

B2B에서 구매의도 파악을 위한 웹 로그 분석

Web Log Analysis for Studying the Intend to Purchasing Under B2B Environment

고재문·서준용·김운식
울산대학교 산업정보경영공학부

요 약

일반적으로 B2C가 불특정 다수에 대한 서비스라면 B2B는 특정 소수에 대한 서비스라고 할 수 있다. 이러한 특성으로 B2C와 B2B에서 고객의 구매의도는 다르게 평가되어야 한다. 또한 B2B는 협상이라는 단계가 있고, 이것은 B2C와 B2B의 구매의도 평가기준에 영향을 미치게 된다. 본 연구에서는 B2B에서 구매의도 파악을 위한 웹 로그 분석 모형을 제시한다. 제시된 모형을 통해 구매의도 파악을 위한 웹 로그 분석 데이터를 추출하고, 추출된 데이터를 기업의 레거시 시스템 데이터와 통합하는 과정을 보여준다. 또한 분석 데이터를 추출하기 위한 웹 마이닝 과정과 추출된 분석 데이터가 데이터베이스에 저장되는 과정을 보여준다.

1. 서 론

e-Business는 거래 주체에 따라 B2C와 B2B 등으로 구분된다. B2C는 불특정 다수의 고객에 대한 서비스를 제공하고, B2B는 특정 소수의 고객에 대한 서비스를 제공한다. B2C에서는 불특정 고객이 구매의도 없이 단순한 관심만으로 사이트 내에서 행위를 할 수 있다. 따라서 B2C에서 불특정 고객이 사이트 내에서 보인 행위를 모두 고객의 구매의도라고 보기는 힘들다. 하지만 B2B는 업무의 연장선으로 정해진 고객이 특정한 목적을 가지고 사이트 내에서 행위를 하게 된다. 즉 일반적으로 협상이라는 과정을 통해서 구매를 하게 되는데, 협상의 결과에 따라 구매가 성사되지 않을 수도 있다. 하지만 구매가 성사되지 않았다고 하더라도 고객이 제품을 구매하려는 본질적인 구매의도가 사라졌다고 볼 수는 없다. 이는 향후 제반 여건이 충족이 되면 다시 구매할 수도 있기 때문이다. 따라서 B2B에서 고객이 사이트 내에서 보인 모든 구매행위는 고객의 구매의도라고 볼 수 있다. 그럼으로 B2C와 B2B에서는 고객

의 구매의도를 평가하는 방법도 달라져야 한다.

일반적으로 기업간 거래에서 업체의 구매의도는 기업간 구매 정보를 통해 파악된다. 기존의 기업정보시스템에서는 기업간 거래가 내부 레거시 시스템에 저장되고, 저장된 데이터의 분석을 통해 고객의 구매의도를 파악한다. 이러한 레거시 시스템은 대부분 실제 트랜잭션만 기록하기 때문에, 고객의 구매의도 파악을 위한 분석은 실제 거래행위에 국한된다. 그러나 온라인 환경의 B2B에서 고객의 구매의도는 실제 거래행위뿐만 아니라 실제 거래는 이루어지지 않았지만 온라인상에서 보인 행위 그 자체도 구매의도로 볼 수 있다. 이러한 행위 정보는 온라인 환경의 웹 로그 파일에 기록된다.

본 연구에서는 수주생산기업 B2B에서의 구매의도 파악을 위한 웹 로그 분석 모형을 제시한다. 제시된 모형을 통해 구매의도 파악을 위한 웹 로그 분석 데이터를 추출하고, 추출된 데이터를 기업의 레거시 시스템 데이터와 통합하는 과정을 보여준다. 또한 분석 데이터를 추출하기 위한 웹 마이닝 과정과 추출된 분석 데이터가 데이터베이스에 저장되는 과정을 보여준다.

2. 이론적 배경

2.1 웹 마이닝(Web Mining)

웹 마이닝이란 웹 시스템 상의 데이터에 대해 적절한 분석기법을 적용하여 업무에 유용한 지식 혹은 패턴을 발견하는 기법을 의미한다. 협의로는 웹 서버의 접속 로그를 분석하여 웹 서비스에 대한 방문자의 행위 특성을 분석하는 기법을 의미하기도 하지만, 광의로는 웹 서비스에 의해 파생되는 제반 정보 및 웹으로 접근 가능한 모든 정보를 대상으로 하는 포괄적인 데이터마이닝 기법을 의미하기도 한다. 웹 마이닝은 어떻게 하면 웹상에 존재하는 데이터를 효율적으로 분석하여 고객의 요구에 맞는 자료를 제공하는가를 연구하는 웹 내용 마이닝(web content

mining) 분야와 웹 서버 로그(web server log), 조회 로그(referral logs), 등록 데이터(Registration data) 등을 이용하여 사용자 경로분석(path analysis), 사용자 이용패턴분석(pattern analysis)을 중심으로 하는 웹 사용 마이닝(web usage mining) 분야로 나눌 수 있다.

