



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0022144
(43) 공개일자 2020년03월03일

- | | |
|--|---|
| <p>(51) 국제특허분류(Int. Cl.)
 <i>G06Q 50/16</i> (2012.01) <i>G06F 40/00</i> (2020.01)
 <i>G06Q 10/06</i> (2012.01) <i>G06Q 30/02</i> (2012.01)</p> <p>(52) CPC특허분류
 <i>G06Q 50/16</i> (2013.01)
 <i>G06F 40/00</i> (2020.01)</p> <p>(21) 출원번호 10-2018-0097923
 (22) 출원일자 2018년08월22일
 심사청구일자 2018년08월22일</p> | <p>(71) 출원인
 울산과학기술원
 울산광역시 울주군 언양읍 유니스트길 50</p> <p>(72) 발명자
 김정섭
 울산광역시 울주군 언양읍 유니스트길 50, 울산과학기술원</p> <p>서정석
 울산광역시 울주군 언양읍 유니스트길 50, 울산과학기술원</p> <p>(74) 대리인
 오위환, 나성곤, 정기택</p> |
|--|---|

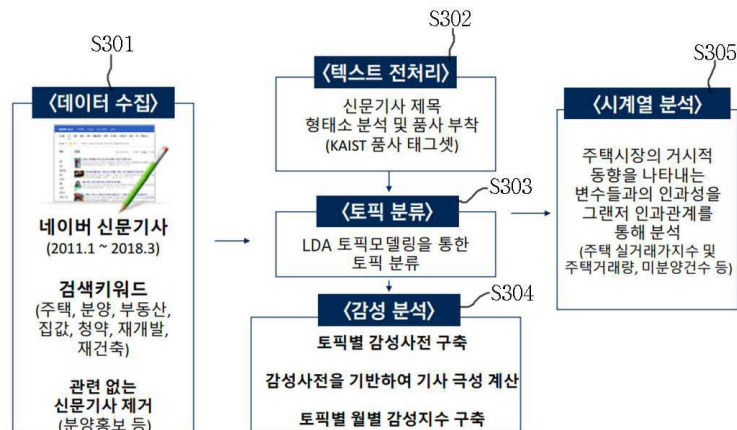
전체 청구항 수 : 총 13 항

(54) 발명의 명칭 **감성사전 구축을 통한 주택시장 동향 분석을 위한 시스템 및 방법**

(57) 요약

본 발명은 토픽별 감성사전 구축 및 감성사전을 기반으로 기사 극성 계산, 토픽별 월별 감성지수 구축을 통하여 시장의 거시적 동향을 나타내는 변수들과의 인과성을 정확하게 분석 및 예측할 수 있도록 한 감성사전 구축을 통한 주택시장 동향 분석을 위한 시스템 및 방법에 관한 것으로, 주택시장 관련 데이터를 수집하는 데이터 수집부; 상기 데이터 수집부에서 수집된 데이터에서 기사 제목의 형태소 분석 및 품사 부착을 하는 텍스트 전처리부; LDA(Latent Dirichlet Allocation) 토픽 모델링을 통한 토픽 분류를 하는 토픽 분류부; 상기 토픽 분류부의 토픽 분류 결과를 이용하여 토픽별 감성사전 구축하고 감성사전을 기반으로 기사 극성 계산 및 토픽별 월별 감성지수 구축을 하는 감성 분석부; 주택시장의 동향을 나타내는 변수들과의 인과성을 분석하는 시계열 분석부;를 포함하는 것을 특징으로 한다.

