

이동평균 개념을 이용한 웹 사이트 사용자 관심도 예측 시스템

박기현* · 유상진**

A Prediction System on User Interest Degree to Web Sites
Using the Concept of the Moving Averages

KeeHyun Park* · Sangjin Yoo**

■ Abstract ■

Now that many organizations have invested a tremendous amount of money and efforts to operate Web sites on the Internet, there is a strong demand to understand the effectiveness of such investments. In other words, one of most frequent and important questions about their Web sites is "Will the current Web site management policy be effective enough to have more visitors come to our Web site?"

In this paper, a system which predicts the degree of user interest in the future to Web sites is constructed. The degree of user interest to a Web site is defined to be the visit counts for the Web site in the system. With higher the visit counts, the related site is considered to be more interesting. However, the figures of the visit counts themselves cannot explain properly the degree of user interest in the future to the related Web sites (i. e., the effectiveness of the related Web sites). Therefore, the system also uses mechanisms which use the concept of the Moving Averages, which have been used frequently in the stock exchanges.

In this paper, two prediction mechanisms are proposed and compared. The first mechanism uses the Golden Cross/the Dead Cross of the Moving Averages, while the second mechanism uses the changes of upward/downward direction of the Moving Averages. Experimental results show that the two prediction mechanisms proposed in this paper predict the degree of user interest in the future to the related Web sites very well in most cases. However, the first one is considered to be better than the second one in the sense that the second one is too much sensitive to the changes of visit counts.

Keyword : Website Management, User Internet, Moving Average

논문접수일 : 2001년 9월 17일 논문게재확정일 : 2002년 4월 30일

* 계명대학교 정보통신대학 정보통신학부

** 계명대학교 경영대학 경영정보학과

1. 서 론

최근들어 급속하게 일어나고 있는 경영환경의 변화는 새로운 패러다임의 정립을 요구하게 되었고, 새로운 패러다임의 정립을 위해서는 신속하고 정확한 정보의 제공이 필연적이다. 이러한 필요성 및 추세는 인터넷 웹 사이트가 기업이 경영에 필요로 하는 정보의 제공을 주도하는 위치에 설 것을 강요하게 되었고, 이에 따라 기업들은 웹 사이트가 필요한 정보의 수집은 물론 기업의 수익창출 도구로서의 역할을 수행할 수 있도록 웹 사이트의 운용에 더욱 많은 경비와 인력을 투입하여 사용자들이 되도록 자주 방문할 수 있도록 노력하고 있다. 그러나, 단순히 많은 노력을 들였다고 해서 해당 웹 사이트가 효율적으로 운영되는 것은 아니기 때문에, 웹 사이트의 사용자 방문횟수를 지속적으로 점검하여 방문자들이 얼마나 관심을 가지고 있는가에 대한 정보를 원하는 기업들이 늘어나고 있다. 이러한 요구에 부응하기 위하여 최근에는 웹 사이트의 효율적인 운영과 방문자들의 관심도를 측정하고 분석하기 위한 웹 분석 기법들이 개발되고 있다[1, 2, 5, 8, 13-16]. 이런 기법들은 사용자들의 방문 횟수나 체류 시간 등을 점검함으로써, 체류 기간, 방문횟수, 관심 웹 페이지 등에 대한 정보를 제공하며, 이러한 웹 분석은 웹사이트의 효율성을 검토하는데 중요한 척도로 사용되고 있다. 그러나, 이런 분석은 과거의 웹사이트 효율성을 점검하는 것으로서, 미래의 웹사이트 효율성을 예측하기는 어렵다. 웹 사이트의 관리자 등은 과거의 실적보다는 미래의 사용자 관심도에 더욱 관심을 가지고 있으므로, 현재의 웹 사이트 분석기법들을 개선할 필요가 있다.

본 논문에서는 웹 사이트에 대한 사용자들의 관심도 변화를 예측할 수 있는 방법을 제안하고 이를 이용한 예측 시스템을 구축한다. 즉, 웹 사이트에 대한 사용자들의 미래의 관심방향을 예측함으로써, 그 사이트의 기존 운영정책을 계속 존속시킬 것인지, 혹은 운영정책을 새로이 수립해야 하거나 콘텐츠 혹은 웹 페이지의 화면 구성 등을 개편할 시점

인지를 판단하는데 필요한 기본 정보를 제공하는 예측 방법을 제안하고, 이 방법을 근간으로 한 예측 시스템을 구현한다. 본 논문에서는 사용자들의 관심도를 방문횟수로 나타낸다. 즉, 방문횟수의 증가는 관심도의 증가를 의미하므로, 해당 웹 사이트의 가치가 높아졌다고 판단될 수 있다. 반대로, 방문횟수가 감소하면 관심도가 감소되어 해당 웹 사이트의 가치가 낮아졌다고 판단될 수 있다.

본 논문이 제안한 예측 방법은 주식 매매에 사용되는 이동평균(Moving Average)[3, 4] 개념을 이용하는데, 특정 웹 사이트의 향후 방문횟수의 방향에 대한 예측이 가능하도록 하기 위하여, 주식 가격(주가) 대신에 방문횟수를 사용한다. 주식 매매에 있어서 가장 중요한 것은 주가의 변화이다. 결국 여러 가지 기술적인 지표들을 이용하여 예측하고자 하는 것은, 과거의 경험에 비추어 앞으로의 주가가 어떻게 변화될 것인가에 대답을 얻고자 하는 것이다. 주가가 오르면 해당 기업의 가치가 높아지는 것과 마찬가지로, 특정 웹 사이트의 방문횟수가 늘어나면 해당 웹 사이트에 대한 관심도가 높아지는 유사점이 있으므로, 본 논문에서는 예측 시스템의 개발에 주식 매매에서 사용하는 주가 예측기법들을 변형하여 사용한다. 본 논문에서 제안하는 예측 시스템의 논리(logic)는 방문횟수에 대한 일봉/주봉 차트 및 이동평균 등에 기초하고 있으며 개발도구로는 C 언어를 사용하였고, MS Windows 운영체제에서 작동하는데 기본적으로 사용자가 웹 브라우저(Browser)를 이용하여 웹 분석 도구를 사용하고 결과를 조회할 수 있도록 하였다.

