

실업률 예측을 위한 인터넷 검색 정보의 활용

권치명 · 황성원 · 정재운*

Application of Web Query Information for Forecasting Korean Unemployment Rate

Chi-Myung Kwon · Sung-Won Hwang · Jae-Un Jung*

ABSTRACT

Unemployment is related to social issues as well as personal economics activity so various policies have been made to reduce the unemployment rate in many countries. Because of delay inherent in the survey mechanism to collect unemployment data, it takes lots of time to acquire survey unemployment data. To develop proper policies for reducing unemployment rate at the right time, it is quite critical to obtain faster and more accurate information concerning about unemployment level. To remedy this problem, recently an advanced analytics utilizing internet queries is suggested. To examine the potential of Web query information, this research investigates the usefulness of internet activity data to predict Korean unemployment rate. One of selected web-query data(unemployment claim) has a quite strong correlation with unemployment rate. This research employes a time series approach of the ARIMA model that utilizes the information of keyword queries provided by the Naver(Korean representative portal site) trend together with unemployment rate data provisioned from Statistics Korea. With respect to model selection guidelines of mean squared error and prediction error, the model with utilizing the web query information shows better results than the model without such information. This suggests that there is a strong potential for the used method, which needs to be further explored.

Key words : Unemployment Rate, Web Query, Time Series Forecast, ARIMA Model

요약

실업은 개인의 경제적 활동뿐 아니라 사회적 문제와 관련되어 있기 때문에 많은 국가들은 실업률을 낮추기 위해 다양한 정책을 추진하고 있다. 기존의 실업 실태 조사 방식에서는 조사시간 지연으로 인해 실업률 자료 확보에 많은 시간이 소요된다. 시의 적절한 실업 정책을 개발하기 위해서는 신속하고 정확한 실업 예측 관련 자료를 확보하는 것이 중요한 문제이다. 이러한 문제를 개선하기 위해 최근에 인터넷 검색 정보를 활용한 분석 기법이 제안되고 있다. 본 연구는 우리나라의 실업률을 예측하는데 인터넷 검색 정보가 어떤 영향을 미치는가를 조사하였다. 선택한 검색어 중에서 '실업급여' 검색어의 트렌드는 실업률과 상당히 높은 상관관계를 보여 주었다. 본 연구는 네이버 트렌드에서 제공하는 인터넷 검색어 정보를 시계열 자료의 분석에 널리 사용되는 ARIMA 모형에 추가하여 검색 정보의 활용이 실업률 예측력에 미치는 영향을 분석하였다. 예측모형의 선택 기준으로 제시되는 예측치의 평균 제공 오차와 예측 오차 측면에서 실업 관련 인터넷 검색어를 활용한 모형이 그렇지 않은 모형보다 우수한 것으로 나타났다. 이러한 결과는 실업률 예측에 있어서 검색 정보의 활용 가능성을 제시하고 있으며 향후 더 많은 연구가 필요할 것으로 판단된다.

주요어 : 실업률, 인터넷 검색어, 시계열 예측, ARIMA 모형

*이 논문은 동아대학교 교내연구비 지원에 의하여 연구되었음.

Received: 30 April 2015, **Revised:** 23 June 2015,
Accepted: 24 June 2015

*Corresponding Author: Jae-Un Jung
E-mail: imhere@dau.ac.kr
BK21Plus Groups, Dong-A University

1. 서론

ILO(International Labour Organization)의 보고에 따르면 전 세계 실업자는 2014년 기준으로 2억 1백만 명이

상이며, 이는 2008년 세계 경제위기 때보다 3천 1백만 명 이상 증가한 수치이다. 또한 2015년부터 향후 5년 동안 실업자 수가 약 1,100만 명이 더 늘어날 것으로 전망하였다. 세계적인 실업문제 해결을 위해서는 2019년까지 일자리 감소를 고려하여 총 2억 8천만 개의 일자리 창출이 필요하다고 밝혔다⁵⁾. 한국은 2014년 기준으로 3.5%의 낮은 실업률을 보이고 있지만, 취업을 원하는 주부와 아르바이트 학생을 고려할 경우 체감실업률은 전체 경제활동인구의 11.2%(320만 명)에 이른다¹⁰⁾. 개인의 실업문제로 인해 경제, 세금, 범죄 등 국가와 사회적 관점에서 다양한 이슈가 제기되고 있어 실업해소를 위한 정책 마련에 안간힘을 쏟고 있다^{7), 11)}. 실업 관련 정책 수립에는 많은 시간이 소요되기 때문에 조기에 정확한 예측이 필요하다.

