

# 소셜 미디어 정보를 이용한 실업률 예측<sup>†</sup>

## (Forecasting Unemployment Rate using Social Media Information)

나 종 화\*, 김 은 섭\*\*  
(Jonghwa Na and Eun-Sub Kim)

**요 약** 소셜 미디어 정보는 실시간으로 가장 최신의 정보의 획득을 가능하게 하며, 확산 속도가 빠르고 형태의 제약 없이 유연하게 생산 및 유통할 수 있다는 강점이 있으며, 최근 경제변수들의 예측에 있어서 소셜 미디어 정보를 이용한 예측의 활용 방안에 대한 논의가 활발히 이루어지고 있다. 본 논문에서는 실업률을 예측함에 있어 소셜 미디어 정보 유형 중 하나인 Google Index 정보를 이용하여 시계열 모형 중 하나인 ARIMAX 모형과 ECM 모형을 적합하였고, 모형의 예측력을 비교하기 위해 기존의 ARIMA 모형과의 비교를 수행하였다. 또한, 소셜 미디어 정보 이용에 있어 Google Index 뿐만 아니라 국내 소셜 미디어 정보인 News Index와의 결합을 통해 예측력의 제고 가능성을 확인하였다. 본 연구에서 다루어진 분석 절차와 통계적 모형의 적합 과정은 실업률 외의 주요 사회, 경제지표에도 적용될 수 있다.

**핵심주제어** : 소셜 미디어 정보, 실업률 예측, Google Index, ARIMAX, ECM

**Abstract** Social media has many advantages. It can gain latest information with real time, be spread rapidly, easily be reproduced and distributed regardless of its form. These advantages can result in real time predictions using the latest information, which is possible due to the increase in social demand for more quick and accurate economic variable predictions. In this paper we adopted ARIMAX and ECM model to predict the unemployment rate and as a social information we used the Google Index provided by Google Trend. Also we used News Index as a domestic social information. The process of fitting statistical model considered in this paper can be adopted to predict various socio/economic indices as well as unemployment rate.

**Key Words** : Social media information, Unemployment rate prediction, Google Index, ARIMAX, ECM

### 1. 서 론

최근 경제변수들의 예측에 있어서 소셜 미디어 정보를 이용한 예측의 활용 방안에 대한 논의가 활발히 이루어지고 있다. 소셜 미디어 정보는 실시간으로 가장 최신의 정보의 획득을 가능하게 하며, 확산 속도가 빠르고 형태의 제약 없이 유연하게 생산 및 유통할 수 있다는 강점이 있다. 이러한 소셜 미디어 정보의 강점은 경제변수들의 예측에 있어 좀 더 빠르고 정확한 예측에 대한 사회적 수요가 급증함에 따라 가장

<sup>†</sup> 본 논문은 미래창조과학부 산업융합원천기술개발사업의 “웹 인텔리전스를 위한 웹 폭증 데이터 분석형 리스닝 플랫폼용 소셜웹 이슈 탐지-모니터링 및 예측 원천 기술 개발”의 지원으로 개발된 것임 (과제번호 10039158)

\* 충북대학교 정보통계학과/ 비즈니스데이터융합학과 (cherin@cbnu.ac.kr)

\*\* 충북대학교 정보통계학과

최신의 정보를 이용한 실시간적인 예측을 가져다 줄 것이다.

실업률의 올바른 예측은 실업률을 기반으로 하여 계산되는 국가 경제지표 및 실업률 기반으로 확립되는 국가 정책수립에서 생기는 오차를 줄일 수 있을 뿐 아니라, 노동시장 및 국가 경제 변화에 국가 정책이 탄력적으로 반응할 수 있도록 도와주기 때문에 올바른 실업률 예측은 중요하다. 하지만 현재의 국내 실업률 예측은 현재 시점을 예측함에 있어 이전 시점의 실업률의 영향만을 고려하는 일변량 시계열 분석 모형을 주로 사용하며 이러한 모형은 시시각각 변하는 경제상황을 즉각적으로 반영하지 못한다는 한계점을 내포하고 있다. 실제로 국외의 많은 연구에서 실업률을 예측함에 있어 실업률을 예측하는 지표로써 소셜 미디어 정보를 활용한 사례가 늘고 있으며, 외국의 경우는 소셜 미디어 정보 유형 중 하나인 구글 검색 통계 사이트인 Google Trend([http:// www.google.com/trends/](http://www.google.com/trends/))에서 제공되는 Google Index(GI) 정보를 이용한 모형 연구가 점차 증가하고 있는 추세이다.