2.2 웹 로그 분석

초기의 웹 로그 분석은 주로 고객이 웹 사이트에서 보인 행위를 분석하고, 이를 웹 사이트의 디자인 전략 등으로 이용하였다. 이화영[6], Borges et al.[10], Srivastava[11]는 로그 분석을 통해서 사용자의 패턴을 분석하여 사용자 특성에 중점을 둔 차별화된 웹 페이지 디자인 전략을 제안하였다. 이를 위해서는 웹 사용자의 패턴을 분석해야 하는데 Fu[13]는 웹 사용자의 접속 패턴을 클러스터링하고 일정한 규칙을 제공하는 방법론을 개발하였고, Igor[14]는 확률 기반의 군집 프레임워크를 제안하고 고객들의 웹 사이트 방문 형태를 다양하게 분류하였다. 또한 Mannila et al.[15]은 웹에서 사용자들의 거래 자료를 순차적인 패턴으로 인식하는 방법론을 개발하였다.

점차적으로 온라인 거래가 활성화 되면서 웹 로그 분석은 고객의 구매행위에 따른 마케팅의 수단으로 이용되었다. 박형수[3]는 일대일 마케팅에서 로그 분석에 필요한 사전처리 과정을 설계하였고 이를 통해 의미있는 규칙을 생성하였다. 최덕원[9]은 B2C에서 상품 추천을 위한 웹 개인화 방안을 제시하였고, 이재신[5]은 개인의 이력경로와 군집의 이동경로를 관찰, 예측함으로써 각 고객과 고객군집에 대한 적절한 마케팅 프로모션을 수립하였다. 또한 전영철[8]은 개별 상품에 따른 구매행위를 파악하여 신규 상품에서의 홍보 전략을 수립하였다. 변숙은[4]은 온라인상의 마케팅에서 사이트 로딩 속도, 사이트 홍보, 고객의 지속적 방문 측면에 대한 문제점과 개선방안을 제시하였다. 장윤경[7]은 온라인상에서 고객의 구매행동을 분류하고 이에 따른 상품구입의 관계를 분석하여, 고객의 가치를 평가할 수 있는 방안을 제시하였다.

지금까지 웹 로그 분석에 대한 연구는 주로 웹 사이트의 관리 및 개선 방향에 대한 연구와 B2C e-CRM 및 고객에 대한 마케팅에 대한 연구로 크게 나누어 볼 수 있다. 하지만 전자상거래 시장의 많은 부분을 차지하는 B2B에서의 로그 분석은 기업간 거래 정보와 기업 내부 정보를 연계하여 분석하는 것이 중요한 의미를 가지지만 이에 대한 연구는 미비한 상태이다.

3. 구매의도 파악을 위한 웹 로그 분석

온라인 거래에서는 기존의 오프라인 거래와는 달리 웹 로그를 통해 고객의 구매이력 및 행동 등을 파악할 수 있다. 이는 고객의 다양한 구매 취향을 파악함으로써, 기존 고객의 이탈 방지와 잠재 고객

의 확보를 위한 마케팅 전략에 활용된다. 이와 같은 고객 행동의 분석은 고객 만족도 향상과 함께 기업의 마케팅 전략에서도 우위를 차지할 수 있는 효과를 얻을 수 있다[1, 2]. 따라서 기업은 고객의 구매의도를 파악하는데 많은 투자를 해야 한다.

온라인 전자상거래에서 B2C가 보편화된 거래 프로세스를 가진다면, B2B는 특정 업종에 따라 조금씩은 다른 프로세스를 가지게 된다. 일반적인 온라인 거래는 업종 특성에 따라 일부 추가되거나 변형된 프로세스를 가지지만, 기본적으로 <그림 1>과 같은 거래 프로세스를 가진다.

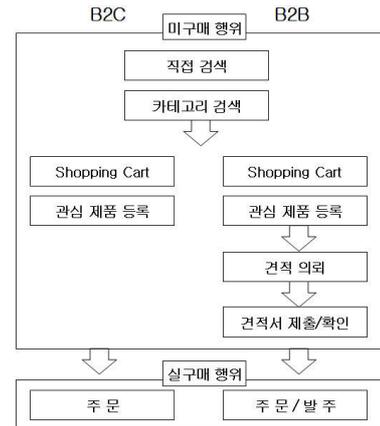


그림 1. 거래 프로세스

일반적으로 구매의도란 고객 또는 소비자가 특정 제품이나 서비스를 구매 또는 이용하고자 하는 의도를 말한다. 이러한 구매의도는 웹 사이트 내에서 제품을 구매하기 위한 일련의 행위로 나타나는데 이를 구매행위라 한다.

본 연구에서는 고객이 구매의도를 가지고 구매여부와 관계없이 웹 사이트에서 보인 모든 행위를 구매행위라고 정의하고, 고객이 웹 사이트에 접속해서 구매행위를 하고 웹 사이트를 이탈할 때까지를 하나의 세션(Session)이라고 한다. 하나의 세션은 한 고객이 웹 사이트에서 보인 여러 구매행위를 가지게 된다. 이러한 고객의 구매행위는 실구매행위와 미구매행위로 나누어지는데, 실구매행위는 구매행위 중 실제 제품 구매로 성사된 행위로서 <그림 1>의 “실구매행위” 부분에 해당한다. 미구매행위는 구매행위 중 실제 제품 구매로 성사되지 않은 행위로서 <그림 1>의 “미구매행위” 부분에 해당한다. 이들에 대한 척도로서 구매율과 미구매율을 생각할 수 있는데, 구매율은 고려 대상이 되는 총 세션 중에서 구매가 성사된 세션의 비율로 정의하고, 미구매율은 (1 - 구매율)로 정의한다.

<그림 2>의 B2B 구매의도 파악 모형은 크게 레저시 시스템 데이터를 분석한 “DB 분석 정보”와 웹 로그 파일 분석을 통해 얻어진 “CBMG 분석 정보” 그리고 이를 통합하여 얻어진 “구매의도 정보”로 나타낼 수 있다.