Choi and Varian(2009)은 미국 노동부에서 제공하는 실업급여 신청자 수(initial claims)와 구글 트렌드(Google Trends)의 실업 관련 인터넷 검색어 자료를 활용하여 실업급여 신청자 수를 조기에 예측하는 연구를 진행하였다. 이들은 복지와 실업(welfare & unemployment) 범주의 인터넷 검색어를 활용하는 예측모형을 제안하였는데 실업급여 신청자 수를 좀 더 정확하게 예측할 수 있었으며 특히 2008년부터 2009년까지의 미국 경기불황 구간(short term)에 대해서 높은 예측력을 보였다⁴⁾. Anvik and Gjelstad (2010)의 경우에는 노르웨이 정부의 실업률 자료와 구글 트렌드의 실업 관련 인터넷 검색어 자료를 활용하여 실업률을 조기에 예측하는 모형을 연구하여, 장기보다 단기적 관점에서 예측력을 높이는 연구 결과를 보였다¹¹⁾. 이와 같이 실업률을 조기에 예측할 수 있는 모델이 개발되면 정부의 실업해소를 위한 정책 지원을 줄일 수 있을 뿐만 아니라 정책의 기대효과도 향상시킬 수 있을 것으로 사료된다. 실업문제는 국가나 사회적 특성에 따라 문제의 양상이 달라지기 때문에 한국 역시 지역 특성에 맞는 실업 예측 모형의 연구가 필요하다.

앞서 살펴본 선행연구 모두 시계열 예측모형 개발에 ARIMA 모형을 사용하였으며, 연구주제와 관련된 인터넷 검색어를 활용하여 모형의 예측력을 개선하는 결과를 보여주었다. 이런 관점에서 본 연구는 한국 통계청에서 제공하는 실업률 시계열 자료를 활용하여 실업률 예측을 위한 ARIMA 기초모형을 개발하고, 한국 인터넷 검색어의 활용이 해당 모형의 예측력을 향상시키는 데 의미가 있는지 검증하고자 한다. 이를 위해 기초모형에 네이버 트렌드(Naver Trend)에서 제공하는 실업 관련 검색어 자료를 변수로 추가하여 예측력 향상 수준을 분석한다. 본 연구는 해외의 기존 연구방법을 활용하여 한국의 실정에

맞는 실업률 변화를 조기에 예측하기 위한 새로운 시도로서 관련 연구의 활성화에 기여할 것으로 사료된다.

2. 실업률 통계자료

우리나라 통계청에 따르면, 실업률 통계자료는 경제활동인구(economically active population) 조사체계에 따라 수집, 관리되고 있다. 경제활동인구조사는 국민의 경제활동(취업, 실업, 노동력 등) 특성을 조사함으로써 거시경제 분석 및 인력자원의 개발정책을 수립하는 데 필요한 기초자료를 수집·제공할 목적으로 1963년부터 조사가 진행되어 오고 있다⁸⁾.

경제활동인구조사의 절차는 Fig. 1과 같다. 매일 조사원들이 전국 표본조사가구를 방문하여 설문(visiting and interviewing)을 진행하고, 그 결과를 컴퓨터에 입력하면 지방통계청별로 조사 결과가 취합된다. 그리고 지방통계청별로 취합된 자료는 다시 중앙의 통계청으로 전달되며, 취합된 자료는 통계청 홈페이지를 통해 공개된다⁸⁾.

실업률 자료는 통계청의 경제활동인구조사 대상 항목 중 실업자(조사대상월 15일 기준 만 15세 이상인 경제활동인구 중 조사대상주간에 수입이 있는 일을 하지 않았고, 지난 4주간 일자리를 찾아 적극적으로 구직활동을 하였던 사람으로서 일자리가 주어지면 즉시 취업 가능한 사람들)와 경제활동인구(만 15세 이상 인구 중 취업자와 실업자)를 대상으로 아래의 Eq. (1)을 활용하여 산출한다⁹⁾.

Table 1. Economically Active Population Survey in Korea

	Description
Started	1963
Cycle	Month
Factors	State Employed/Unemployed, Demographic Factors, etc.
Method	Visiting and Interviewing

※ Source: Statistical Explanation DB of Statistics Korea, 2015



Fig. 1. Investigation Process for Economically Active Population in Korea

Table 2. Unemployment Rate (Unit: %)

	2010	2011	2012	2013	2014	2015
January	5.0	3.8	3.5	3.4	3.5	3.8
February	4.9	4.5	4.2	4.0	4.5	
March	4.1	4.3	3.7	3.5	3.9	
April	3.8	3.7	3.5	3.2	3.9	
May	3.2	3.2	3.1	3.0	3.6	
June	3.5	3.3	3.2	3.1	3.5	
July	3.7	3.3	3.1	3.1	3.4	
August	3.3	3.0	3.0	3.0	3.3	
September	3.4	3.0	2.9	2.7	3.2	
October	3.3	2.9	2.8	2.8	3.2	
November	3.0	2.9	2.8	2.7	3.1	
December	3.5	3.0	2.9	3.0	3.4	

※ Source: Statistical Explanation DB of Statistics Korea, 2015

$$\text{실업률}(\%) = (\text{실업자} / \text{경제활동인구}) \times 100 \quad (1)$$

Table 2는 최근 5년(2010년 1월~2015년 1월) 동안 통계청에서 조사한 실업률을 나타내고 있으며, 해당 자료를 통해 실업률이 매년 유사한 패턴으로 증감하는 계절성(seasonality)이 존재함을 알 수 있다.

