce 구매 패턴과 유사블록 그룹핑 및 기타 마케팅에 유용한 외부 공개 자료의 검색 기능을 포함한다. 또한, 마케팅 캠페인에 프로토타입의 활용도를 높이기 위해서 상품번호도 기반 검색, 라이프스타일 기반 검색 및 복합정보 기반 검색 모듈 등의 다양한 서비스를 제공한다.

1. 서론

일반적으로 다이렉트 마케팅(direct marketing)이란 잠재 또는 기존 고객에게 우편을 통한 상품/서비스 홍보를 의미한다. 대형유통업체의 광고, 각종 고지서, 편지.엽서.소포 등의 우편물은 물론 전단의 배포, 또는 실물견본이나 정기간행물 등도 모두 다이렉트 마케팅 분야에 포함된다. 이러한 마케팅 방법은 우편배달주소를 이용한 CRM, GIS 및 시간, 공간정보 분석기술을 활용하여 원청 고객이 요구하는 특정 배달지역과 배달주소를 분석, 제공하는 새로운 우편서비스이다[1]. 다이렉트 마케팅의 성공 요인은 내/외부 고객리스트를 활용한 대상고객의 선정, 상품 제안 방법, 광고물의 형태, 제품의 가격이며 이중 적절한 고객 리스트의 확보가 홍보에 대한 성패의 40%~80%를 차지한다고 알려져 있다[2]. 따라서 적절한 고객 리스트의 선정을 위하여 시간 및 공간 정보를 포함한 정확하고 효율적인 분석 기술이 다이렉트 마케팅에 필요하다[3].

본 논문에서는 기존의 마케팅 서비스에 IT기술인 GIS, 시간, 공간 마이닝 기술을 접목한 온라인 다이렉트 마케팅 서비스 시스템을 제안한다. 제안한 시스템은 마케팅 캠페인을 원하는 기업체 또는 원청업체를 대상으로 순차패턴, 주기성 분석을 통한 구매패턴 예측과 GIS 기반의 공간 클러스터 분석을 통한 유사 라이프스타일 지역 정보를 제공한다. 이를 위해서 (그림 1)과 같이 시스템은 다양한 내/외부 데이터를 연계하여 공간 정보화를 통한 시간, 공간 마

이닝을 수행하며, 다이렉트 마케팅에 유용한 정보를 사용자에 제공한다.



(그림 1) 다이렉트 마케팅 시스템 개념도

본 논문의 다이렉트 마케팅 시스템에서 제공되는 기능은 다음과 같다.

- **상품번호도 기반 지역 검색**
 인터넷 쇼핑물의 주소지별 구매이력 분석을 통한 상품 품목별 선호도와 구매패턴 정보를 제공한다.
- **라이프스타일 기반 지역 검색**
 블록 단위 인구/사회/경제학적 특성 정보와 클러스터 분석을 통한 유사 라이프스타일을 가진 블록들의 그룹 정보를 제공한다.

1) 본 연구는 지식경제부 우정사업본부의 우정기술연구개발사업의 일환으로 수행하였음 [2006-X-001-02, 실시간 우편물류 운영기술 개발]

▪ 복합 정보 기반 지역 검색

상품번호도와 라이프스타일 정보를 통합한 검색 기능으로 라이프스타일 기반의 블록 탐색 후, 특정 상품의 선호도 정보를 기반으로 마케팅 지역 검색 기능을 제공한다.

2. 시스템 설계

제안한 시스템은 통합 DB부터 웹서비스까지의 4계층 구조로 설계한다. 최상위 Web Service 층은 지역 검색, 사용자 인증, 주문/접수 등의 기능과 GIS 활용을 위한 GIS ActiveX 엔진을 포함한다. 내부/외부 데이터 및 공간 데이터의 조작/관리 및 공간연산, 질의처리, 공간/속성 데이터 전송 등의 Geo 서버 기능은 GIS 미들웨어 서비스 층에서 담당한다. Spatial/Customer DB 서비스 계층은 공간 데이터 관리 및 고객 정보 관리를 포함한다. 대용량의 고객 주소 정보는 효율적인 관리를 위해서 독립적인 DB로 분류하여 설계한다.