국외에서 소셜 미디어 정보를 이용한 인플루엔자의 전염성 및 확산의 패턴을 연구하는 논문으로 Dukić 등[9], Ginsberg 등[11], Kelly와 Grant [12], Valdivia[13]가 있으며 이들은 각 미국과 뉴질랜드, 호주, 스페인의 인플루엔자 패턴을 분석한 결과 GI를 이용한 분석이 인플루엔자 전염성 및 확산의 패턴 연구에 상당한 도움이 되는 것을 확인한 바 있다. 또한, Askitas와 Zimmermann [5]은 독일의 실업률 자료를 분석함에 있어 GI 중 실업 및 경제 검색 데이터를 오차수정모형(ECM) 적합에 사용하였으며 그 결과 GI 변수를 사용하기 전보다 사용한 후에 모형이 향상됨을 확인하였다. D'Amuri와 Marcucci[8]는 미국 실업률의 예측함에 있어 AR, ARMA 모형과 GI를 외생변수로 추가한 ARX, ARMAX 모형 등을 총 51개 주에 대한 집단별 분석한 결과 총 500여개의 모델 중 가장 작은 MSE를 보이는 대부분의 모델은 GI변수가 외생변수로 들어간 ARX, ARMAX 인 것을 알 수 있었다.

또한, D'Amuri[7]는 이탈리아 실업률 자료를 분석함에 있어 GI를 이용하여 단기 실업률 예측 ARMAX 모형에 적합하였다. 그 결과 실업률 예측 모형이 개선된 것을 확인하였으며 더 나아가 이탈리아의 짧은 기간 경제변동을 반영시키기에는 GI변수가 적합하다는 점을 확인한 바 있으며, Choi와 Varian[6]은 초기 실

업수당 신청자를 예측함에 있어 GI를 이용하여 초기 실업수당 신청자를 예측함에 있어 AR모형으로 예측한 결과보다 GI를 사용한 ARIMAX모형이 더 좋은 예측력을 나타내고 있음을 확인하였다.

김희재[1]는 국내 실업률 예측에 있어 Box -Jenkins의 계절형 ARIMA, 분해법, 자기회귀오차모형, Winters의 계절지수 평활법을 사용하여 실업률 예측력이 가장 우수한 모형을 찾는 연구를 진행하였다. 이 연구에 의하면 위에서 언급한 4가지 모형 중 국내 실업률을 예측하는 방법으로는 계절형 ARIMA 모형이 가장 바람직하다는 결론을 얻을 수 있었다.

이와 같이 과거에는 쓸모없는 정보로 취급되거나 활용방안을 찾지 못해 흘려보냈던 소셜 미디어 정보들은 사회적 트렌드를 반영시킬 새로운 소셜 지표로의 발전가능성을 인정받아 활발히 연구중이다. 또한 소셜 미디어 분석은 단순히 독감이나 질병의 확산 및 패턴 등을 분석하던 초창기와 달리 경제 및 사회 지표 분석, 마케팅, 소비자의 심리 분석을 하는 데에 활용되는 등 그 활용폭이 점차 확대되고 있으며, 이에 따라 본 논문에서도 이미 국외에서 선행되었던 소셜 미디어 정보를 이용한 실업률 예측을 국내 실업률 예측에 적용시켜 보았다.

본 논문에서는 실업률을 예측함에 있어 소셜 미디어 정보 유형 중 하나인 GI 정보를 이용하여 시계열 모형 중 하나인 ARIMAX 모형과 ECM 모형을 적합하였고, 모형의 예측력을 비교하기 위해 기존의 실업률 예측 모형 중 우수성이 확인된 계절형 ARIMA 모형과의 비교를 통해 본 연구의 활용 가능성 및 개선 여지에 대해 살펴보았다. 또한, 소셜 미디어 정보 이용에 있어 GI뿐만 아니라 한국전자통신연구원(ETRI)에서 제공한 국내 소셜 미디어 정보인 News Index(NI)와의 결합을 통한 실업률 예측이 예측력에 있어 개선의 여지가 있는지 살펴보았다.