3. 키워드 선정

실업과 관련된 인터넷 검색어를 선정하기 위해 1차적으로 2015년 1월 25일부터 2015년 2월 25일까지 최근 1개월 동안 생성된 SNS와 블로그 자료를 분석하였다. 이 중 총 8,005건(트위터: 5,291건, 블로그: 2,714건)에서 청년, 청년실업, 급여, 실업급여, 경제, 문제, 구직, 정부, 일자리, 기업 등의 키워드가 실업의 연관어로 검색되었으며, 총 7,442건(트위터: 5,232건, 블로그: 2,210건)에서 구인, 구인구직, 사이트, 취업, 구직단념자, 급여, 실업, 알바, 서류, 회사 등의 키워드가 구직의 연관어로 검색되었다. 이후 2차로 실업과 구직에 관한 연관 검색어 상위 12개를 선정하여 네이버 트렌드 상에서 검색 트렌드 값을 추출하였으며, 이 중 키워드 값이 0에 가깝게 나타나거나 실업률과 상관성이 낮은 키워드는 제외하였다. 최종적으로 ‘실업급여’, ‘구인구직’, ‘취업’ 3개의 키워드를 실업과 관련된 연관어로 선정하였으며 Fig. 2는 해당 키워드의 네이버 검색 트렌드를 보여주고 있다.

선정된 인터넷 검색어 자료는 주 단위로 생성되어 있기 때문에 월 단위의 실업률 자료와 함께 사용하기 위해서는 인터넷 검색어 자료를 월 단위로 변환하는 과정이 필요하다. 이에 본 연구에서는 Anvik and Gjelstad(2010)에서 사용했던 가중치 부여 방식을 활용하여 인터넷 검색

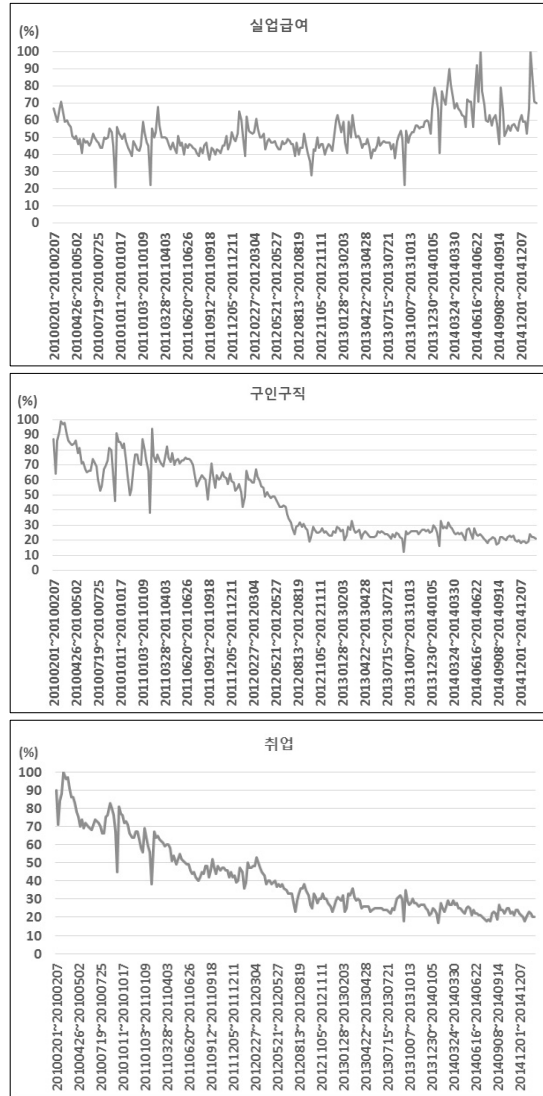


Fig. 2. Web Queries (Naver)

어 자료와 실업률 자료의 단위를 월 단위로 일치시켰다.

Table 3은 주 단위의 인터넷 검색어 값(key word value)을 월 단위의 값(transformed value)으로 변환하는 과정을 설명해 주고 있다. 통계청의 경우 조사대상월 15일 기준으로 실업률을 조사·발표하고 있기 때문에 인터넷 검색어를 월간 자료로 변환할 때 15일 이전 2주와 15일 이후 2주에 대한 자료를 활용하여 월 단위의 자료로 변환하였다(예-2010년 2월). 다만 1개월에 포함되는 날의 수가 매월 달라지기 때문에 이를 고려하기 위해 변환 월의 초와 말에 포함되는 주단위의 자료는 해당 월에 포함되는 날차

Table 3. Transformation of Weekly Key Word Data (Unit: %)