3.1 CBMG 분석 정보

본 연구에서 사용한 웹 로그 분석은 넓은 의미에서의 웹 로그 분석으로, 웹 로그 데이터와 기업 백엔드시스템 데이터를 포함하여 분석을 실시한다 [4, 6]. 웹 로그 데이터는 기업의 B2B 전자상거래 사이트 서버에 저장된 로그 데이터를 사용한다.

서버에 기록되는 로그파일은 크게 get 방식과 post 방식의 두 가지 방식에 따라 다르게 기록된다. 이는 프로그램 구현단계에서 정의된다. get 방식은 url(cs-uri-stem)의 뒤에 질의어(cs-uri-query)를 붙여 보내는 방식이고, post 방식은 질의어를 http 헤더 안에 포함시켜 전송하는 방식이다. 이 두 방식의 가장 큰 차이점은 질의어를 보내는 방식의 차이인데, get 방식으로 질의어를 보낼 경우에는 로그파일

표 1. 거래 프로세스에 따른 백엔드데이터 및 질의 항목

업무 프로세스	제품 검색	관심 제품 (Cart)	견적서의뢰	견적서 발행	주문	수주	보증 요청
백엔드 데이터	제품 DB	Cart DB	견적의뢰 DB	견적서 DB	주문 DB	수주 DB	보증 DB
질의어	공통	고객 코드, 사용자 ID, 요청 제품					
	개별		견적의뢰 No	견적서 No	주문 No	수주 No	보증 No

<표 1>에서 백엔드데이터는 일반적인 B2B 전자 상거래 사이트에서 사용되는 데이터베이스를 나타낸

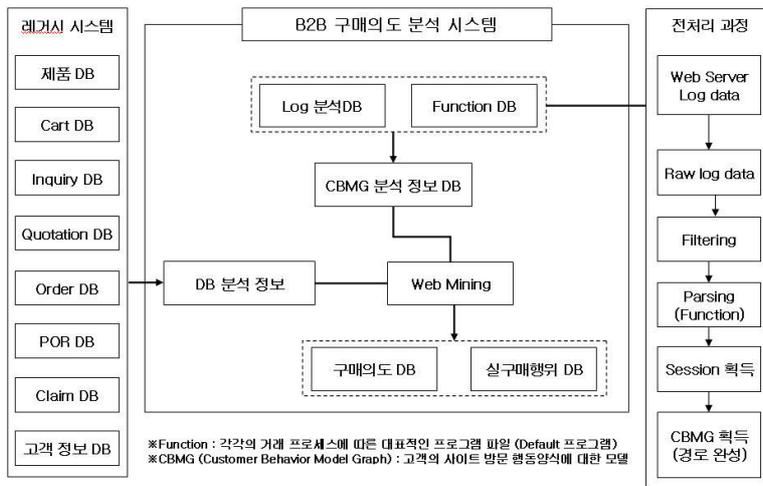


그림 2. B2B 구매의도 파악 모형

에 cs-uri-query에 해당 질의어가 기록이 된다는 것이고, post 방식은 기록이 되지 않는다는 것이다.

웹 로그 데이터 분석에 필요한 로그 데이터 항목은 W3C 확장 로그파일 형식에 정의된 기본값을 사용한다. 정의된 기본값 중 필요한 항목은 날짜, 시간, client ip 주소, uri-stem, uri-query이다. uri-stem은 client가 서버에 요청한 파일명을 의미하고, uri-query는 client가 서버에 요청한 파일에 따른 질의어를 의미한다.

예를 들어 고객이 “32420”이라는 제품 코드를 가지고 검색하였다면, get 방식의 로그데이터에는 “uri-stem : uri-query = search.asp : query=32420”와 같이 기록되며, 브라우저 주소 창에는 “http://abc.com/search.asp?query=32420”와 같이 나타나게 된다. post 방식은 “uri-stem : uri-query = search.asp : -”으로 웹 로그에 기록되며, 브라우저 주소 창에는 “http://abc.com/search.asp”와 같이 나타난다.

본 연구에서는 웹 로그 분석의 효율성을 위해서 전체 거래프로세스의 각 기능을 get 방식으로 정의한다. 백엔드데이터는 거래 프로세스에 따른 데이터를 사용하게 된다. <표 1>은 거래 프로세스에 따른 백엔드데이터와 get 방식의 질의 항목을 보여준다.

다. 기업의 백엔드데이터는 세부적인 거래 프로세스에 따라 조금씩 차이는 있지만, 전체 프로세스를 놓고 보면 <표 1>의 범주 안에 포함된다.

본 연구에서는 W3C 확장 로그파일 형식을 사용한다. 로그파일은 서버 관리자가 정의한 방법에 따라 일자별로 생성된다. 생성된 로그파일은 분석의 효율성을 위해 로그분석 DB에 원하는 로그 데이터가 기록되고 난 후 삭제된다. 웹 로그 분석을 위해서 로그 데이터를 적절한 형태로 가공하는 전처리 작업이 필요하다[11].

일자별로 로그 분석 DB에 저장된 로그 데이터에서 고객의 거래정보를 추출하기 위해 <그림 3>과 같은 획득 모형을 이용한다. 그림에서 merge and filter 단계는 데이터 추출의 효율성을 가져오기 위해 각각의 세부 단계로 나누어서 실행한다.