대 표 도 - 도3



(52) CPC특허분류

G06Q 10/0637 (2013.01)

G06Q 30/0201 (2013.01)

G06Q 30/0203 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호 1345280528

부처명 교육부

연구관리전문기관 한국연구재단

연구사업명 인문사회기초연구

연구과제명 텍스트 마이닝을 활용한 부동산 뉴스 감성분석 및 주택경기 예측 모형에의 적용

기 여 율 1/1

주관기관 울산과학기술원

연구기간 2018.05.01 ~ 2019.04.30

명세서

청구범위

청구항 1

주택시장 관련 데이터를 수집하는 데이터 수집부;

상기 데이터 수집부에서 수집된 데이터에서 기사 제목의 형태소 분석 및 품사 부착을 하는 텍스트 전처리부;

LDA(Latent Dirichlet Allocation) 토픽 모델링을 통한 토픽 분류를 하는 토픽 분류부;

상기 토픽 분류부의 토픽 분류 결과를 이용하여 토픽별 감성사전 구축하고 감성사전을 기반으로 기사 극성 계산 및 토픽별 월별 감성지수 구축을 하는 감성 분석부;

주택시장의 동향을 나타내는 변수들과의 인과성을 분석하는 시계열 분석부;를 포함하는 것을 특징으로 하는 감성사전 구축을 통한 주택시장 동향 분석을 위한 시스템.

청구항 2

제 1 항에 있어서, 감성 분석부는,

토픽 분류부의 토픽 분류 결과를 이용하여 토픽별 감성사전 구축하는 감성 사전 구축부와,

감성 사전 구축부에서 구축된 감성사전을 기반으로 기사 극성 계산을 하는 극성 계산부와,

극성 계산부에서 계산된 극성값을 이용하여 토픽별 월별 감성지수 구축을 하는 감성 지수 산출부를 포함하는 것을 특징으로 하는 감성사전 구축을 통한 주택시장 동향 분석을 위한 시스템.

청구항 3

제 1 항 또는 제 2 항에 있어서, 감성 분석부는 감성 분석을 통한 감성 사전 구축 및 지수 산출을 위하여,

신문 기사 제목을 기반으로 긍정, 부정, 중립으로 논조를 분류하여 논조를 결정짓는 단어 조합을 하여 감성 사전을 구축하고,

감성 사전 구축시에 동시 출현 매트릭스를 기반으로 하고, 잘못된 분류가 있는 경우에는 해당 기사를 제외하고, 매트릭스로 분류가 안된 기사는 직접 단어 조합을 결정하여 신문 기사별 감성 지수를 정량화하는 것을 특징으로 하는 감성사전 구축을 통한 주택시장 동향 분석을 위한 시스템.

청구항 4

제 2 항에 있어서, 감성 지수 산출부에서 산출된 감성 지수는,

기사의 논조 (감성, sentiment) = $\frac{\text{긍정(excitement) 패턴} - \text{부정(anxiety) 패턴}}{\text{신문기사 제목}(r)\text{의 형태소수}}$ 으로 정의되는 것을 특징으로 하는 감성사전 구축을 통한 주택시장 동향 분석을 위한 시스템.

청구항 5

제 1 항에 있어서, 시계열 분석부는 주택시장의 동향을 나타내는 변수들과의 인과성을 분석하기 위한 예측 모델 (Autoregressive and Distributed Lags;ARDL)을,

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_1 x_t + \cdots + \beta_q x_{t-q} + e_t \text{ 으로 정의하고,}$$

여기서, y는 주택동향 변수, x는 신문기사 논조의 감성 지수, q는 교차상관 분석 및 SIC(Schwarz Information Criteria)의 결과값이 최소화되는 시차이고, α_0 는 상수항(intercept)이고, α_1 은 종속변수의 시차변수(1개월 시차)의 추정 계수값, $\beta_1 \sim \beta_q$ 는 독립변수(보도논조 지수)의 시차변수(1~q개월 시차)의 추정 계수값이고,

신문기사 논조의 모든 과거값이 y 에 장기적으로 미치는 영향($\frac{\beta_1 + \dots + \beta_q}{1 - \alpha_1}$)을 고려하는 것을 특징으로 하는 감성사전 구축을 통한 주택시장 동향 분석을 위한 시스템.