Month (Days)	Period (yyyymmdd)	Days	Weighted Key Word Rate(a)	Key Word Value(b)	(a)×(b)	Transformed Value
2010 February (28)	20100201~20100207	7	7/28	67	16.8	63.8
	20100208~20100214	7	7/28	62	15.5	
	20100215~20100221	7	7/28	59	14.8	
	20100222~20100228	7	7/28	67	16.8	
⋮		⋮				⋮
2010 April (30)	20100329~20100404	4	4/30	58	7.7	52.6
	20100405~20100411	7	7/30	56	13.1	
	20100412~20100418	7	7/30	51	11.9	
	20100419~20100425	7	7/30	49	11.4	
	20100426~20100502	5	5/30	51	8.5	

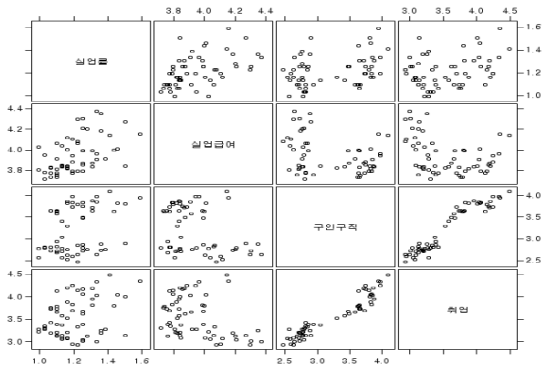


Fig. 3. Scatterplot Matrix for Unemployment Rate and Web Queries

만큼만 가중치(weighted rate)를 부여하여 월 단위의 자료로 변환하였다(예-2010년 4월).

이후, 실업률 자료와 연관검색어(실업급여, 구인구직, 취업) 자료의 상관관계를 Fig. 3과 같이 분석하였다. 분석 결과, 실업률과 연관검색어의 상관계수는 ‘실업급여’가 0.517, ‘구인구직’이 0.337, ‘취업’은 0.323으로 조사되었다. 이 중 연관검색어 ‘구인구직’과 ‘취업’의 상관계수가 0.952로 아주 높은 선형관계를 보여주고 있어, ‘취업’ 키워드를 제외한 ‘실업급여’와 ‘구인구직’ 검색어 자료만 실업률 추정모형에 입력변수로 활용하고자 한다.

4. 예측모형 및 결과분석

Choi and Varian(2009)과 Anvik and Gjelstad(2010)의 연구결과에 의하면 구글 트렌드의 인터넷 검색어 자료

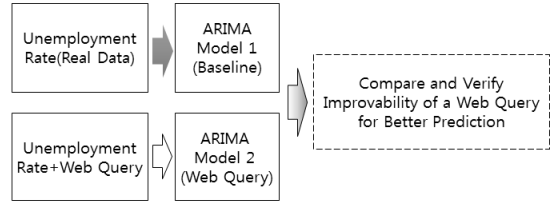


Fig. 4. Research Design

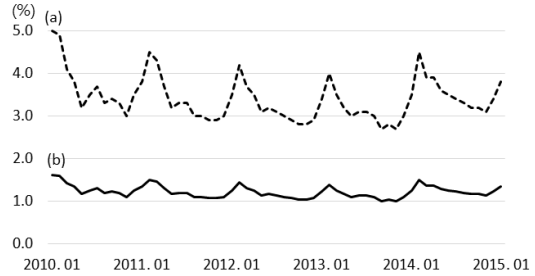


Fig. 5. Unemployment Rate (Statistics Korea)

는 실업률 등의 경제 관련 시계열 자료의 예측력을 제고하는 데 기여하는 것으로 조사되었다. 본 연구는 이러한 결과를 바탕으로 우리나라 검색시장에서 점유율이 76.9%(2014년 9월 4주차 기준)로 가장 높은 네이버 검색엔진을 사용하여 획득한 실업 관련 검색어 정보를 실업률을 예측하는 모형에 활용하는 방안을 연구하고자 한다.

이러한 목적으로 본 연구는 시계열 자료의 예측 모형에 적합한 ARIMA 모형을 활용하여 실업률 예측 모형을 개발하고자 한다.

Fig. 4는 본 연구의 예측모형 도출 과정을 체계적으로 나타내고 있다. 먼저 최근 5년 동안의 실업률을 기반으로 향후 실업률을 예측하는 ARIMA 모형 Model 1(baseline)과 Model 1에 사용된 동일한 실업률 자료에 네이버 트렌드의 실업 관련 인터넷 검색어 자료를 추가적으로 활용하는 Model 2를 제안한다. 그리고 이 두 모형의 실업률 예측력을 비교 분석하여 실업률 예측에 실업 관련 인터넷 검색어의 활용성결과를 평가하고자 한다.