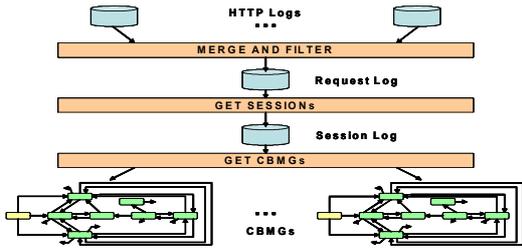


그림 3. CBMG 획득 모형

CBMG는 접속한 고객에 따라서 여러 가지 행태를 가질 수 있다. 전자상거래 사이트에서 CBMG 분석을 통한 로그분석은 고객의 구매행태 등을 파악하는데 더 높은 신뢰성을 제공한다. CBMG 분석과정은 Filtering, Parsing & Merging (Function), Session 획득, CBMG 획득 과정으로 이루어진다 [12].

1) Filtering 과정

filtering 과정은 웹 로그파일에서 불필요한 그림 파일이나 css, js 파일 등을 제거한다. 이것은 다음 단계에서 수행할 parsing & merging 단계의 효율성을 증가시키기 위한 과정이다.

2) Parsing & Merging (Function) 과정

parsing & merging (function) 과정은 로그파일에 요청된 콘텐츠 파일명이 어느 콘텐츠에 대한 요청인지를 쉽게 구분하기 위한 것이다. 이 과정에서 filtering 과정을 통해서 정제된 로그파일은 거래 프로세스에 따른 function으로 변환된다. <그림 4>는 거래 프로세스에 대한 function을 정의한 예이다. <그림 4>의 거래 프로세스 function은 filtering된 로그파일을 parsing 시키면 <그림 5>와 같은 로그파일을 얻을 수 있다. 또한 parsing 단계에서는 거래 프로세스 function에 정의되지 않은 로그파일을 제거한다. 이 단계를 통해 정제된 로그파일은 거래 프로세스 function에 따라서 변경된 로그파일이 된다. 이렇게 parsing된 로그파일을 request log라고 불리기도 한다[12].

PROC	LOG	FUNCTION	DESC
A	/DEE/A2A00.ASP	SEARCH	제품 검색
A0	/DEE/A2A800.ASP	SEARCH_VIEW	제품 검색 후 상세 보기
B	/DEE/A2AC00.ASP	INQUIRY	견적 의뢰
C	/DEE/A2AD00.ASP	QUOTATION	견적서 보기
D	/DEE/A2AE00.ASP	ORDER	주문
E	/DEE/A2AF00.ASP	POR	수주
Z	/DEE/A2AFT0.ASP	호선정보	호선에 대한 기본 정보

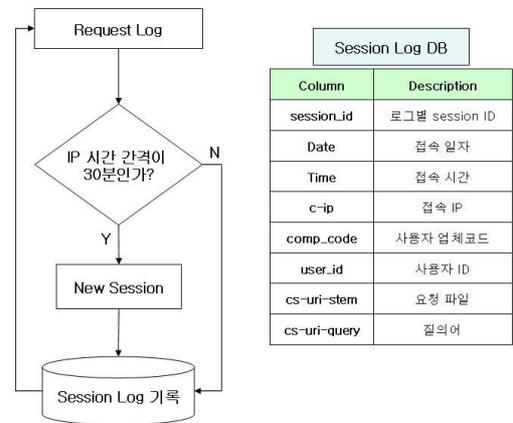
그림 4. Function의 예



그림 5. Parsing 된 로그파일

3) Session 획득 과정

이 과정은 각 사용자의 데이터들을 식별하는 사용자 식별, 각 사용자들의 정보를 세션으로 식별하는 세션식별 과정을 걸쳐게 된다. 웹 서버는 접속한 클라이언트의 브라우저에 유일한 세션값을 할당하게 된다. 동일한 클라이언트에서 동시에 2개의 브라우저로 접속을 하더라도 각각 다른 세션값을 할당 받는다. 세션의 종료는 클라이언트가 브라우저를 종료하거나 다른 웹 사이트로 이동을 할 경우이다. 또한 클라이언트의 브라우저에서 일정 시간 반응이 없으면 서버는 클라이언트의 세션을 종료시킨다. 일반적으로 세션은 20분 동안 클라이언트에서 반응이 없으면 종료되게 된다. 보통 접속 ip를 통해서 사용자들의 세션을 구분하지만, 본 연구에서는 접속 ip와 더불어 get 방식을 통해 얻은 사용자 정보를 통합해서 사용한다. 즉 질의어에 업체 코드와 user id 등을 같이 보냄으로써, 로그파일에 해당 접속 ip와 더불어 해당 사용자의 정보를 쉽게 파악하기 위해서이다. <그림 6>은 session을 얻는 알고리즘으로, session log는 사용자별 거래 행태를 나타낸다.



※SessionID는 "YYYYMMDD" + 업체코드 + 일련번호로 정의한다.

그림 6. Get session logs

4) CBMG 획득 과정

CBMG를 획득하기 전에 경로완성 과정을 먼저

수행해야 한다. 경로 완성은 한 세션에서 클라이언트의 브라우저와 프락시 서버의 캐시로 인해 로그파일의 중간 중간에 기록되지 않은 웹 페이지들을 채워주는 과정이다. 경로완성을 하기 위해서는 사이트 내에서 고객이 행동을 보일 수 있는 CBMG를 먼저 작성해야 한다. CBMG는 고객이 사이트 내에서 원하는 작업을 하기 위해서 이동할 수 있는 경로라고 이야기 할 수 있다. <그림 7>은 B2B 표준 거래 프로세스에 따른 CBMG를 보여준다.