다이렉트 마케팅 시스템의 구조도는 (그림 2)와 같고, 개발환경은 <표 1>과 같다.



(그림 2) 시스템 구조도

<표 1> 시스템 프로토타입 개발환경

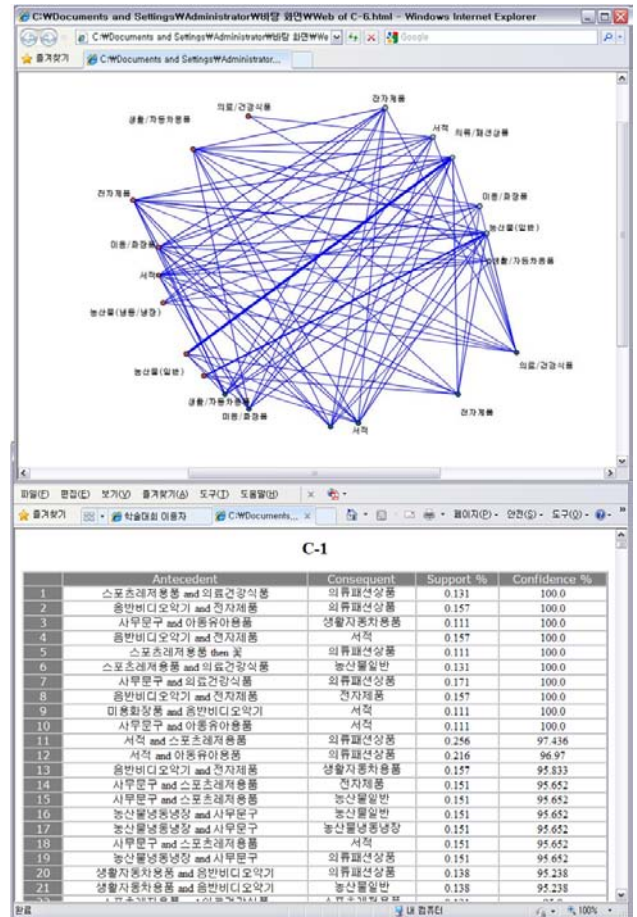
구성	세부항목
서버	Unix 서버
웹 개발 언어	HTML, JSP
웹 개발 툴	Eclipse J2EE
UX 컴포넌트	Flex, 익스트림 빌더
DBMS	Oracle 10g, PostGIS
GIS 엔진	GeoBus
웹서버/서블릿 컨테이너	Apache/Tomcat, Weblogic

3. 시간, 공간 마이닝을 이용한 구매패턴 및 클러스터 분석

상품번호도 분석을 위한 시간 마이닝으로 순차패턴[4] 및 주기성 분석[5], [6]을 수행하며, 블록 단위 라이프스타일 분석을 위해 인구통계, 사회, 경제학적 유사 지역 그룹핑을 위해 k-means[7] 알고리즘을 통한 클러스터 분석을 수행한다.

▪ 순차패턴 마이닝

e-Commerce를 통한 2008년 및 2009년 택배접수 이력 데이터로부터 고객의 구매패턴 분석을 [4]에서 제안한 AprioriAll 알고리즘을 사용한다. 11,036,940건의 구매 트랜잭션에 대해, 순차패턴의 주요 임계값인 지지도와 신뢰도는 각각 1%, 80%로 설정하여 블록 단위별 순차패턴들을 추출하였고, (그림 3)과 같이 웹을 통해 가시화 하였다.



(그림 3) 순차패턴 마이닝 결과 예

▪ 선호상품에 대한 구매 주기성 분석

구매 주기성 분석을 위해서는 [5], [6]에서와 같이 시간의 단위 설정이 필요하다. 이 논문에서는 트랜잭션의 구매일자(예: 2008.12.22.) 속성값을 "월 | 주 | 요일"로 변환하여 각 시간 단위에서의 주기성을 탐색한다. 예를들어, 2008년 12월 22일에 농산물과 서적이 판매되었을 경우의 변환된 데이터 형식은 <표 2>와 같고 블록단위 주기성 분석 결과는 <표 3>의 예시와 같다.