청구항 6

주택시장 관련 데이터를 수집하는 데이터 수집 단계;

기사 제목의 형태소 분석 및 품사 부착을 하는 텍스트 전처리 단계;

LDA(Latent Dirichlet Allocation) 토픽 모델링을 통한 토픽 분류를 하는 토픽 분류 단계;

토픽 분류 단계를 통한 토픽 분류 결과를 이용하여 토픽별 감성사전 구축하고 감성사전을 기반으로 기사 극성 계산 및 토픽별 월별 감성지수 구축을 하는 감성 분석 단계;

주택시장의 동향을 나타내는 변수들과의 인과성을 분석하는 시계열 분석 단계;를 포함하는 것을 특징으로 하는 감성사전 구축을 통한 주택시장 동향 분석을 위한 방법.

청구항 7

제 6 항에 있어서, 토픽 분류 단계는 신문기사의 토픽이 단어의 집합으로 이루어져있다고 가정하고,

신문기사 제목을 형태소 단위로 분리하여 설정한 토픽수 만큼 토픽을 확률적으로 구하는 머신러닝 기법으로 토픽 모델링을 하는 것을 특징으로 하는 감성사전 구축을 통한 주택시장 동향 분석을 위한 방법.

청구항 8

제 6 항에 있어서, 감성 분석 단계는,

토픽 분류 단계의 토픽 분류 결과를 이용하여 토픽별 감성사전 구축하는 감성 사전 구축 단계와,

감성 사전 구축 단계에서 구축된 감성사전을 기반으로 기사 극성 계산을 하는 극성 계산 단계와,

극성 계산 단계에서 계산된 극성값을 이용하여 토픽별 월별 감성지수 구축을 하는 감성 지수 산출 단계를 포함하는 것을 특징으로 하는 감성사전 구축을 통한 주택시장 동향 분석을 위한 방법.

청구항 9

제 6 항에 있어서, 감성 분석 단계에서 감성 분석을 통한 감성 사전 구축 및 지수 산출을 위하여,

신문 기사 제목을 기반으로 긍정, 부정, 중립으로 논조를 분류하여 논조를 결정짓는 단어 조합을 하여 감성 사전을 구축하고,

감성 사전 구축시에 동시 출현 매트릭스를 기반으로 하고, 잘못된 분류가 있는 경우에는 해당 기사를 제외하고, 매트릭스로 분류가 안된 기사는 직접 단어 조합을 결정하여 신문 기사별 감성 지수를 정량화하는 것을 특징으로 하는 감성사전 구축을 통한 주택시장 동향 분석을 위한 방법.

청구항 10

제 8 항에 있어서, 감성 지수 산출 단계에서 산출된 감성 지수는,

기사의 논조 (감성, sentiment) = $\frac{\text{긍정(excitement) 패턴} - \text{부정(anxiety) 패턴}}{\text{신문기사 제목}(T)\text{의 형태소 수}}$ 으로 정의되는 것을 특징으로 하는 감성사전 구축을 통한 주택시장 동향 분석을 위한 방법.

청구항 11

제 6 항에 있어서, 시계열 분석 단계에서 주택시장의 동향을 나타내는 변수들과의 인과성을 분석하기 위한 예측 모델(Autoregressive and Distributed Lags;ARDL)을,

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_1 x_t + \dots + \beta_q x_{t-q} + e_t \text{ 으로 정의하고,}$$

여기서, y 는 주택동향 변수, x 는 신문기사 논조의 감성 지수, q 는 교차상관 분석 및 SIC(Schwarz Information Criteria)의 결과값이 최소화되는 시차이고, α_0 는 상수항(intercept)이고, α_1 은 종속변수의 시차변수(1개월 시차)의 추정 계수값, $\beta_1 \sim \beta_q$ 는 독립변수(보도논조 지수)의 시차변수(1~ q 개월 시차)의 추정 계수값이고,

신문기사 논조의 모든 과거값이 y 에 장기적으로 미치는 영향($\frac{\beta_1 + \dots + \beta_q}{1 - \alpha_1}$)을 고려하는 것을 특징으로 하는 감성사전 구축을 통한 주택시장 동향 분석을 위한 방법.