웹 사이트들의 로그(log) 화일을 이용하여 본 논문에서 개발한 예측 시스템을 실험한 결과 방문횟수에 대한 이동평균을 이용한 기법의 예측 능력이 매우 우수함을 알 수 있었다. 본 시스템에서는 2가지 예측 방법들을 제안하는데, 첫 번째는 방문횟수 이동평균선들의 골든크로스(Golden Cross)와 데드크로스(Dead Cross)를 이용하는 방법이며, 두 번째는 방문횟수 이동평균이 상승 전환하는 시점과 하락 전환하는 시점을 이용한 예측 방법이다. 첫 번째 예측

방법은 대부분의 경우에 방문 횟수의 상승과 하락을 예측하여서 예측 능력이 매우 우수함을 알 수 있었다. 두 번째 방법은 예측 능력이 우수하지만, 작은 변화에 대해서도 너무 민감하게 반응하기 때문에 실용적인 측면에서 개선되어야 할 점들이 있었다.

본 논문은 다음과 같이 구성된다. 2절에서는 웹 분석 및 주가 예측 기법들에 대한 관련 연구를 살펴보고, 3절에서는 본 논문에서 제안한 사용자 방문횟수 예측 기법을 바탕으로 한 웹 분석 시스템에 대한 구현에 대해 설명한다. 4절에서는 실제 자료에 근거하여 수행한 실험에 대하여 결과를 분석하고, 마지막으로 5절에서는 본 논문의 결론을 다룬다.

2. 관련 연구

2.1 웹 분석 도구

현재 다양한 웹 분석 도구들이 개발되고 있다. 웹 분석 도구란, 웹 페이지에 대한 방문자들의 사용 빈도, 사용 시간, 사용 행동 등을 분석하여 의사결정에 도움이 되도록 하는 인터넷 소프트웨어이다.

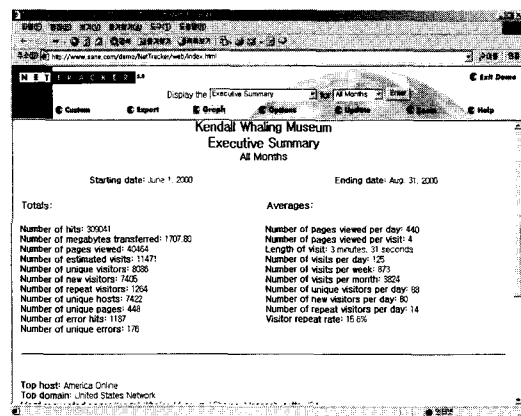
2.1.1 NetTracker

Sane Solutions 회사에서 만든 log file 분석 도구로서, 웹 사이트의 흐름(Traffic) 보고 및 웹 마이닝(Mining)을 수행한다[13]. Nettracker는 다음과 같은 흐름 분석을 한다.

- 개인 방문자의 클릭 흐름(Click stream) 분석 : 개인 방문자가 웹 사이트의 어느 곳을, 어떤 순서로 마우스를 클릭하면서 돌아다니는지 분석한다.
- 방문자 분석 : 반복 방문자, 새 방문자 등에 대한 분석을 통하여 웹 사이트가 방문자를 얼마나 오랫동안 머물게 하는지를 분석한다.
- 웹 사이트 유용성 분석 : 방문자가 요구한 웹 페이지에 대한 응답시간 등을 분석한다.

- 쿠키(Cookie) 번역 : 쿠키를 통한 방문자의 신상명세(이름, E-mail 주소 등)에 대한 정보를 제공한다.
- 내용별 그룹화 및 인기도 분석 : 방문자가 관심을 두는 내용을 분석하여 방문자들을 그룹으로 묶어서 인기도(Popularity)를 분석한다.
- 방문자 분류 : 방문자의 프로파일(Profile)을 작성하고 웹 사이트 내에서의 행동을 분류한다.
- 배너(Banner) 광고 분석 : 배너를 이용한 광고 효과를 분석한다.
- 수입 분석 : E-mail 광고, 배너 광고, 웹 사이트 연결(Link), 서치엔진(Search engine) 등이 웹 사이트에서의 구매와 직결되는지를 분석한다.
- 개인 방문자의 클릭 흐름(Click stream) 고급 분석 : 상품을 구매한 방문자와 그렇지 않은 방문자를 구분하여 그들의 웹 페이지 조회 경로를 분석한다.

<그림 1>은 Nettracker가 시행한 Kendall Whaling Museum에 대한 분석 화면을 보여준다.



<그림 1> NetTracker 분석 화면(예)

2.1.2 SurfReport

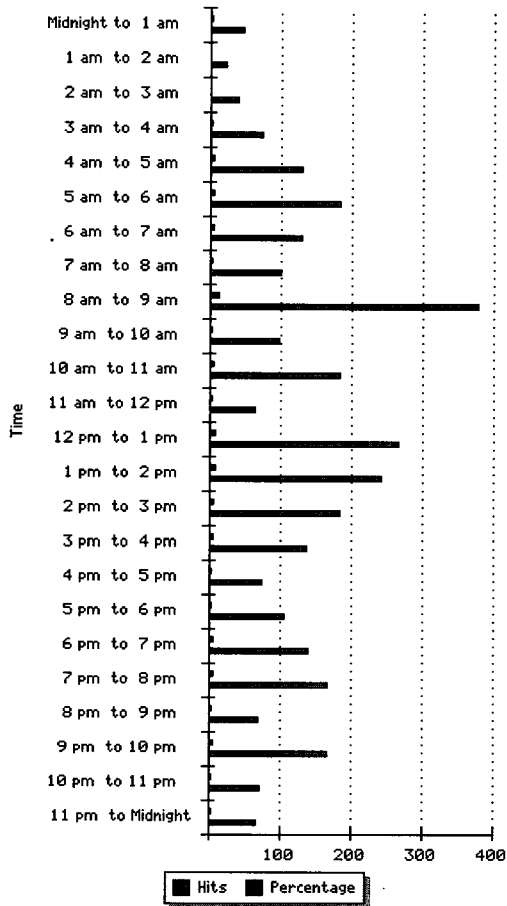
Netrics.com 회사가 개발한 웹 사이트의 흐름 분석 도구이다[10]. SurfReport가 분석 결과로서 보여주는 사항들은 다음과 같다.