<표 2> 주기성 분석을 위한 입력 데이터 형식

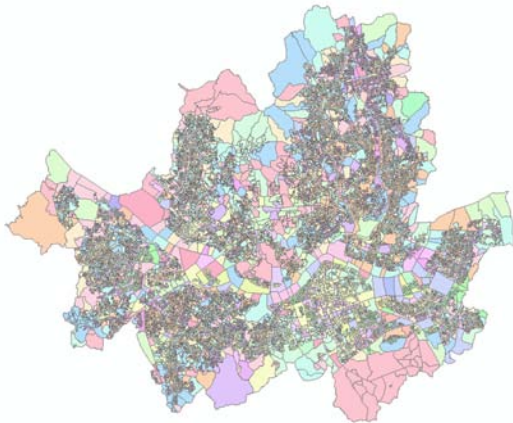
TID	Month	Week	Day	Items
100	12	4(주)	월	농산물, 서적

<표 3> 블록 단위의 주기성 분석 결과

빈발항목	시간 주기성	블록 ID
{의류/패션, 서적, 꽃}	<매월, 2주, 월요일>, <매월, 2주, 수요일>, <1월, 매주, 화요일>	C1, C12, C14
{의류/패션, 미용/화장품}	<매월, 3주, 월요일>, <매월, 3주, 화요일>, <매월, 3주, 금요일>, <1월, 매주, 수요일>	C1, C2, C162
...
{농산물(일반), 서적}	<매월, 3주, 월요일>, <매월, 3주, 화요일>	C4, C8, C110, C1123, C1782, C25014, C26542

■ 라이프스타일 분석을 위한 클러스터 분석

블록 단위에서 유사한 연령, 학력, 종교 등의 인구·사회적 특성과 거주지, 주택, 신용정보에 관한 경제적 특성을 이용한 지역 그룹화를 위해 k-means[7] 알고리즘을 사용한다. k-means 분석 결과 서울시 16,357개 블록지역에 대해, (그림 4)와 같이 36개 유사 지역으로 그룹화 하였다.



(그림 4) 라이프스타일 유사 그룹(동일 색은 같은 라이프스타일이 같은 그룹임)

4. 시스템 프로토타입

■ 상품선호도 기반 검색 모듈

사용자는 최초 마케팅 대상 영역을 설정(반경)한 후, 상품선호도 검색 조건 창을 통해 관심있는 상품품목을 선택하고 상품구매 횟수와 최근성을 입력할 수 있다. 시스템은 입력 조건에 만족하는 홍보물 배송주소지를 디스플레이하며, 그 지역의 순차패턴, 구매 주기성을 사용자에게 제공한다. (그림 5)는 강남구 일대, 의류/패션 상품을 5회 이상 구매한 이력의 배송지와 구매패턴, 주기성을 검색한 결과이며, 서비스 시나리오는 다음과 같다.

- ① 서비스 접속 후에, 초기 지도 화면의 중심 위치 정보와 마케팅 대상 지역을 지도상에 영역을 설정한다.
- ② 홍보 상품과 동일 품목이나 유사성이 있는 품목을 카탈로그에서 선택하고 검색한다.
- ③ 검색 결과인 상품 선호도 조건을 충족한 배송지역과 선호도 정보, 대상 지역의 구매패턴 및 주기성 정보를

제공한다.

- ④ 원청고객은 제공된 배송지역에 홍보물을 웹을 통해 접수한다.

■ 라이프스타일 기반 검색 모듈

클러스터 분석을 통한 36개 군집의 프로파일에 기반하여, 인구/사회/경제학적 속성 정보 및 라이프스타일 정보를 제공한다. (그림 6)은 송파구 일대에 대해, 첫 번째 라이프스타일 군집 유형이면서, 고령인구 밀집지역, 아파트 밀집지역을 조회한 결과이다. 서비스 시나리오는 다음과 같다.