## 2. 소셜 미디어 정보를 이용한 실업률 예측 모형

### 2.1 분석 자료

본 연구에서 실업률 예측에 사용된 자료는 통계청에서 매월 공표하는 국내 실업률 자료와 구글 검색

통계 사이트에서 제공되는 월별 검색 키워드별 지수(GI)이며, 모형 적합 및 예측에 사용된 자료의 기간은 2004년 1월부터 2011년 10월까지의 94개의 월별 자료를 이용하였다. 실업률과 연관성이 있는 많은 키워드들 중에서 본 연구에서는 취업정보 획득과 관련 있는 키워드인 '채용'에 대한 GI를 실업률 예측에 사용하였다. 본 연구에서 키워드 "채용"이 외생변수로 채택된 이유는 40여개의 실업률 연관 키워드의 분석 결과 가장 연관성이 높은 것으로 나타났기 때문이다. 다양한 통계적 예측과 관련된 연구로는 이현규와 신호호[4], 나중화 등[2], 유혜경 등[3]이 있다.

## 2.2 ARIMAX 모형

조사계열인  $Y_t$ 와 연관이 있는 외생계열  $X_t$ 가 존재할 때,  $X_t$ 를  $Y_t$ 를 설명할 때 설명변수로 사용할 수 있다.  $Y_t$ 가 ARIMA 모형을 따르고,  $X_t$ 와 연관이 있다고 할 때,  $Y_t$ 는 ARIMA 모형에 외생계열이 결합된 형태인 ARIMAX 모형으로 표현할 수 있으며 식 (2.1)과 같이 정의된다.

$$\phi_p(1-B)^d Y_t = \theta_q(B)a_t + \omega x_t' \quad (2.1)$$

일반적으로 ARIMAX의 가정으로는 입력시계열(외생계열)이 출력시계열(반응계열)에 영향을 미치나 반대로 출력시계열은 입력시계열에 영향을 미치지 않는 것을 가정으로 한다. 만약 분석하고자 하는 두 시계열 사이에 서로 영향을 주는 관계가 성립되는 경우 ARIMAX 모형 대신 다변량 시계열 모형인 벡터 ARIMA 모형을 이용한 분석이 더 바람직하다고 할 수 있다.

## 2.3 오차수정모형(ECM)

변수들간의 공적분 관계가 밝혀진 상황에서 차분된 자료를 사용하여 AR모형을 추정하는 것은 유용한 정보를 버리는 오류를 범하게 된다. 오차수정모형의 경우 AR 모형에 공적분 제약이 추가된 모형으로 공적분항에서는 변수들간의 장기적 인과관계를, 차분항에서는 단기적 인과관계의 효과를 살펴볼 수 있다.

“특정 경제변수 사이에 공적분이 존재한다면 이를

대변하는 오차수정모형이 존재한다”는 대표이론(Representation Theorem)이 Engle과 Granger[10]에 의해 개발되었는데, 오차수정모형의 일반모형식은 다음과 같다.

$$\Delta Y_t = \alpha + \delta e_{t-1} + \sum_{i=1}^n \beta_i \Delta Y_{t-i} + \sum_{k=1}^m \gamma_k \Delta X_{t-k} + \epsilon_t \quad (2.2)$$

여기서  $e_{t-1} = Y_{t-1} - \lambda X_{t-1}$ 이며  $e_{t-1}$ 은 공적분이 존재할 때  $X_{t-1}$ 과  $Y_{t-1}$ 사이의 불균형 오차를 반영한 오차수정항(Error Correction Term)이다. 따라서  $e_{t-1} \sim I(0)$ 이며,  $Y_t$ 의 변화가  $X_t$ 의 변화뿐만 아니라 두 변수간의 불균형의 정도에 의해서도 영향을 받는 것으로 표현하고 있다. 이러한 ECM은 장기균형의 특징을 파악함과 동시에 단기적 조정과정을 동태적인 측면에서 파악할 수 있게 해준다.