Top 20 Companies Accessing NETRICS.COM's Site

3992	Unresolved Address
1282	netrics.com
478	proxy.aol.com
475	sd.intuit.com
429	connectnet.com
353	ix.netcom.com
326	cts.com
270	adnc.com
203	4dcom.com
198	copleypress.com
196	compuserve.com
153	apple.com
137	san-diego.ca.interraop.com
129	erols.com
109	paloar.edu
101	wantree.com.au
101	west.saic.com
98	cache.nlanr.net
97	sndg02.eacbell.net
96	inetworld.net

<그림 2> SurfReport 분석 화면1(예)

Hourly Hit Averages for SurfReport Test Site



<그림 3> SurfReport 분석 화면2(예)

- 웹 사이트 방문자 수 및 방문 횟수
- 웹 페이지 방문 회수
- 방문자별 평균 웹 페이지 조회 수
- 일별 평균 방문자 수
- 최다 방문자 Top X
- 최다 방문기업 Top X
- 최다 방문 국가 Top X
- 특정 파일 혹은 키워드 방문자
- 패턴매칭(Pattern matching) 및 필터링(Filtering) 기능

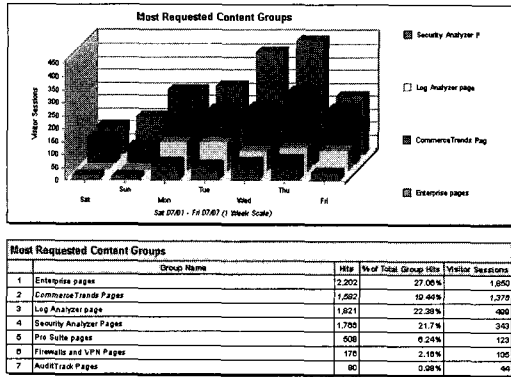
<그림 2> 및 <그림 3>은 SurfReport가 시행한 분석 화면을 보여준다.

2.1.3 Webtrends 제품

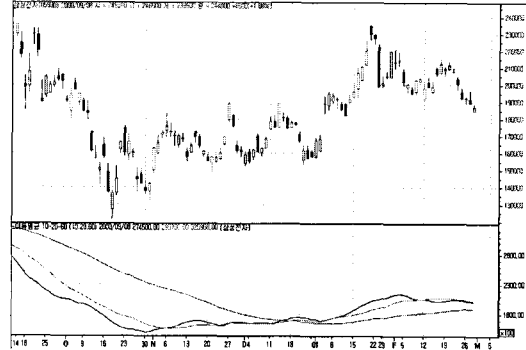
WebTrends 회사는 WebTrends Enterprise Reporting Server, WebTrends Log Analyzer, WebTrends Live 등을 개발하고 있다. 이들은 모두 웹 분석 도구로서, 다음과 같은 분석 결과를 보여준다[15].

- 웹 사이트 방문자 수 및 방문 횟수
- 웹 페이지 방문 회수
- 웹 사이트 체류 기간
- 메시지 전송량
- 1회 방문자
- 2회 이상 방문자
- 빈번하게 조회되는(진입) 웹 페이지
- 빈번하게 조회되는(진입) 웹 내용
- 빈번하게 download되는(동적) 파일
- 빈번하게 조회하는 경로
- 빈번하게 사용되는 동적 화일 변수
- 광고화면 조회 수
- 빈번하게 방문하는 조직, 국가
- 쿠키 정보 처리
- 본 웹 사이트를 경유하는 목적 웹 사이트
- DNS(Domain Name Service) 역조회

<그림 4>는 WebTrends Log Analyzer가 시행한 분석 화면을 보여준다.



<그림 4> WebTrends Log Analyzer 분석 화면(예)

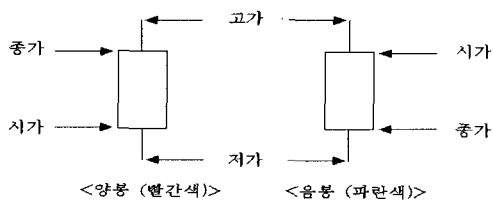


<그림 6> 주가 차트(일봉) 및 이동평균선들

2.2 주식 매매의 기술적 분석지표

2.2.1 주가 차트(Chart)

일반적으로 해당 기간의 주가 변화를 봉(Bar 혹은 Candle)으로 표현한다. 따라서 주가의 변화를 일별로, 주별로, 월별로 그래프를 이용하여 표현한 것을 각각 일봉 차트, 주봉 차트, 월봉 차트라고 부른다[3, 4]. <그림 5>는 한 개의 봉에 대한 설명을 나타낸 그림이다. 일봉의 경우에는, 양봉은 당일의 최종 가격이 시초 가격(시가)보다 높은 경우를 표현하며 빨간색으로 나타낸다. 이에 반하여 음봉은 당일의 최종 가격(종가)이 시초 가격보다 낮은 경우를 표현하며 파란색으로 나타낸다. 주봉의 경우에는 해당 주의 월요일 종가가 주봉의 시가로 나타나고 금요일 종가가 주봉의 종가로 표현된다. <그림 6>의 상단 그래프는 일봉 차트의 예를 나타낸다. 이 그래프에서 X축은 일자를 나타내고 Y축은 주가를 표현하며, 한 개의 봉은 당일의 주가 변화를 나타낸다.