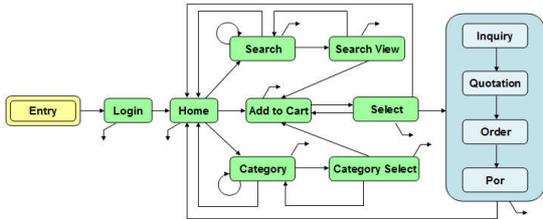


그림 7. B2B 표준 거래 프로세스의 CBMG 예

위의 CBMG에서 화살표가 각 function과 연결되어 있지 않는 것은 이탈을 의미한다. 여기서 이탈은 다른 사이트로 이동을 하였거나, 브라우저를 종료시키는 등의 경우에 해당한다. 위의 CBMG를 통해서 session log 데이터에서 경로가 완성되지 않은 session 들을 논리적으로 추가/삭제해야 한다. <그림 8>은 이러한 경로 완성에 대한 예를 보여준다.

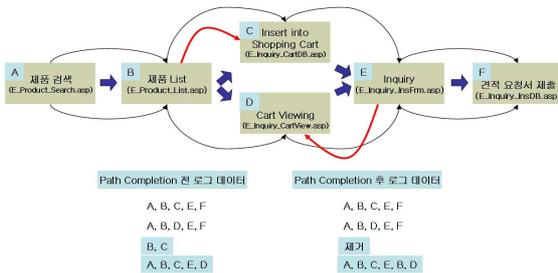


그림 8. 경로를 논리적으로 재구성하는 사례

<그림 8>의 논리적인 경로완성 단계를 통해서 최종적으로 session 별로 완성된 CBMG를 얻을 수 있다. 이렇게 얻은 session 별 CBMG와 기업의 백엔드시스템 데이터를 연결하여 “미구매행위”, “실구매행위” 등의 정보를 획득할 수 있다. 이를 위해서는 웹 로그에 기록하는 방법을 get으로 설정해야 하고, 기록되는 내용의 형식을 정해야 한다. 즉 질의어를 통해서 보낼 내용에 대한 정의를 선행해야 한다. <표 2>는 각 거래 프로세스 단계에서 질의어에 대해 정의한 예를 보여준다.

표 2. CBMG 데이터에 기록된 모습

화면	기능	Function (cs-uri-stem)	질의어 (cs-uri-query)
shopping cart ⇒ 견적의뢰	견적의뢰 (B)	/DEE/A2AC00.ASP	COMP_CODE=D0412&USER_ID=WOORISYS&PARTS=7RT84A,543043,1,7RT84A,653040,1,7RT80A,45300,1
견적서 ⇒ 주문	주문 (D)	/DEE/A2AE00.ASP	COMP_CODE=D0412&USER_ID=WOORISYS&PARTS=7RT84A,543043,1,7RT84A,653040,1,7RT80A,45300,1
주문 ⇒ 수주	수주 (E)	/DEE/A2AF00.ASP	COMP_CODE=D0412&USER_ID=WOORISYS&PARTS=7RT84A,543043,1,7RT84A,653040,1,7RT80A,45300,1

<표 2>의 거래 프로세스에 따른 URL이 CBMG 데이터에 기록된 모습을 보면 <표 3>과 같다.

표 3. CBMG 데이터에 기록된 모습

화면	기능	Function (cs-uri-stem)	질의어 (cs-uri-query)
shopping cart ⇒ 견적의뢰	견적의뢰 (B)	/DEE/A2AC00.ASP	COMP_CODE=D0412&USER_ID=WOORISYS&PARTS=7RT84A,543043,1,7RT84A,653040,1,7RT80A,45300,1
견적서 ⇒ 주문	주문 (D)	/DEE/A2AE00.ASP	COMP_CODE=D0412&USER_ID=WOORISYS&PARTS=7RT84A,543043,1,7RT84A,653040,1,7RT80A,45300,1
주문 ⇒ 수주	수주 (E)	/DEE/A2AF00.ASP	COMP_CODE=D0412&USER_ID=WOORISYS&PARTS=7RT84A,543043,1,7RT84A,653040,1,7RT80A,45300,1

이렇게 다음 단계로 링크를 통해 이동을 할 때 get 방식으로 전송을 하기 위해서 일련의 규칙을 정해 놓았다. 이를 토대로 시스템 구현을 하게 된다. <표 3>에서 “질의어” 항목의 COMP_CODE는 “업체 코드”, USER_ID는 “사용자 ID”, PARTS는 “제품 타입, 제품 코드, 수량”의 형태로 정의한 것이다.