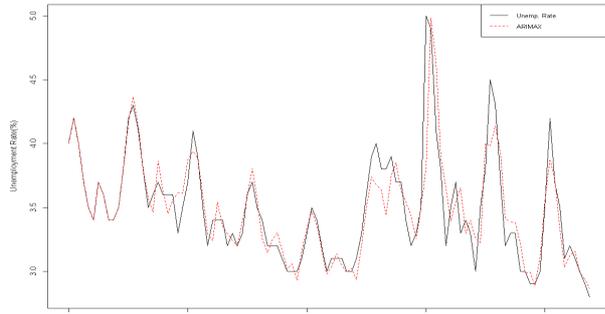
## 3. 모형 적합 및 평가

### 3.1 적합모형

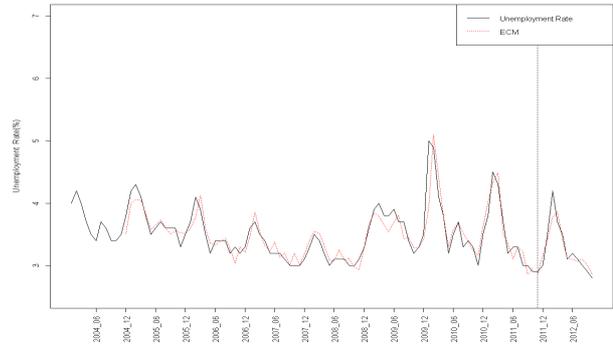
소셜 미디어 정보를 이용한 실업률 예측의 적용 가능성을 확인하기 위해서 실업률과 채용 GI자료에 대한 실업률 예측 모형으로 ARIMAX 모형과 오차수정모형을 적합하였다.

$$(1 - 0.5148B)(1 - B)(1 - B^{12})Y_t = (1 + 0.7605B)(1 + 0.7637B^{12})a_t + 0.0002x_t' \quad (3.1)$$

먼저, ARIMAX 모형의 적합 결과를 식으로 표현하면 위 식 (3.1)과 같이 표현할 수 있으며, 적합 결과와 표본기간 이후의 12달 기간의 실업률을 예측한 결과를 그래프로 나타내면 <그림 3.1>과 같이 나타낼 수 있다.



<그림 3.1> 실업률 예측 적합모형 - ARIMAX 모형



<그림 3.2> 실업률 예측 적합모형 - ECM

다음의 <표 3.1>은 실업률과 채용 GI에 대한 오차수정모형의 적합 결과이며, 적합 결과와 표본기간 이후의 12달 기간의 실업률을 예측한 결과를 그래프로 나타내면 <그림 3.2>과 같이 나타낼 수 있다.

<표 3.1> 오차수정모형의 적합결과

독립변수	추정계수	표준편차	t-stat
$\alpha$	0.2627	0.1318	1.9930*
$e_{t-1}$	-0.1190	0.0585	-2.0350**
$\Delta X_{t-1}$	0.0297	0.0446	0.6660
$\Delta X_{t-2}$	-0.0338	0.0442	-0.7640
$\Delta X_{t-3}$	0.0609	0.0468	1.3010
$\Delta X_{t-4}$	0.0860	0.0451	1.9050*
$\Delta X_{t-5}$	0.1341	0.0470	2.8560***
$\Delta X_{t-6}$	0.0448	0.0497	0.9010
$\Delta X_{t-7}$	0.0854	0.0468	1.8260*
$\Delta X_{t-8}$	-0.0084	0.0480	-0.1740
$\Delta X_{t-9}$	0.0929	0.0420	2.2120**
$\Delta X_{t-10}$	-0.0353	0.0436	-0.8110
$\Delta Y_{t-1}$	-0.0893	0.1271	-0.7030
$\Delta Y_{t-2}$	-0.2901	0.1167	-2.4860**
$\Delta Y_{t-3}$	-0.3321	0.1304	-2.5460**
$\Delta Y_{t-4}$	-0.5795	0.1295	-4.4750***
$\Delta Y_{t-5}$	-0.0636	0.1450	-0.4390
$\Delta Y_{t-6}$	-0.2639	0.1293	-2.0420**
$\Delta Y_{t-7}$	-0.3761	0.1257	-2.9920***
$\Delta Y_{t-8}$	-0.1633	0.1333	-1.2250
$\Delta Y_{t-9}$	-0.3059	0.1409	-2.1720**
$\Delta Y_{t-10}$	-0.2815	0.1490	-1.8890*

1) \*, \*\*, \*\*\* 는 10%, 5%, 1%에서 통계적 유의성이 있음을 나타냄.