- ① 초기 지도 중심 위치 정보와 마케팅 대상 지역을 지도 위에 영역을 표시한다.
- ② 홍보물에 관심이 높을 가능성이 있는 잠재 고객을 라이프스타일 통계 조건을 입력하고 검색한다.
- ③ 검색 결과인 라이프스타일 통계 항목을 블록 단위의 통계값 및 등급별로 지도 위에 표시한다.
- ④ 원청업체가 지도 위에서 반경, 다각형 설정 및 지역 선택 등으로 마케팅 캠페인 대상 지역을 선택하면 해당 지역에 대한 배송지 개수, 라이프스타일 상세 분석 정보 등을 가시화한다.
- ⑤ 원청업체가 캠페인 지역을 확정하면, 홍보물을 웹을 통해 접수한다.

■ 복합 정보 기반 검색 모듈

라이프스타일 검색과 선호도 검색 기능을 통합한 형태로 관심있는 지역의 군집정보, 인구/사회/경제학적 정보를 먼저 검색한 후에, 높은 통계 수치 지역에 대한 상품선호도를 재검색 한다. 아래 그림은 송파구 일대에 대한, 미혼 여성이 많고, 의류/패션 상품을 최근 1년 이내 구매한 배송지를 검색한 화면이다. 상세 서비스 시나리오는 다음과 같다.

- ① 초기 지도 화면의 중심 위치 정보와 마케팅 캠페인 대상 지역을 지도상에 표시한다.
- ② 홍보물에 관심이 높을 것 같은 잠재 고객 선정을 위해 라이프스타일 통계 조건을 입력하고 검색한다.
- ③ 라이프스타일 검색 범위 안에서 홍보물과 동일 품목이나 유사성이 있는 품목을 카탈로그에서 선택하고 검색한다.
- ④ 검색 결과인 상품 선호도 조건을 충족한 배송지역과 선호도 정보, 대상 지역의 구매패턴 및 주기성 정보를 제공한다.
- ⑤ 원청고객이 지도상에서 검색된 배달점에 대해서 반경, 다각형, 지역 선택 등으로 캠페인 대상 지역들을 선택하면 해당 지역에 대한 배송지 정보, 라이프스타일 군집의 특성정보 등을 가시화 한다.
- ⑥ 원청업체가 캠페인 지역을 확정하면, 홍보물을 웹을 통해 접수한다.

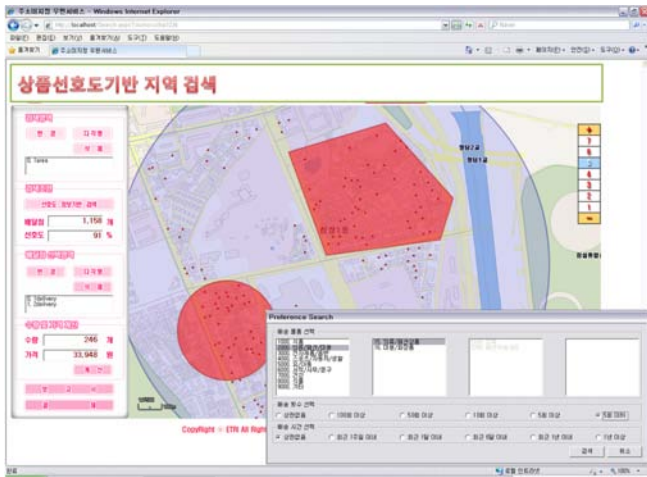
5. 결론 및 향후연구

본 논문에서는 효율적인 마케팅 캠페인 지원을 위한 시스템 개발을 목적으로 시간 및 공간 마이닝에 기반한 다이렉트 마케팅 시스템 프로토타입을 개발하였다. 본 시스템은 순차패턴 및 주기성 분석과 라이프스타일 클러스터 분석을 통해, 상품번호도 기반 검색, 라이프스타일 기반 검색 및 복합정보 기반 검색 등의 다양한 서비스를 제공한다.