3.2 DB 분석 정보

기업의 레거시 시스템에는 기업간 거래에서 발생한 트랜잭션에 대한 모든 정보를 가지고 있다. 트랜잭션 데이터는 방대한 양의 데이터를 가지고 있고, 이 데이터는 사용자들이 바로 사용할 수 있게끔 제공되지 않는다. 이러한 트랜잭션 데이터를 일반 사용자들이 쉽게 이용할 수 있게끔 특정한 목적에 맞게 재구성하는 것이 데이터베이스 분석이다. 일반적으로 DB 분석을 하기 위해서는 데이터웨어 하우스 또는 데이터 마트 등의 솔루션을 이용한다. 하지만 이러한 솔루션은 비용이 많이 들고 웹 로그 데이터를 포함한 분석이 미흡하다는 단점이 있다. 따라서 본 연구에서는 구매의도 파악과 연관된 질의를

통해 별도의 데이터베이스를 만든다. 데이터베이스 질의는 뷰(View) 또는 프로시저(Procedure)로 작성된다. DB 분석 정보를 별도의 데이터베이스로 만드는 것은 레거시 시스템의 Raw 데이터를 직접 이용함으로써 오는 오류 및 데이터베이스 시스템의 안정성을 유지하기 위해서이다. DB 분석 정보는 "Quotaion DB", "Order DB", "POR DB", "Claim DB" 등으로 요약되어 진다.

3.3 구매의도 파악 정보

앞에서 얻은 "CBMG 분석정보"와 "DB 분석정보"를 이용하여 구매의도 정보를 얻을 수 있다. 다음 <그림 9>는 "구매의도"를 획득하기 위한 모형이다.

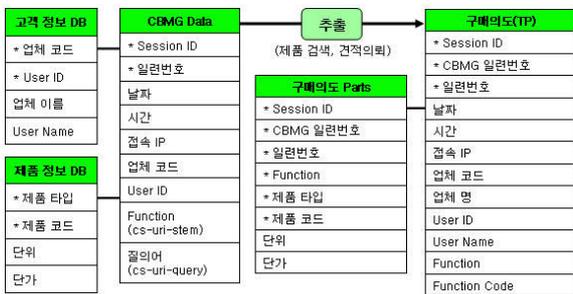


그림 9. 구매의도 획득 모형

구매의도는 구매행위로 나타나므로 CBMG 데이터에서 제품검색과 견적의뢰 정보만을 추출하면 된다. 그러나 제품검색의 경우는 동일한 품목에 대해서 반복해서 검색을 할 수 있다. 따라서 동일한 세션에서 동일한 품목을 반복적으로 검색 할 경우는 하나의 품목을 검색한 것으로 규정한다. 이와 같은 규칙을 적용하여 구매의도를 살펴보면 "구매의도 = 제품 검색 + 견적의뢰" 로 나타낼 수 있고, CBMG 데이터에서 function이 "제품 검색(A, A0)"과 "견적의뢰(B)"인 것만을 추출하면 된다.

"구매의도" 정보를 추출하였으면, CBMG 데이터에서 "실구매행위" 정보를 추출해야 한다. 실구매행위 정보는 우선 거래 프로세스에서 "견적의뢰(Inquiry) - 견적서(Quotation) - 주문(Order) - 수주(Por)"에 해당하는 function을 추출한다. 그리고 거래 프로세스에 따른 function 중 "주문" 단계에 있는 로그 데이터를 추출하면 그 정보가 실제 거래로 이루어진 정보가 된다. <그림 10>은 "실구매행위" 정보를 추출하는 모형이다.

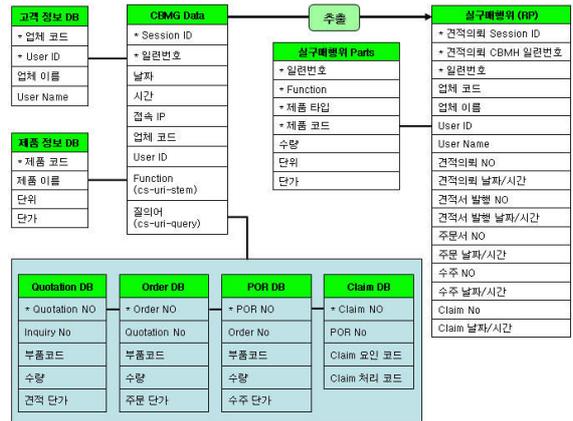


그림 10. 실구매행위 획득 모형

실구매행위 획득 과정은 다음과 같다. 먼저 고객은 견적의뢰와 주문의 function을 이용하고, 해당 기업 담당자는 견적서 발행과 수주 확정 function을 이용한다. CBMG 로그에서 각 function에 따라서 실구매행위 DB에 기록하는 방법은 조금씩 다르다. 각 function에 대한 추출 방법을 살펴보면 다음과 같다.

- 견적의뢰 : 견적의뢰는 구매의도 DB에 바로 기록한다.

- 견적서 : 견적서는 기업의 백엔드시스템 데이터의 Quotation DB에서 해당 견적서 번호를 통해서 견적의뢰 번호를 알아온다. 그 다음에 구매의도 DB에서 견적의뢰 번호에 해당하는 데이터를 찾아서 수정/삽입한다.

- 주문 : 주문은 주문 번호를 통해서 견적서 번호, 견적의뢰 번호를 찾은 다음 실구매행위 DB에서 견적의뢰 번호에 해당하는 데이터를 찾아서 수정/삽입한다.

- 수주 : 수주는 수주 번호를 통해서 주문 번호, 견적서 번호, 견적의뢰 번호를 찾은 다음 실구매행위 DB에서 견적의뢰 번호에 해당하는 데이터를 찾아서 수정/삽입한다. 이렇게 수정된 실구매행위 DB는 그 자체만으로 고객의 수익성, 거래이력 등의 정보를 획득할 수 있다.