<그림 5> 양봉 및 음봉

2.2.2 이동평균(Moving Average)

시장의 추세를 파악하기 위한 지표이다. 매일의 주가 흐름은 매우 불규칙하지만 이러한 일일주가를 며칠동안 묶어서 평균 주가를 구하면 일시적으로 급등하거나 급락하는 비정상적인 시장 흐름을 바로 잡아 시장의 큰 흐름을 알 수 있다는 생각에서 만든 지표이다. n일간의 이동평균은 n일간의 종가들의 합을 n으로 나눈 값이다. 이러한 이동평균값들을 그래프로 연결한 것이 이동평균선인데, 이것을 이용한 주식 매매 방법은 이동평균선의 교차를 이용하는 것이다. 단기 이동평균선이 장기 이동평균선을 상향 돌파(골든크로스, Golden Cross)할 때 매수하고, 단기 이동평균선이 장기 이동평균선을 하향 돌파(데드크로스, Dead Cross)할 때 매도한다. 골든크로스가 발생하였다 함은 최근의 가격 추세가 그 동안 하락하던 가격의 큰 흐름을 전환시키려고 한다는 것을 의미하기 때문에 시장의 중요한 질적 변화를 암시한다. 따라서 이동평균선의 교차 시점이 중요한 매매 타이밍이 되는 것이다[3, 4]. 단기 이동평균선이 장기 이동평균선을 상향 돌파한다는 것은, 특정 일자의 단기 이동평균값이 장기 이동평균값 보다 작다가(즉, 주가 차트에서 단기 이동평균선이 장기 이동평균선의 아래쪽에 위치하다가), 그 다음 날에 단기 이동평균값이 장기 이동평균값 보다 커지는 경우를 의미한다. 하향 돌파의 의미는 그 반대가 된다.

본 논문에서는 이 방법을 사용하여 특정 웹 사

이트에 대한 사용자 관심도의 향후 변화를 예측한다. <그림 6>의 하단 그래프는 상단 그래프의 일봉들에 대한 주가 이동평균선들(10일, 20일, 60일)을 보여준다. 이 그래프에서 가장 위쪽에 있는 것이 60일 이동평균선이고, 가장 아래쪽에 있는 것이 10일 이동평균선이다. <그림 2>에서는, 10일 이동평균선이 60일 이동평균선을 상향돌파하는 시점이 12월 13일이다.

3. 시스템 구성 및 구현

3.1 시스템 구성

본 연구에서 개발한 시스템은 로그화일 작성 모듈, 방문 횟수 통계처리 모듈, 방문 횟수 차트 처리 모듈 등으로 구분된다.

- 로그화일 작성 모듈 : HTTP(Hyper Text Transfer Protocol)[7]의 패킷(Packet)으로부터 방문에 관한 정보를 수집하여 로그 화일을 작성한다.
- 방문횟수 통계처리 모듈 : 작성된 로그 화일로부터 방문 정보를 읽어서 방문횟수를 생성한다.
- 방문횟수 차트처리 모듈 : 방문횟수 정보를 여러 형태의 차트들(주가 일봉/주봉 차트 및 이동평균선)로 표현한다. 사용자로 하여금 해당 방문자의 방문 특성을 차트로 표현함으로써, 방문자의 관심도 변화 추이 및 향후의 방문 회수를 어느 정도 예측할 수 있도록 한다. 방문횟수에 대한 골든크로스, 데드크로스, 상향 전환, 하향 전환 등의 발생을 감지하고 웹 사이트 관리자에게 알려주는 역할을 한다.

위에서 언급한 여러 가지 기술적 예측 지표들을 사용하기 위하여 본 연구에서는 주가 대신에 방문횟수를 사용한다. 즉, 주봉의 경우에는 해당 주의 첫 번째 요일(여기서는 일요일)의 방문횟수를 주봉의 시가로 대신하고, 마지막 요일(여기서는 토요

일)의 방문횟수를 주봉의 종가로 대신해서 사용한다. 물론 관리자의 의도에 의해서, 주봉의 시가와 종가에 해당하는 요일을 바꿀 수 있도록 구현되어 있다. 그러나, 일봉은 주봉과 다르기 때문에 봉 차트를 만들지 못한다.

주식 매매에 있어서 여러 가지 기술적인 지표들을 이용하여 예측하고자 하는 것은, 과거의 경험에 비추어 앞으로의 주가가 어떻게 변화될 것인가에 대답을 얻고자 하는 것이다. 한편, 웹 사이트 운영에 있어서 가장 중요한 것은 방문횟수의 변화이다. 방문횟수가 늘수록 해당 웹 사이트의 가치가 늘어나는 것이다. 따라서 주가 대신에 방문횟수를 사용하게 되면, 주식 매매의 여러 가지 지표들을 이용해서 앞으로의 웹 사이트 방문횟수의 변화에 대한 예측을 어느 정도 할 수 있으리라고 생각한다.