### 3.2 모형평가

기존의 실업률 예측 방법과 본 논문의 실업률 예측 방법과의 비교를 통해 소셜 미디어 정보를 이용한 실업률 예측에 있어 예측력의 개선 여지를 알아보자. 기존의 실업률 예측 방법으로는 김희재[9]의 연구결과 실업률 예측에 있어 AIC(Akaike Information Criterion), SSE(Error Sum of Square)를 기준으로 가장 바람직하다고 판단되었던 계절형 ARIMA 모형을 이용한 예측방법을 사용하였다. 모형의 추정 결과에 기초한 해당 모형의 설명력 또는 예측력은 추정치와 관측치의 차이, 즉 잔차에 대한 여러 가지 통계량의 산출을 통해 모형의 추정 및 예측 과정에서 발생하는 오차의 크기와 방향을 살펴볼 수 있다.

다음의 <표 3.2>는 실업률 예측 모형들의 예측력을 다양한 평가지표를 사용하여 비교한 결과를 제시하였다. 그 결과 기존의 실업률 예측 방법으로 사용되어져 왔던 계절형 ARIMA 모형의 결과보다 채용 GI 정보를 이용한 ARIMAX 모형 및 ECM의 결과가 모든 평가지표에서 실업률 예측력에 개선이 있음을 확인할 수 있다. 모형 비교에 사용된 평가지표로는 MSE(Mean Square Error), RMSE(Root Mean Square Error), MAE(Mean Absolute Error), MPE(Mean Percentage Error), MAPE(Mean Absolute Percentage Error)가 사용되었다.

<표 3.2> 실업률 예측 모형의 예측력 비교

모형	MSE	RMSE	MAE	MPE	MAPE
ARIMAX	0.0124	0.1114	0.0899	-1.0589	2.6665
ECM	0.0239	0.1547	0.1148	0.0971	3.3374
계절형 ARIMA	0.0289	0.1701	0.1299	3.0547	3.7701

- 1) ARIMAX의 표본기간은 2004/1~2011/10.
- 1) ECM의 표본기간은 2004/1~2011/10.
- 2) 계절형 ARIMA 모형의 표본기간은 1996/6~2011/10.

#### 4. 국내 소셜 미디어 정보를 이용한 실업률 예측모형

3절에서는 기존의 실업률 예측 방법과 본 논문의 실업률 예측 방법과의 비교를 통해 소셜 미디어 정보를 이용한 실업률 예측에 있어 예측력의 개선이 있음을 확인하였다. 4절에서는 3절에서 실업률 예측에 이용한 GI 중 2011. 11. ~ 2012. 10. 기간의 GI를 국내 소셜 미디어 정보인 NI 정보와 결합을 통한 GI 정보의 개선이 실업률 예측력에 있어 개선의 여지가 있는지 알아보자.

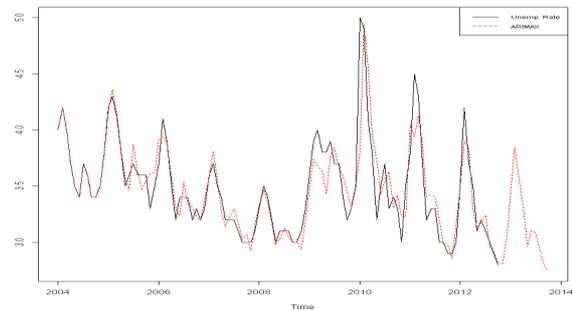
##### 4.1 예측모형 - ARIMAX 모형

우선, GI와 국내 소셜 미디어 정보의 결합을 통한 GI 정보의 개선이 실업률 예측력에 개선을 가져다주는지 알아보자. 2004. 1. ~ 2012. 10. 기간의 실업률 자료와 동 기간의 GI 중 2011. 11. ~ 2012. 10. 기간이 GI와 NI의 다양한 혼합 비율의 가중합에 의해 개선된 GI 자료를 이용하여 ARIMAX 모형을 적합하였으며, <표 4.1>은 각 혼합 비율에 따른 적합모형의 예측력 비교를 통해 모형의 성능 개선 여부를 확인하였다. 그 결과 국내 소셜 미디어 정보인 NI 정보의 적절한 활용이 예측력 제고에 기여함을 알 수 있다.