4. 적용 사례

4.1 대상 기업

본 연구에서는 "H"사의 e-Service 시스템에 적용한 사례를 보인다. "H"사는 세계 최대의 선박용 엔진 생산기업으로, 글로벌 환경에서 기업경쟁력 강화를 위한 전략적 수단으로 e-Service 시스템을 구축하였다. e-Service 시스템은 선박용 엔진부품의 판매 및 조달 그리고 판매된 엔진의 사후관리 등을 위한 목적으로, 엔진부품 구매업체와 부품 및 원자재의 공급업체, 그리고 사후관리를 위한 용역업체 등 선박용 엔진에 관련되는 다양한 고객을 대상으로 한다.

4.2 구매의도 분석 시스템

“H”사는 지금까지 거래이력 정보를 기준으로 데이터베이스 질의를 통해서 “견적 대비 수주”, “수주 대비 발주”, “납품 대비 클레임”, “납품 대비 수금” 등의 정보를 추출하여 영업활동에 사용하였다. 하지만 “견적 대비 수주”의 비율이 현저하게 낮다는 것을 파악하였고, 또한 견적의뢰를 하지 않고 단지 제품만 검색하는 비율이 높다는 것을 웹 로그 분석을 통해서 파악하였다. 따라서 “H”사는 향후 영업 전략의 근간을 세우기 위해서 구매의도를 가지고 사이트에 접속하였지만, 실제 구매로 이어지지 않은 제품에 대한 파악 및 고객의 구매 패턴 분석을 통해서 시장 장악을 꾀하는 전략을 추구하게 되었다. 이를 위해서 레거시시스템 데이터와 웹 로그 데이터를 통합한 새로운 구매의도 분석 정보가 필요하게 되었다. 이를 위해 “H”사는 B2B 구매의도 파악 시스템을 구축하였다.

본 연구에서는 “H”사의 e-Service 시스템에서 기업간 거래 시스템을 이용하는 고객을 대상으로 웹 로그 분석을 실시한다. “H”사의 e-Service 시스템은 기업정보시스템과 연동되어 실시간 정보를 제공하며, 고객은 인터넷을 통해 기업간 비즈니스를 처리한다. “H”사의 웹 서버는 IIS 5.0을 사용하고, 시스템 운영체제는 Windows 2000 Server이다. 기업의 백엔드시스템은 오라클 데이터베이스를 사용하고, 실시간 정보제공을 위해 ODBC Connection으로 연동되어 운영된다. 웹 서버에서 생성되는 로그파일은 MS-SQL 7.0 서버에 저장되고, 로그파일 기록은 IIS 5.0의 웹 로깅 저장 방법 중 ODBC 연동을 통해 자동으로 MS-SQL 서버로 기록된다.

“H”사는 B2B 구매의도 분석 시스템을 위해 기존의 e-Service 시스템에 다음 <그림 11>과 같은 시스템을 구축하였다. 웹 서버의 웹 로그 데이터는 파일에 기록되지 않고 데이터베이스에 저장된다. 이것은 분석의 효율성을 높이기 위해서이다.

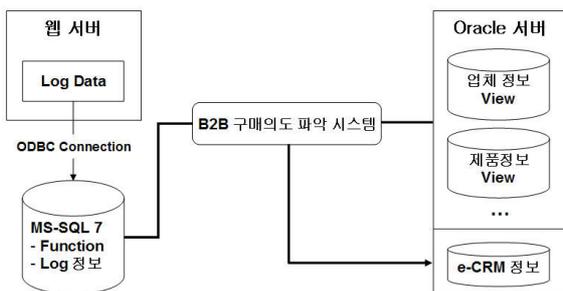


그림 11. “H”사 시스템 구조

B2B 구매의도 파악 시스템을 통해서 “구매의도”와 “실구매행위” 정보를 얻게 되었다. 이를 통해서 고객별, 제품별 구매의도를 파악할 수 있다. <그림 12>는 “업체별 구매율”을 나타낸다.

업체별 구매율

조건 : [분기별 집계] 기간 : 2003년 1사분기 [조회]

날짜	업체 코드	업체명	구매	미구매	비율
2003년 1사분기	D001	동진산업기술(주)	2,400	2,500	48.98%
	D002	케이프기공(주)	1,100	1,000	52.38%
	D003	J.O ENGINEERING	1,200	1,900	38.71%
	D004	대화엔지니어링 서비스(주)	900	1,000	47.37%
	D005	오션마린서비스(주)	2,000	2,200	47.62%
	D006	현대상선(주)	1,500	800	65.22%
					49.19%

최근 6개월간 “동진산업기술(주)” 평균 구매비율

2003-3	2003-2	2003-1	2002-12	2002-11	2002-10
38.34%	43.23%	30.23%	35.03%	46.22%	39.44%

그림 12. 업체별 구매율

업체별 구매율에서 미구매율은 제품검색과 견적의뢰를 한 제품이 구매로 이루어지지 않은 비율을 나타낸다. <그림 13>은 부품별 구매율을 나타낸다.