4.1 ARIMA 모형

최근 5년간의 실업률을 나타내는 Fig. 5의 (a)에서 t 월의 실업률 y_t 에 대한 시계열 자료 $\{y_t : t = 1, 2, \dots, T\}$ 는 2010년부터 2013년까지 다소 감소하는 경향을 보이다가 2014년에 다시 증가하는 추이를 보여주고 있다. 연중으로는 2월에 수치가 가장 높게 나타나며, 12월 정도부터

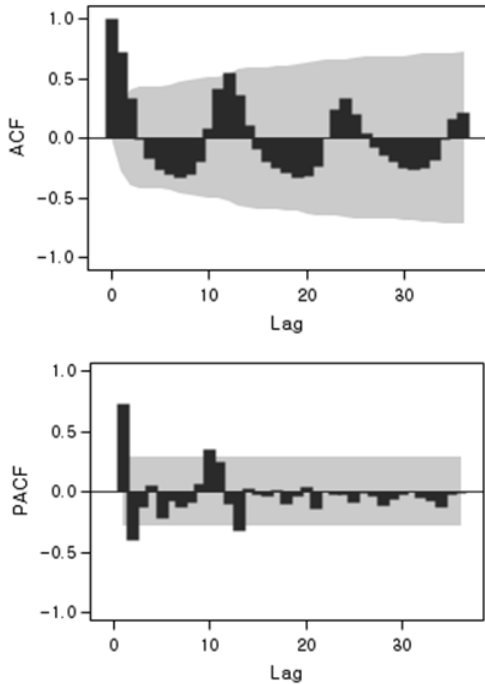


Fig. 6. ACF and PACF of $\log(y_t)$

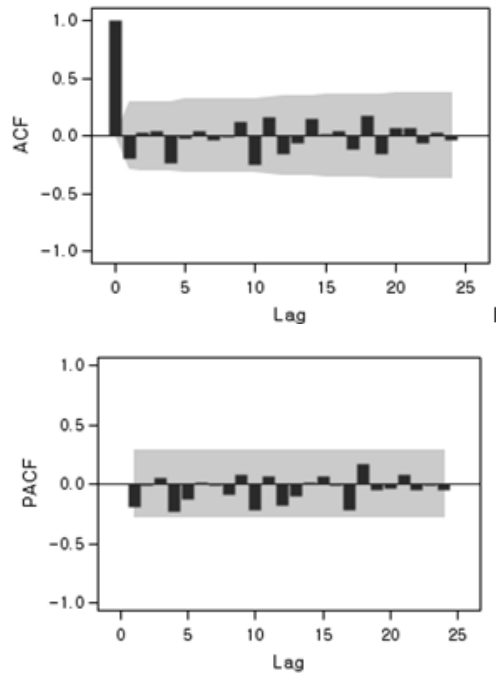


Fig. 7. ACF and PACF of ARIMA (1,1,0)s

이듬해 2월까지 증가한 뒤 다시 감소하는 계절적 특성을 나타내고 있다. 시계열 자료의 변동성이 이와 같이 나타나는 경우 원 자료를 사용하여 시계열 분석을 하는 것보다 $\log(y_t)$ 값으로 변환한 자료를 사용하여 분석하는 방법이 제안된다. 이는 $\log(y_t)$ 변환 자료가 원 자료에 비해 상대적으로 적은 변동을 보여주기 때문이다³¹. 실업률에 대한 시계열 $\log(y_t)$ 는 Fig. 5의 (b)와 같다.

시계열 자료 $\log(y_t)$ 을 살펴보면 12개월 단위로 주기를 형성하며, 하나의 주기 내에서는 원 시계열 $\{y_t\}$ 와 비슷한 경향을 보이고 있으며 변동량은 $\{y_t\}$ 보다 적게 나타난다. 감소 또는 증가 추세를 보이는 시계열에서 정상성(stationarity)을 고려한 적합한 ARIMA(p,d,q) 모형의 식별은 자기상관함수(ACF: Auto Correlation Function)과 편자기상관함수(PACF: Partial Auto Correlation Function) 등을 참조하여 차분의 필요성 여부 및 차수 d, AR 차수 p와 MA 차수 q가 결정된다³¹.

Fig. 6은 SAS 9.3을 이용하여 얻은 $\log(y_t)$ 시계열 자료의 ACF와 PACF이다. Fig. 6에서 ACF는 단조적으로 순환하면서 0을 향해 감소하고 있으므로 정상 시계열을 얻기 위해서는 적어도 1차 차분이 필요하다고 판단된다⁶¹.

또한 ACF는 12개월 단위로 값의 크기가 주기를 형성하고 있어 lag 12의 계절 차분 s가 추가로 필요하다고 사료된다. 본 연구는 계절 차분이 있는 ARIMA(p,d,q)s 시계열모형을 식별하기 위해 SAS 9.3에서 제공하는 SCAN(smallest canonical) 상관 방법을 사용하여 유의수준 5%에서 ARIMA(1,1,0)s 모형이 적절한 것으로 조사하였다.