<표 4.1> ARIMAX 모형 예측력 비교

GI(%):NI(%)	100:0	75:25	50:50	25:75	0:100
MSE	0.0395	0.0394	0.0392	0.0390	0.0390
RMSE	0.1988	0.1984	0.1979	0.1976	0.1974
MAE	0.1195	0.1187	0.1178	0.1176	0.1176
AIC	-514.8	-515.2	-515.6	-515.6	-515.5
BIC	-504.7	-505.1	-505.4	-505.5	-505.4

2004. 1. ~ 2012. 10. 기간의 실업률 자료와 GI 자료를 이용하여 ARIMAX 모형을 적합하였다. 최근 1년(2011. 11. ~ 2012. 10.) 기간의 GI 정보는 동 기간의 NI 정보와 50:50의 가중 지표값을 사용하였다. 아래의 <그림 4.1>은 모형 적합 결과와 함께 향후 1년간의 예측값을 나타낸 것이다. ARIMAX 모형의 예측값을 구하는 과정에서 GI 또는 GI\*NI의 가중지표값을 ANN (Artificial Neural Network) 방법으로 예측한 값을 예측 기간(2012. 11. ~ 2013. 10)의 외생변수로 사용하였다.



<그림 4.1> ARIMAX 예측 모형

##### 4.2 예측모형 - ECM

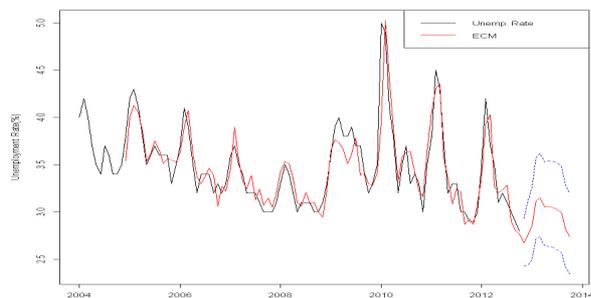
4.1절과 마찬가지로 GI와 국내 소셜 미디어 정보의 결합을 통한 GI 정보의 개선이 ECM 적합모형을 통한 실업률 예측력에 개선을 가져다주는지 알아보자. 2004. 1. ~ 2012. 10. 기간의 실업률 자료와 동 기간의 GI 중 2011. 11. ~ 2012. 10. 기간이 GI와 NI의 다양한 혼합 비율의 가중합에 의해 개선된 GI 자료를 이용하여 ECM을 적합하였으며, <표 4.2>은 각 혼합 비율에 따른 적합모형의 예측력 비교를 통해 모형의 성능 개선 여부를 확인하였다. 그 결과 MSE, RMSE, MAE 등의 평가 지표에서 GI:NI 가 100:0인 경우, 다시 말

해 GI만 사용한 경우가 가장 작은 값을 보이나, AIC 등의 기준에서 NI 정보의 적절한 개입이 모형의 성능 제고에 기여함을 알 수 있다.

<표 4.2> ECM 예측력 비교

GI(%):NI(%)	100:0	75:25	50:50	25:75	0:100
MSE	0.0330	0.0331	0.0333	0.0340	0.0344
RMSE	0.1816	0.1819	0.1825	0.1843	0.1854
MAE	0.1302	0.1314	0.1324	0.1333	0.1339
AIC	-257.0	-267.6	-265.2	-243.0	-208.5
BIC	-182.3	-192.9	-190.5	-168.3	-133.9

2004. 1. ~ 2012. 10. 기간의 실업률 자료와 GI 자료를 이용하여 ECM을 적합하였다. 최근 1년(2011. 11. ~ 2012. 10.) 기간의 GI 정보는 동 기간의 NI 정보와 50:50의 가중 지표값을 사용하였다. 아래의 <그림 4.2>는 모형 적합 결과와 함께 향후 1년 간의 예측값을 나타낸 것이다.



<그림 4.2> ECM 예측 모형

## 5. 결론 및 향후 연구방향

본 논문에서는 실업률 예측 방법에 있어 소셜 미디어 정보를 이용한 예측 방법의 적용 가능성에 대한 시사점을 제시하였다. 실업률 예측을 위해 사용된 주된 소셜 미디어 정보로는 GI를 이용하였고, 국내 정보인 NI를 적절히 활용하여 GI의 예측력 제고에 기여하였다. 고려된 예측 모형으로는 ARIMAX, ECM을 주된 모형으로 하여 분석 하였다.