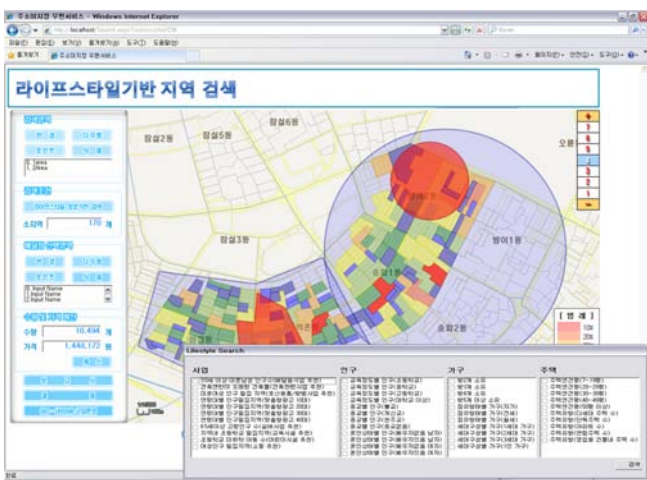
향후연구로 현재 개발된 프로토타입을 기반으로 전국 서비스로의 확대를 통해 기업의 다이렉트 마케팅 지원 및 상권분석, 입지선정, 고객 타겟팅 등의 서비스에 활용 가능하도록 확대 개발이 필요하다.

참고문헌

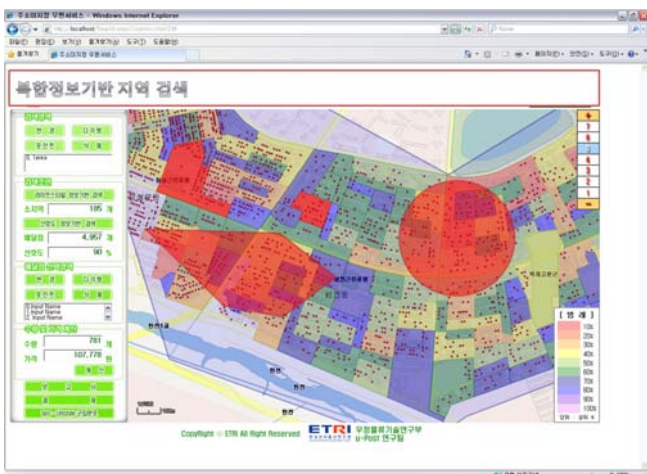
- [1] 나동길, 이현규, 정훈, 박종홍, “홍보우편 서비스 및 기술 동향,” 전자통신동향분석 제25권 제4호, 2010년 8월.
- [2] 이훈영, “CRM 해법강화를 위한 Direct Marketing의 특성과 문제점 및 개선방안, 마케팅 4월호, pp. 30-37, 2002.
- [3] 정동한, 최원길, 김응모, “고객 중심의 기업 경영 및 마케팅을 위한 데이터 마이닝의 활용: 멀티플렉스에 적용,” 한국정보처리학회 2008년 추계학술발표대회, 제15권 제2호, 2008, pp. 311-314.
- [4] R. Agrawal and R. Srikant, “Mining Sequential Patterns,” In Proceedings of the 11th International Conference on Data Engineering, pp. 3-14, 1995.
- [5] Y. Li, P. Ning, “Discovering Calendar-based Temporal Association Rules,” In proceedings of Int'l Symposium on Temporal Representation and Reasoning, 2001.
- [6] 이현규, 노기용, 서성보, 류근호, “캘린더 패턴 기반의 시간 연관적 분류 기법”, 정보과학회 논문지, 제32권 제6호, 2005, pp. 567-584.
- [7] G. W. Milligan, “Clustering validation: Results and implications for applied analyses,” in Clustering and Classification World Scientific, 1996, pp. 341-375.



(그림 5) 상품번호도 지역 검색 화면



(그림 6) 라이프스타일 지역 검색 화면



(그림 7) 복합정보 지역 검색 화면