SAS에서는 여러 가지 적합 가능한 모형들을 비교하여 최적모형을 선택하는 기준으로 사용하는 정보기준(Information Criterion)으로 AIC(Akaike Information Criterion)와 SBC(Schwarz's Bayesian criterion) 값을 제공한다²¹. AIC와 SBC의 정보기준 값이 작을수록 모형의 적합도가 높다고 볼 수 있는데 ARIMA(p,d,q)s 모형에서 적합도가 높은 모형을 대상으로 수행한 차분의 차수(d)와 AR의 차수(p) 그리고 MA의 차수(q) 중에서 p=1, d=1, q=0, s=12일 때 AIC(-135.20), SBC(-131.46)가 가장 적게 나타나 다른 모형에 비해 모형의 적합성이 가장 높게 나타남을 확인할 수 있었다. Fig. 7은 ARIMA(1,1,0)s 모형에 대한 ACF와 PACF를 나타낸 것으로 ACF와 PACF가 0에 가까워 연구 모형이 적절하다고 판단된다.

위의 결과를 바탕으로 t 번째 달의 실업률 예측모형을 계절적 자기회귀(seasonal AR) 모형으로 수립하고 다음

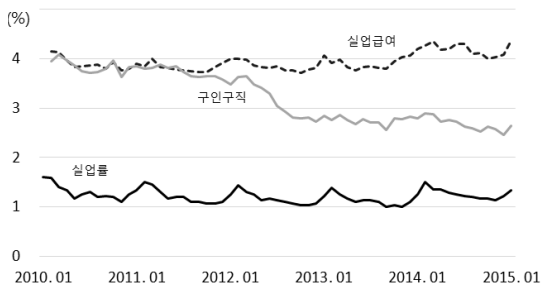


Fig. 8. Web Queries and Unemployment Rate

의 Eq. (2)와 같이 설정하였다. 여기서 e_t 는 모형의 오차항이다.

$$Model\ 1: \log(y_t) = \alpha_0 + \alpha_1 \log(y_{t-1}) + \alpha_2 \log(y_{t-12}) + e_t \quad (2)$$

Model 1의 LSE(least squares estimate)에 따른 추정 회귀모형은 다음 식과 같이 나타났다.

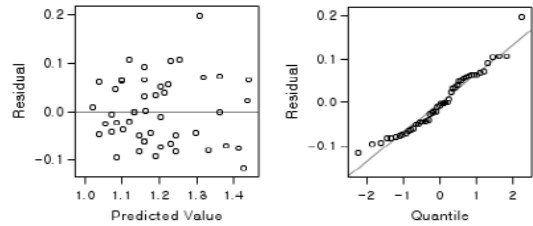
$$Model\ 1: \log(y_t) = 0.08563 + 0.461977 \log(y_{t-1}) + 0.46360 \log(y_{t-12}) \quad (3)$$

4.2 Web Query ARIMA 모형

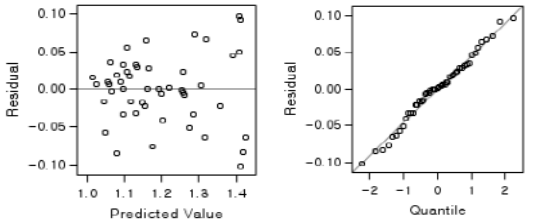
실업률을 예측하기 위해 개발한 Eq. (2)의 Model 1에 실업률과 연관성이 있는 네이버 트렌드의 시계열 자료를 추가하여 Web Query ARIMA 모형을 설정하고, 설정된 모형이 실업률 예측을 얼마나 개선할 수 있는지를 조사하고자 한다. 시점 t 에서 ‘실업’에 대한 n 개의 연관 검색어 자료, $w_t^{(k)}$ ($k = 1, 2, \dots, n$)을 Model 1에 추가하여 실업률을 추정하는 모형을 Model 2와 같이 수립하였다.

$$Model\ 2: \log(y_t) = \alpha_0 + \alpha_1 \log(y_{t-1}) + \alpha_2 \log(y_{t-12}) + \sum_{k=1}^n \beta_k \log(w_t^{(k)}) + e_t \quad (4)$$

Model 2에 포함될 연관 검색어를 선정하기 위해 본 연구는 연관 검색어인 실업급여, 구인구직에 대한 자료와 통계청의 실업률 자료(log 값으로 변환)를 시계열 plot으로 작성하였다(Fig. 8 참조). ‘실업급여’는 실업률 자료와 유사한 추세(계절적 특성)를 보이지만, ‘구인구직’은 시간 흐름에 따라 실업률과 서로 다른 패턴을 보여 본 연구에



(a) Model 1



(b) Model 2

Fig. 9. Regression Diagnostics Plot

서는 이를 제외하고, ‘실업급여’ 자료만 변수로 활용하여 Eq. (4)에서 $k = 1$ 으로 정하고 모형을 설정하였다.

Model 2의 LSE(least squares estimate)에 따른 추정 회귀모형은 Eq. (5)와 같이 나타났다.

$$Model\ 2: \log(y_t) = -0.97052 + 0.269531 \log(y_{t-1}) + 0.54667 \log(y_{t-12}) + 0.30016 \log(w_t^{(1)}) \quad (5)$$

추정된 회귀모형 Model 1과 Model 2의 적합성을 평가하기 위하여 작성된 추정 값과 잔차의 산점도, 잔차의 정규성 검토를 위한 회귀 진단 plot은 Fig. 9와 같다.