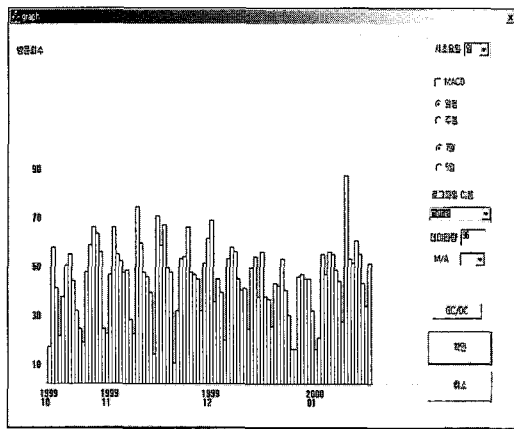
3.2 시스템 구현

본 시스템은 웹을 기반으로 동작하도록 구현되었다. 즉, 본 시스템은 웹 서버에 설치된다. 웹 서버는 웹 사이트를 운영하는 서버를 사용해도 무방하고 별도의 서버를 사용해도 된다. 웹 사이트 관리자는 웹 클라이언트의 일반적인 브라우저를 사용하여 본 시스템의 분석 및 예측 자료 등을 점검할 수 있다. 이를 위하여 CGI(Common Gateway Interface) 기법[6]을 사용한다. 앞에서 언급한 바와 같이 시스템을 구현하는데 사용한 프로그래밍 언어는 C언어이며, 운영체제로는 MS Windows를 사용하였다.

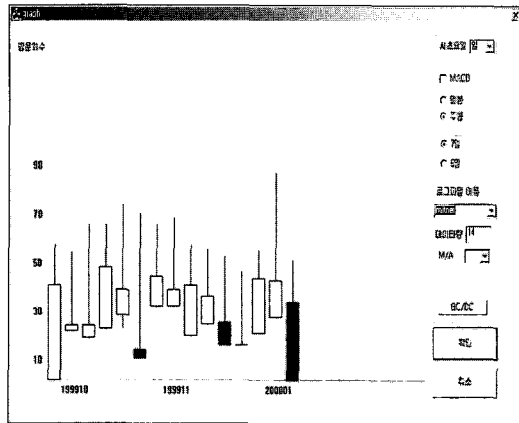
3.3 실험 및 결과 분석

본 연구를 위하여 2곳의 회사(회사 A 및 회사 B)들이 운영하는 웹 사이트 방문횟수 자료를 수집한 후, 본 연구가 제작한 시스템을 이용하여 실행하였다. 이 절에서는 본 시스템을 실행한 결과 화면들에 대해서 설명한다. <그림 7> 및 <그림 8>은 회사 A에 대한 방문횟수의 일봉 차트 및 주봉

차트를 각각 보여준다. 일봉 차트는 막대 그래프로 표시되어 있는데, 이것은 위에서 설명한 바와 같이 일봉 차트에서는 음봉 및 양봉을 표현할 수가 없으므로 방문횟수를 단순한 막대 그래프로 표현하였기 때문이다. 일봉 차트를 보면 토요일 및 일요일의 방문횟수는 다른 요일들의 방문횟수에 비해서 현격히 낮다는 것을 알 수 있다.



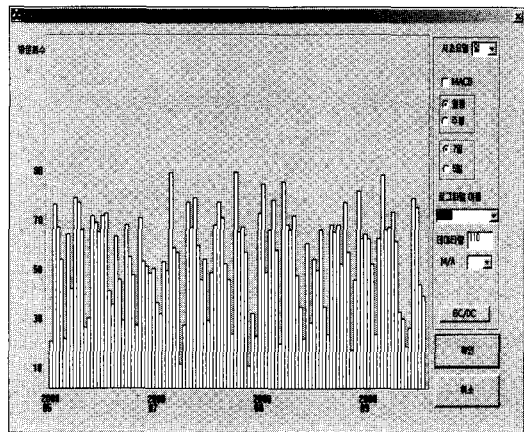
〈그림 7〉 회사 A의 일봉 차트



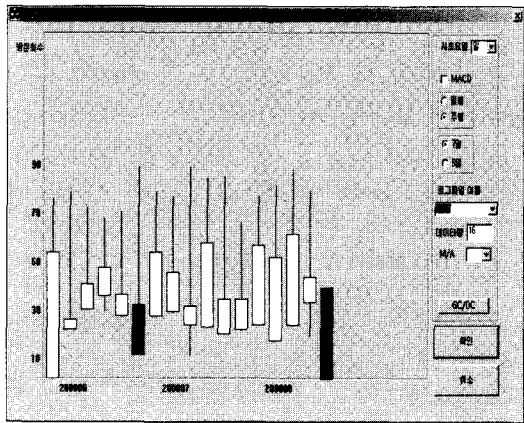
〈그림 8〉 회사 A의 주봉 차트

주봉은 목요일의 방문횟수가 시가에 해당되고 그 다음주 수요일의 방문횟수가 종가가 되게 표시하였다. 주봉 차트에서 흰색은 양봉을, 검은색은 음봉의 색깔을 각각 표시한다. 〈그림 9〉 및 〈그림 10〉은 회사 B에 대한 방문횟수의 일봉 차트 및 주

봉 차트를 각각 보여준다. 물론 다른 요일을 주봉의 시가로 선택할 수 있는 기능도 추가되었다. 〈그림 10〉에서는 수요일의 방문횟수가 주봉의 시가가 되고, 그 다음 주 화요일의 방문횟수가 주봉의 종가가 되도록 변경된 주봉 차트를 보여준다.