부품별 미구매율

조건 : [분기별 집계] 기간 : 2003년 1사분기 [조회]

Ship Selection	Project No	Ship Name	Shipowner	Ship Operator	Hull No	Engine Type	Engine Delivery	Ship Delivery
		: AROSIA	: 현대상선(주)	: 현대상선(주)		: 6L70MC	: 1984.08.23	: 1985.06.12

Engine Type	Part Code	Part Name	구매	미구매	비율(%)
6L70MC	90101-30K-018	DISTANCE PIECE	2,400	2,500	51.02%
6L70MC	90101-30K-055	CYLINDER COVER	1,100	1,000	47.62%
6L70MC	90101-30K-082	HEXAGON SCREW	1,200	1,900	61.29%
6L70MC	90101-30K-199	STUD FLEXHAUST VALVE	900	1,000	52.63%

항 목 : [기간별 평균 미구매비율] 최근 6개월간 “90101-30K-018” 평균 미구매비율

2003-3	2003-2	2003-1	2002-12	2002-11	2002-10
47.44%	54.32%	39.13%	64.03%	36.22%	31.44%

그림 13. 부품별 미구매율

부품별 구매율은 월별, 분기별 등으로 분석하여 해당 제품의 구매행태를 파악하게 된다. 이렇게 파악된 정보는 영업 전략 및 재고 유지 계획의 일부로 활용하게 된다.

7. 결 론

지금까지 B2B는 고객 관리를 위한 방법으로 고객의 구매량, 고객의 전략적 가치, 고객의 수익성에 기반을 두었다. 그러나 기업의 업무 프로세스가 변화하고 인터넷 의존도가 더욱 높아짐에 따라, 이러한 정보만으로는 고객에 대한 평가를 제대로 하기 어렵다. e-Business 시대의 효과적인 고객 평가 및 관리는 기존의 오프라인 정보뿐만 아니라, 온라인 정보를 포함한 분석이 이루어져야 한다. 이러한 온라인 정보에서 가장 중요한 것은 고객이 웹 사이트에서 보인 구매의도를 분석하는 것이다. 따라서 본 연구에서는 B2B 환경에서 고객의 구매의도를 파악할 수 있는 방안으로 웹 로그 데이터 분석 방안을 제시하였다.

본 연구는 B2B 기업의 구매의도를 파악하기 위해서 기업의 레거시시스템 데이터와 웹 로그 데이터를 이용하였다. 이를 통해 웹 사용자들의 구매의도를 파악할 수 있었다. 구매의도를 파악하기 위한 방안으로 B2B 구매의도 파악 방안을 제시하였다. 제시된 방안은 웹 로그 데이터를 전처리 과정을 통해

실제 활용이 가능한 CBMG 정보로 추출하였고, 기업의 레거시 시스템 데이터를 데이터베이스 질의를 통해서 뷰 형태의 정보로 추출하였다. 이렇게 추출된 CBMG 정보와 기업이 레거시시스템 데이터를 이용하여 구매의도 정보를 파악하였다.

본 연구는 B2B 기업 중 구매기업에 대한 연구이다. 향후 기업의 공급사슬에 관련된 여러 기업 형태를 분석한다면, 기업은 업체별 세분화된 시스템을 구현할 수 있는 기초 자료를 얻을 수 있으며, 이를 통한 업체관리를 위한 더욱 많은 정보를 얻을 수 있을 것이다.

참고 문헌

- [1] 강태근, 전해성, 박상신, 이경미, 이종우, 김영국, 유관중, "Multiple Intelligent Agent Architecture for Personalized Advertisement and Information Services on Internet Shopping Mall", [한국경영정보학회 국제학술대회](1998), pp. 117-126.
- [2] 김원주, 김재일, 주유진, [마케팅정보론], 박영사, 1996.
- [3] 박형수, "일대일 마케팅을 위한 웹 로그 분석 Preprocessor 설계 및 구현", [한국과학기술원], 석사논문(2000)..
- [4] 변숙은, "Log Data 분석을 통한 웹 사이트의 효율적 관리 방안", [한국과학기술원], 석사논문(2001).
- [5] 이재신, "데이터마이닝을 활용한 동적인 고객 분석에 따른 고객관계관리 기법", [한국지능정보시스템학회], 9권, 3호(2003).
- [6] 이화영, "표준 로그파일을 이용한 웹 마이닝에 관한 연구", [한국과학기술원], 석사논문(1999).
- [7] 장윤경, "로그분석을 통한 고객행동과 고객 가치와의 관계에 관한 연구", [아주대학교], 석사학위논문(2002).
- [8] 전형철, "로그파일을 이용한 인터넷 쇼핑물 사용자의 구매행동 연구", [한국과학기술원], 석사논문(2002).
- [9] 최덕원, "전자상거래에서 상품 추천을 위한 웹 개인화 방안에 관한 연구", [한국경영과학회, 춘계학술대회 논문집](2001).
- [10] Borges, J. and M. Levene, "Mining Navigation Pattern Discovery from Internet Data, ", <http://www.almaden.ibm.com/cs/quest>(1999).
- [11] Cooley, R., Mobasher, Bamshad, Srivastava, J., "Data Preparation for Mining World Wide Web Browsing Patterns", *The Journal of Knowledge and Information Systems*, Vol. 1, No. 1(1999).
- [12] Daniel A. Menasce & Virgilio A.F. Almeida, *Scaling for E-Business : Technologies, Models, Performance, and Capacity Planning*, Prentice Hall, 2000.
- [13] Fu, Y., K. Sandhu and M.Y. Shih, "Clustering of Web Users Based on Access Patterns", *Proceedings of the Workshop on Web Usage Analysis and User Profiling (WEBKDD'99)*, San Diego, CA, August(1999).
- [14] Igor V. C., Scott G. and P. Smyth, "A general probabilistic framework for clustering individuals and objects", *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining(2000)*, pp. 140-149.
- [15] Mannila, H. and C. Meek, "Global partial orders from sequential data", *Proc. of the ACM SIGKDD international conference on Knowledge discovery and data mining(2000)*, pp. 161-168.