두 모형에서 잔차가 랜덤하게 흩어져 있고, 각 케이스의 점들이 정규직선에 근접하고 있어 모형의 선형성, 등분산성, 정규성이 만족한다고 볼 수 있다. Model 2는 Model 1에 비해 케이스의 점들이 정규직선에 더욱 가까이 위치해 있으며, 정규성 가정에 이상이 없음을 알 수 있다.

4.3 결과분석

실업률에 대한 예측모형의 정확도를 확인하기 위하여 Eq. (6)과 같이 예측치의 평균제곱오차 MSE(mean square error)를 계산하였다.

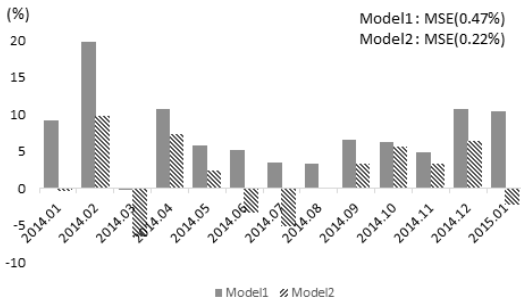


Fig. 10. Prediction Error Plot

$$MSE = \frac{1}{T} \sum_{t=1}^T PE_t^2 \quad (6)$$

여기서 T 는 전체기간이며, $PE_t = \log(\hat{y}_t) - \log(y_t)$ 이다. Model 1과 Model 2로부터 구한 MSE는 각각 0.47%와 0.22%로 나타난 평균제곱오차 측면에서 Model 2가 Model 1 보다 우수한 모형으로 추정된다.

Fig. 10은 예측 오차(prediction error) plot으로 연관검색어가 포함되어 있는 Model 2의 예측오차가 대부분 Model 1보다 적고, 2014년 2월을 제외한 모든 예측오차는 10% 미만임을 확인할 수 있었다.

Fig. 11은 실업률 실측치와 Model1 및 Model 2의 예

측 값에 대한 시계열 plot이다. 실선은 월별 실업률에 대한 실측치를, 점선은 Model 1로 예측한 결과를, 라인점선은 Model 2로 예측한 결과를 각각 나타낸다. Model 2로 예측한 결과가 실측치 실업률에 근접해 있어, ‘실업급여’ 연관 검색어가 포함되어 있는 Model 2가 Model 1보다 실업률 예측 정확도 측면에서 우수한 것으로 파악되었다.

5. 결 론

본 연구는 중요한 경제관련 지표 중 하나인 실업률을 보다 정확히 예측하는 방안으로 인터넷 검색 정보를 활용하는 방안을 조사하였다. 이를 위해 통계청의 실업률 자료를 기반으로 우선 월별 실업률을 예측하는 모형으로 계절적 자기회귀(seasonal AR) 모형을 설정하였다. 네이버 트렌드의 실업 관련 인터넷 검색어는 실업률과 관련성이 높은 두 개의 검색어를 선정하고 이 중에서 실업률 시계열 자료와 유사한 변화 추이를 보이는 ‘실업급여’ 검색어의 시계열 자료를 추정된 자기회귀 모형에 추가하여 최종적으로 우리나라의 실업률을 예측하는 모형을 개발하였다. 예측모형을 개발할 때는 자료를 실험용과 평가용으로 분할해서 사용하는 것이 일반적이다. 하지만 본 연구는 우수한 예측모형을 개발하는 것이 목적이 아니라 모형의 예측력을 개선하는 데 인터넷 검색어를 활용하는 것이 의미가 있는지를 확인하는 것이 목적이기 때문에 실업률 자