본 논문에서는 실업률 자료에 대한 연구를 수행하였으나, 본 연구에서 다루어진 분석 절차와 통계적 모

형의 적합 과정은 실업률 외의 주요 사회, 경제지표에도 적용될 수 있다. 또한, 실업률 자료의 경우 월 단위로 제공되는 반면, 주 단위 또는 일 단위로 제공되는 다양한 사회, 경제 지표에도 적용될 수 있다. 2004년부터 제공되는 GI에 비해 국내 소셜 미디어 정보는 최근의 정보만을 DB로 구축하고 있어 이용에 제한이 따르는 것이 사실이다. 그러나 보다 정확하고 신속하게 정보를 제공하는 장점이 있어, 본 연구에서와 같이 이들 정보를 결합하여 사용하는 것도 좋은 시도라 생각된다. 또한 국내 소셜 미디어 자료의 경우 키워드 검색빈도가 일정 수준 이상으로 확보하는데 용이하며, 감성분석의 결과를 활용하는 데도 장점을 가진다. 다만 감성분석의 경우 분석 결과에 대한 보다 세밀한 검증 과정이 요구된다.

## 참 고 문 헌

- [1] 김희재, “계절형 시계열의 예측방법 비교”, 경성대학교 기초과학연구 논문집, 제21집, pp. 115-124, 2010.
- [2] 나중화, 유혜경, 남은미, 조완섭, “질적변수에 대한 계량화를 통한 사면방귀 예측모형”, 한국산업정보학회지, 제14권 제5호, 281-288, 2009.
- [3] 유혜경, 이진영, 나중화, “매장문화재 예측을 위한 통계적 분류 분석”, 한국산업정보학회지, 제14권 제3호, 106-113, 2009.
- [4] 이현규, 신용호, “센서스 정보 및 전력 부하를 활용한 전력 수요 예측”, 한국산업정보학회논문지, 제18권 제3호, 35-46, 2013.
- [5] Askatas, N. and Zimmermann, K. F., “Google Econometrics and Unemployment Forecasting”, IZA Discussion Paper, 4201, 2009.
- [6] Choi, H. and Varian, H., “Predicting Initial Claims for Unemployment Benefits”, Google Technical Report, 2009.
- [7] D’Amuri, F., “Predicting Unemployment in Short Samples with Internet Job Search Query Data”, MPRA paper 18403, 2009.
- [8] D’Amuri, F. and Marcucci, J., “Google it! Forecasting the US Unemployment Rate with a Google Job Search Index”, Bank of Italy, 2009.

- [9] Dukić, V., Lopes, H. F. and Polson, N. G., "Tracking Epidemics with State-Space SEIR and Google Flu Trends", Journal of the American Statistical Association, Vol. 107, 2012.
- [10] Engle, R. F. and Granger, C. W. J., "Co-Integration and Error Correction : Representation, Estimation, and Testing", Econometrica, Vol. 55, pp. 251-276, 1987.
- [11] Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S. and Brilliant, L., "Detecting Influenza Epidemics using Search Engine Query Data", Nature, Vol. 457, pp. 1012-1014, 2009.
- [12] Kelly, H. and Grant, K., "Interim Analysis of Pandemic Influenza(H1N1) 2009 in Australia: Surveillance Trends, Age of Infection and Effectiveness of Seasonal Vaccination", EURO SURVEILLANCE, Issue 31, pp.1-5, 2009.
- [13] Valdivia, A., "Diseases Tracked by Using Google Trends, Spain", Emerging Infectious Diseases, Vol. 16, No. 1, 2010.



**나 중 화 (Jonghwa Na)**

- 중신회원
- 서울대학교 계산통계학과 이학석사
- 서울대학교 계산통계학과 이학박사
- 충북대학교 자연과학대학 정보통계학과 교수
- 관심분야 : 빅데이터, 데이터마이닝, 통계 모델링/예측, 소셜 데이터 분석



**김 은 섭 (Eun-Sub Kim)**

- 충북대학교 정보통계학과 학사
- 충북대학교 통계학과 석사
- 관심분야 : 빅데이터, 소셜 미디어 분석, 데이터마이닝, 시계열 분석

논문 접수 일 : 2013년 08월 26일  
 1차수정완료일 : 2013년 10월 05일  
 게재 확정 일 : 2013년 10월 08일