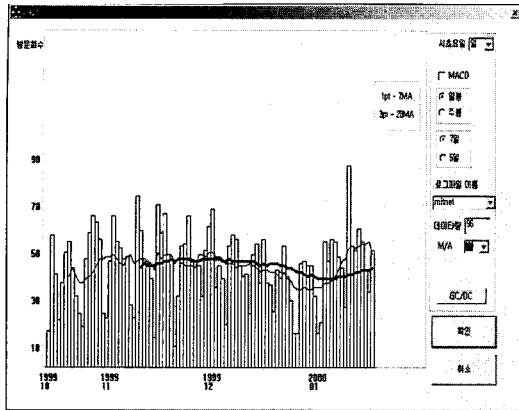
〈그림 9〉 회사 B의 일봉 차트



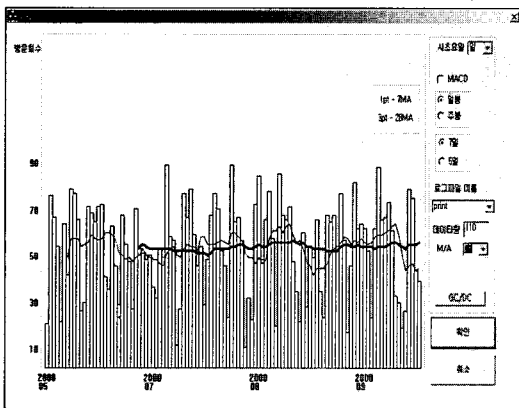
〈그림 10〉 회사 B의 주봉 차트

〈그림 11〉 및 〈그림 12〉에는 회사 A 및 회사 B의 일봉 차트들에 대한 이동평균선들이 표시되어져 있다. 이들은 7일 이동평균선(7MA)과 28일 이동평균선(28MA)들로서 각각 1주일간의 평균 방문횟수와 1개월간의 평균 방문횟수를 의미한다. 그림에서는 7MA는 가는 선으로, 28MA는 굵은 선으로 표시되어 있다. 물론 이동평균을 계산하는 기간을

임의로 정할 수 있는 기능도 추가하였다.



〈그림 11〉 회사 A의 일봉 차트에 대한 이동평균선



〈그림 12〉 회사 B의 일봉 차트에 대한 이동평균선

골든크로스(GC)와 데드크로스(DC)가 방문횟수

변화에 대한 예측에 사용될 수 있는지를 알아보기 위하여 <표 1>과 <표 2>를 작성하였다. <표 1> 및 <표 2>는 회사 A 및 B에 대한 GC 및 DC의 발생 시점과 일일 평균 방문횟수의 증감율을 각각 보여준다. 증감율은 직전 GC 시점 및 DC 시점부터 각각 다음 DC 시점 직전 및 GC 시점 직전까지의 기간 동안에 발생한 일일 방문횟수의 평균에 대한 증감율로 정의한다. GC 및 DC가 방문횟수 변화에 대한 예측에 사용될 수 있기 위해서는 직전 GC 시점부터 다음 DC 시점 직전까지의 기간(GC 기간) 동안에 발생한 평균 일일 방문횟수의 증감율은 증가하여야 되고, 직전 DC 시점부터 다음 GC 시점 직전까지의 기간(DC 기간) 동안에 발생한 평균 일일 방문횟수의 증감율은 감소하여야 한다.

<표 1>을 살펴보면, 회사 A의 경우에는, 조사 기간 동안에 각각 3회씩의 GC와 DC가 발생하였다. 1999년 11월 14일에 7일 이동평균선이 28일 이동평균선을 하향 돌파하는 DC가 처음 발생하였고, 1999년 11월 15일에 7일 이동평균선이 28일 이동평균선을 상향 돌파하는 GC가 발생하였다. 따라서 DC 기간은 11월 14일 하루밖에 없으며, 이 기간 동안의 평균 일일 방문횟수는 14회이다. GC 기간은 11월 15일부터 19일까지가 되며, 이 기간 동안의 평균 일일 방문횟수는 62회이다. 따라서, 11월 15일부터의 GC 기간 동안에는, 직전 DC 기간에 대비하여, 평균 342.9%의 일일 방문횟수 증가가 있음을 알 수 있다. 전체적으로, <표 1>에서는 모든

〈표 1〉 GC 및 DC를 이용한 방문횟수 예측 비교표(회사 A)

기 간	일일 평균 방문 횟수	상 태	평균 일일 방문 횟수 증감율(%)	비 고
1999/11/10~1999/11/13	50	시작~다음 DC 직전	-----	DC : 1999/11/14
1999/11/14~1999/11/14	14	DC~다음 GC 직전	- 72.0	GC : 1999/11/15
1999/11/15~1999/11/19	62	GC~다음 DC 직전	+ 342.9	DC : 1999/11/20
1999/11/20~1999/11/26	46	DC~다음 GC 직전	- 25.8	GC : 1999/11/27
1999/11/27~1999/12/05	49	GC~다음 DC 직전	+ 6.5	DC : 1999/12/06
1999/12/06~2000/01/05	41	DC~다음 GC 직전	- 16.3	GC : 2000/01/06
2000/01/06~2000/01/17	66	GC~마지막	+ 61.0	

<표 2> GC 및 DC를 이용한 방문횟수 예측 비교표(회사 B)

기 간	일일 평균 방문 횟수	상 태	평균 일일 방문 횟수 증감율(%)	비 고
2000/06/27~2000/07/06	55	시작~다음 GC 직전	-----	GC : 2000/07/07
2000/07/07~2000/07/07	60	GC~다음 DC 직전	+ 9.1	DC : 2000/07/08
2000/07/08~2000/07/09	19	DC~다음 GC 직전	- 68.3	GC : 2000/07/10
2000/07/10~2000/07/27	63	GC~다음 DC 직전	+ 231.6	DC : 2000/07/28
2000/07/28~2000/08/03	50	DC~다음 GC 직전	- 20.6	GC : 2000/08/04
2000/08/04~2000/08/13	58	GC~다음 DC 직전	+ 16.0	DC : 2000/08/14
2000/08/14~2000/08/22	52	DC~다음 GC 직전	- 10.3	GC : 2000/08/23
2000/08/23~2000/09/11	60	GC~다음 DC 직전	+ 15.4	DC : 2000/09/12
2000/09/12~2000/09/17	47	DC~마지막	- 21.7	

GC 기간이 직전 DC 기간보다 평균 일일 방문횟수가 증가함을 알 수 있다. 즉, 회사 A의 경우에는 GC가 발생하면 다음 DC가 발생할 때까지 기간 동안에 직전 기간에 비해서 일일 방문횟수가 평균적으로 증가함을 알 수 있다. 반대로, 모든 DC 기간이 직전 GC 기간보다 평균 일일 방문횟수가 감소함을 알 수 있다. 따라서, 향후 방문횟수 변화의 예측에 GC 및 DC가 의미있는 예측 지표 역할을 할 수 있음을 알 수 있다.