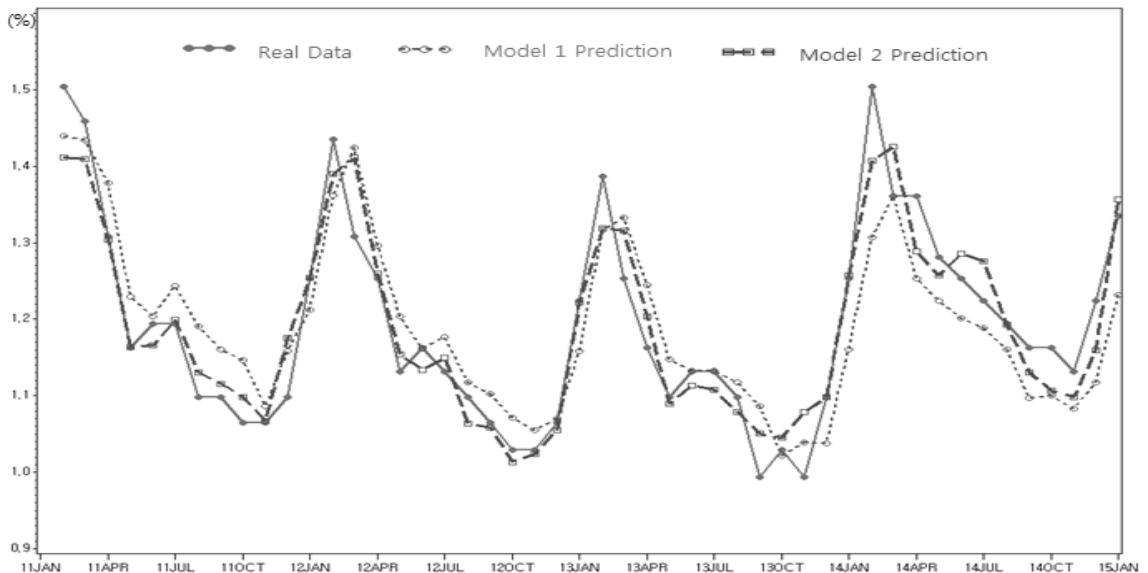


Fig. 11. Unemployment Rate Predictions based on Model 1 and Model 2

료를 실험용과 평가용으로 구분하지 않고 사용하였다.

모형 진단 결과, 계절적 자기회귀 모형에 ‘실업급여’ 연관 검색어를 추가한 월별 실업률 추정 모형은 타당한 것으로 나타났다. 추정된 모형으로부터 예측력을 비교·검증한 결과에서 연관 검색어(실업급여)가 포함된 모형이 그렇지 않은 모형보다 실업률 예측 정확도가 높게 나타났다. 이로부터 실업과 관련한 인터넷 검색어 자료를 활용하는 것이 한국의 실업률을 좀 더 정확하게 예측하는 데 도움이 된다는 것을 확인하였다. 이러한 연구결과로부터 향후 다양한 형태의 경제관련 지표의 작성이나 예측에 인터넷 검색어를 활용하는 연구가 필요할 것으로 사료된다.

한편, 본 연구에서는 실업 관련 인터넷 검색어의 실업 예측력 개선 유무를 확인하는 데 목적이 있었기 때문에 제한된 기간과 범위 내에서 관련 검색어를 탐색해서 선정하였다. 하지만 향후 우수한 실업 예측 모형을 개발하기 위해서는 인터넷 검색어의 탐색 기간과 범위를 넓혀 선택 키워드의 질을 개선하는 방안에도 연구가 필요하다.

References

1. Anvik, C. and Gjelstad, K., Just Google it: Forecasting Norwegian Unemployment Figures with Web Queries, Master's Thesis, BI Norwegian School of Management, 2010.
2. Beal, D. J., "Information Criteria Methods in SAS for Multiple Linear Regression Models", Proc. of the Southeast SAS User Group(SESUG), Hilton Head, SC, 2007.
3. Box, G. E. P., Jenkins, G. M. and Reinsel, G. C., Time Series Analysis: Forecasting and Control, 4th Edition, Wiley, 2008.
4. Choi, H. and Varian, H., Predicting Initial Claims for Unemployment Benefits, 2009. (<http://static.googleusercontent.com/media/research.google.com/en/archive/papers/initialclaimsUS.pdf>)
5. International Labour Organization, World Employment Social Outlook: Trends 2015, Geneva, 2015.
6. Montgomery, D. C and Johnson, L. A., Forecasting and Time Series Analysis, McGraw-Hill, 1976.
7. Park, S.H., Kim, C.Y. and Shin, Y.J., "The Effects of Employment Status Changes on Mental Health", Journal of Critical Social Welfare , Vol. 27, pp.79-120, 2009.
8. Statistical Explanation DB of Statistics Korea (<http://meta.narastat.kr/metascv/index.do?confmNo=10104&inputYear=2015>)
9. Statistics Korea, Economically Active Population Survey, 2015. (http://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=1063)
10. Upkorea (<http://www.upkorea.net/news/articleView.html?idxno=38400>)
11. Yun, W. S., "Testing the Effects of Unemployment and Income Gaps on Suicide: Time Series Analysis by Using Monthly Data from 1995 to 2008", Korean Association of Public Safety and Criminal Justice Review, Vol. 42, pp. 151-185, 2011.



권 치 명 (cmkwon@dau.ac.kr)

1978 서울대학교 산업공학과 학사
1981 서울대학교 산업공학과 석사
1991 VPI & SU Dept. of ISE 공학박사
1983~현재 동아대학교 교수

관심분야 : 시스템 모델링, Output Analysis, Simulation Optimization



황 성 원 (sense64@dau.ac.kr)

1986 동아대학교 응용통계학과 학사
1993 경성대학교 정보공학과 석사
2006 동아대학교 경영정보학과 박사
2011~현재 동아대학교 조교수

관심분야 : 정보시스템 평가, CRM, 데이터마이닝, 빅데이터



정 재 운 (imhere@dau.ac.kr)

2005 신라대학교 경영정보학과 학사
2008 동아대학교 경영정보학과 석사
2010 동아대학교 경영정보학과 박사
2013~현재 동아대학교 계약교수

관심분야 : 시스템다이나믹스, 물류정보시스템, 클라우드 컴퓨팅 & 빅데이터