<표 2>에서는 회사 B에 대한 GC, DC 및 평균 일일 방문횟수 증감율을 보여준다. 조사 기간 동안에 각각 4회씩의 GC 및 DC가 발생하였는데, 모든 기간에 걸쳐서 GC 기간이 직전 DC 기간에 비하여 평균 일일 방문횟수가 증가함을 알 수 있으며, DC 기간이 직전 GC 기간에 대해서는 반대의 현상을 보여준다. 따라서 회사 B의 경우에서도, 회사 A의 경우처럼, 향후 방문횟수 변화의 예측에 GC 및 DC가 의미있는 예측 지표 역할을 할 수 있음을 알 수 있다.

완벽한 수학적 증명이나 모든 인터넷 웹 사이트들에 대한 실증을 통하여 예외 없는 결론을 내릴 수는 없어도, 위의 사례들을 통하여 다음과 같은 경험적인 방문횟수 예측 법칙을 도출할 수 있을 것이다. 즉, 인터넷 웹 사이트의 관리자는 방문횟수에 대한 GC 및 DC의 발생 여부를 주의 깊게 살펴

보되, 만약에 GC가 발생한다면 향후 방문횟수가 평균적으로 상승하고 방문자의 관심도가 높아질 것이므로, 현재의 웹 사이트 관리정책을 크게 수정할 필요가 없을 것이다. 만약 DC가 발생한다면 향후 방문횟수가 평균적으로 감소할 것이고 방문자의 관심도가 낮아질 것이므로, 현재의 웹 사이트 관리정책의 방향을 크게 수정하여야 할 것이다.

일일 방문횟수에 대한 이동평균선의 상승 전환과 하락 전환을 이용하여 방문횟수 변화를 예측하기 위해서는 이동평균선의 상승 기간과 하락 기간을 구분해야 한다. 일반적으로, 단기 이동평균선보다는 중기 혹은 장기 이동평균선이 일시적인 추세의 변화에 너무 예민하게 대응하지 않는 장점이 있으므로, 중기 혹은 장기 이동평균선을 선호한다. <표 3>은 일일 방문횟수에 대한 28일 이동평균선에 대한 상승 및 하락 정보를 보여주는데, 회사 A에 대해서는 13번의 상승 기간과 12번의 하락 기간이 존재함을 알 수 있다.

<표 3>에서는 기간별 증감율을 표시하지 않았으나 상승과 하락을 기준으로 하는 방법이기 때문에 상승(하락) 기간 동안에는 방문횟수가 당연히 상승(하락)함을 예상할 수 있다. 그러나, 이 방법을 사용하게 되면 너무나 많은 상승 및 하락 기간이 존재할 수 있으므로 웹 사이트 관리자로서 하여금 관리정책을 자주 변경하도록 강요할 가능성이 있다.

〈표 3〉 28일 이동평균선의 상승 전환과 하락 전환 비교표(회사 A)

기 간	비 고	기 간	비 고
1999/11/10~1999/11/11	상 승	1999/11/12~1999/11/14	하 락
1999/11/15~1999/11/19	상 승	1999/11/20~1999/11/20	하 락
1999/11/21~1999/11/22	상 승	1999/11/23~1999/11/26	하 락
1999/11/27~1999/11/29	상 승	1999/11/30~1999/11/30	하 락
1999/12/01~1999/12/01	상 승	1999/12/02~1999/12/04	하 락
1999/12/05~1999/12/06	상 승	1999/12/07~1999/12/10	하 락
1999/12/11~1999/12/12	상 승	1999/12/13~1999/12/15	하 락
1999/12/16~1999/12/16	상 승	1999/12/17~1999/12/17	하 락
1999/12/18~1999/12/18	상 승	1999/12/19~1999/12/29	하 락
1999/12/30~1999/12/30	상 승	1999/12/31~2000/01/02	하 락
2000/01/03~2000/01/03	상 승	2000/01/04~2000/01/05	하 락
2000/01/06~2000/01/10	상 승	2000/01/11~2000/01/11	하 락
2000/01/12~2000/01/17	상 승		하 락

따라서 주식 매매에서는 주가 변화 예측 지표로서 많이 쓰이는 방법들 중의 하나이지만 방문횟수 변화 예측에는 크게 실용적인 방법이 못된다고 볼 수 있다. 회사 B의 경우에도 비슷한 결론을 얻을 수 있으므로 여기에서는 구체적인 언급은 생략하기로 한다.

주식 매매에 있어서는 위에서 언급한 2가지 예측 방법들 모두 효율적이지만 방문횟수에 대한 예측에 있어서는 두 번째 방법이 그다지 실용적이지 못함을 알 수 있다. 이에 대한 이유는 다음과 같이 설명될 수 있다. 주식 매매에 있어서는 주가가 한 동안 하락한 후에 투자자들이 주가가 충분히 낮아 졌다고 인정되면 지속적으로 주식을 매매하게 된다. 이럴 경우에는 이동평균이 오랫동안 지속적으로 상승하게 되기 때문에 예측 지표로서 손색이 없다. 그러나, 방문횟수에 관해서는 이런 논리가 성립되지 않는다. 즉, 사용자는 단순히 특정 웹 사이트의 방문횟수가 적다고 오랫동안 지속적으로 방문하지 않는다. 웹 사이트의 내용에 관심이 있으면 방문횟수의 많고 적음에 관계없이 방문하게 되기 때문에 이동평균의 지속적인 상승이나 하락을 기대하기 어렵다.

지금까지 방문횟수의 일봉을 이용하여 미래의

방문자 방문횟수 변화를 예측하는 방법에 대하여 설명하고 실험 결과를 분석하였다. 마찬가지로, 방문횟수의 주봉도 예측에 사용할 수 있다. 이 경우에는 주봉의 증가를 이용하여 이동평균을 계산하면 일봉의 경우와 같은 환경이 되므로 여기에서는 자세한 설명을 생략하기로 한다.

5. 결 론

최근에 많은 경비와 노력을 기울인 웹 사이트의 효율성에 대하여 관심을 가지기 시작했으며, 웹사이트의 사용자 방문횟수를 지속적으로 점검하여 방문자들이 얼마나 관심을 가지고 있는가에 대한 정보를 원하는 기업들이 늘어나고 있다. 본 논문은 이러한 요구에 부응하기 위하여 웹 사이트에 대한 사용자들의 관심도 변화를 예측할 수 있는 방법을 제안하였다. 즉, 웹 사이트 홈페이지에 대한 사용자 방문횟수를 분석하여서 그 웹 사이트의 관심도가 향후 어떻게 변할 것인가에 대한 예측 방법을 제안하고 시스템을 개발하였다. 본 예측 시스템에서 사용한 예측 방법은 주식 매매에서 사용하는 주가 예측 기법들을 변형한 것이다. 본 시스템은 주식의 가격에 방문횟수를 대응하여 사용했으며, 방

문횟수에 대한 일봉/주봉 차트 및 이동평균등을 이용하여 해당 웹 사이트의 향후 방문횟수 변동을 예측하는데 필요한 정보를 제공한다. 만약에 예측 시스템이 방문횟수의 증가를 예측한다면 웹 사이트에 대한 사용자들의 관심도가 증가될 것이라고 예측할 수 있으므로 웹 사이트 운영자는 현재의 운영 정책을 강화하면 될 것이다. 한편, 예측 시스템이 방문횟수의 감소를 예측한다면 사용자들의 관심도가 감소될 것이라고 예측되므로 웹 사이트 운영정책이나 콘텐츠 혹은 웹 페이지의 화면 구성 등의 개편을 고려해야 할 것이다.

본 시스템은 두 가지 예측 방법들을 사용한다. 첫 번째 방법은 방문횟수에 대한 이동평균의 골든 크로스와 데드크로스 개념을 사용하고, 두 번째 방법은 이동평균의 상승 전환과 하락 전환의 개념을 이용한다. 2개의 웹 사이트들에 대한 실험 결과, 두 가지 방법들 모두 방문횟수 변화에 대한 예측 능력이 우수함을 알 수 있었다. 그러나, 두 번째 방법은 작은 변화에 대해서도 너무 민감하게 반응하기 때문에 첫 번째 방법보다 실용적이지 못하다고 믿어진다.

본 논문은 아직 미비한 점들이 많으므로 많은 사항들에 대해서 향후 보안 개선해야 한다고 생각된다. 우선, 본 논문에서 응용한 주가 예측 기법들 이외에도 많은 기법들이 있으므로 이들에 대해서도 예측시스템을 구현하고 실험하여 어느 기법이 방문횟수 예측에 보다 효율적인지를 밝히는 것이 필요하다고 생각한다. 마지막으로, 본 논문에서는 큰 대내외적인 변화가 없다고 가정하고 일상적인 환경에서의 사용자들에 대한 방문횟수를 기초로 하여 논리를 전개하였다. 따라서, 충격적인 환경 변화에 의한 방문횟수의 급격한 변화에 대해서는 본 논문의 논리가 적용되지 않을 수 있음을 밝혀두고자 한다.

참 고 문 헌

- [1] 김 산, 웹 사이트의 효율적 관리를 위한 로그 분석기에 관한 연구, 계명대학교 산업기술대학원 컴퓨터공학과 석사학위논문, 2001.
- [2] 변숙은, Log Data 분석을 통한 웹 사이트의 효율적 관리 방안-ABC Jeans사 사례분석을 중심으로, 한국과학기술원 석사학위논문, 2001.
- [3] 제일투자신탁증권, CJ시스템 트레이딩 매뉴얼, 제일투자신탁증권, 2000.
- [4] 최정일, 박선오, 임병준, 2000년대 주식·선물 시장 기술적 분석 -알기 쉬운 사례 분석 중심으로-, 도서출판 진리탐구, 1995.
- [5] Department of Information and Computer Science, University of California at Irvine, wwwstat, <http://www.ics.uci.edu/pub/websoft/wwwstat>.
- [6] Felton, M., "CGI : Internet Programming with C++ and C," Prentice Hall Pub. Co., 1997.
- [7] IETF Working Groups, Hypertext Transfer Protocol - HTTP/1.1, <http://www.ietf.org/rfc/rfc2068.txt?number=2068>, Jan. 1997.
- [8] Microsoft Co., Usage Analyst, <http://www.interse.com>.
- [9] Monticino, M., "Web analysis : stripping away the hype," IEEE Computer, Vol.31, No.12(Dec. 1998), pp.130-132.
- [10] Netrics.com, SurfReport, <http://software.bienlogic.com/SurfReport>.
- [11] Pirolli, P., R. Rao and J. Pitkow, "Silk from a Sow's ear : extracting usable structures from the Web," Proceedings of 1996 Conference on Human Factors in Computing Systems(1996) pp.118-125.
- [12] Pitkow, J., "In search of reliable usage data on the WWW," Proceedings of Sixth International World Wide Web Conference, 1997.
- [13] Sane Solutions, NetTracker, <http://www.sane.com/products/NetTracker>.
- [14] Statistical Laboratory, University of Cam-

[1] 김 산, 웹 사이트의 효율적 관리를 위한 로그

- bridge, Analog, <http://www.statslab.cam.ac.uk/~sret1/analog>.
- [15] WebTrends, WebTrends, <http://www.webtrends.com>.
- [16] Wu, K.L., P.S. Yu and A. Ballman, "Speed-Tracer : a Web usage mining and analysis tool," IBM Systems Journal, Vol.37, No.1 (1998), pp.89-105.