청구항 12

제 6 항에 있어서, 감성 분석 단계에서 단어의 출현 빈도로 긍정 또는 부정을 판단하지 않고,

동시출현 단어의 관계를 분석하여 긍정 또는 부정을 판단하는 것을 특징으로 하는 감성사전 구축을 통한 주택시장 동향 분석을 위한 방법.

청구항 13

제 12 항에 있어서, 동시출현 단어의 관계를 분석하는 과정에서,

시,도 단위 이상의 지역명, 검색어 키워드, 화폐 단위, 시간 단위에 관한 용어를 제외시키는 것을 특징으로 하는 감성사전 구축을 통한 주택시장 동향 분석을 위한 방법.

발명의 설명

기술 분야

[0001] 본 발명은 시장 동향 분석에 관한 것으로, 구체적으로 토픽별 감성사전 구축 및 감성사전을 기반으로 기사 극성 계산, 토픽별 월별 감성지수 구축을 통하여 시장의 거시적 동향을 나타내는 변수들과의 인과성을 정확하게 분석 및 예측할 수 있도록 한 감성사전 구축을 통한 주택시장 동향 분석을 위한 시스템 및 방법에 관한 것이다.

배경 기술

[0002] 정보통신기술의 발달로 인해 상거래, 소셜미디어(social media), 인터넷 블로그, 사물인터넷(internet of things) 등을 통해 생성된 수많은 빅데이터(big data)가 전자정보 형태로 실시간으로 기록 및 축적되고 있다.

[0003] 이 중 텍스트 데이터(text data), 사진, 동영상 등과 같은 비정형적(unstructured) 데이터가 빅데이터의 상당수를 차지하고 있는데 특히 텍스트 데이터에 포함된 정보를 분석 활용하기 위한 시도가 다양한 분야에서 활발히 이루어지고 있다.

[0004] 이처럼 텍스트 데이터로부터 유용한 정보를 추출해내는 과정을 텍스트 마이닝(text mining)이라 한다. 텍스트 마이닝 기법은 최근 다양한 분석 방법론의 개발, 정보처리 속도의 향상, 저장장치의 발달 등을 기반으로 급속히 발전하고 있으며 이에 대한 관심도 높아지고 있다.

[0005] 텍스트 데이터로부터 의미 있는 정보를 찾아낼 수 있을 것이라는 기대는 "통계적 의미론 가설(statistical semantics hypothesis)"에 근거한다.

[0006] 텍스트 마이닝의 기틀이 되는 이 가설은 "사람들의 글, 말 등에서 드러나는 단어 사용의 통계적 규칙성으로부터 사람들이 말하고자 하는 바를 찾아낼 수 있다"(Turney and Pantel, 2010)는 전제를 바탕으로 한다.

[0007] 하지만 텍스트 데이터는 벡터, 행렬 등의 수치형 자료와 같은 정형적인 형태를 가지고 있지 않아 기존 수치형 데이터(numerical data)에 비해 분석이 어렵다.

[0008] 또한 텍스트 데이터에서는 다양한 단어들이 복잡하고 미묘한 문법적 구조 안에서 표현되고, 동일한 문자로 기록된 단어도 표현 방식에 따라 다른 의미를 지닌다. 이러한 문법적인 구조와 의미의 다양성을 모두 수치형 자료로 표현하기 위해서는 매우 큰 차원의 벡터 또는 행렬이 필요하고, 이는 비효율적인 저장공간 활용, 계산시간의 기하급수적 증가 등의 문제를 야기하게 된다.

- [0009] 따라서, 의미의 손실이 다소 발생하더라도 일정한 가정 아래에서 텍스트 데이터를 가능한 간략한 수치형 자료로 표현하는 방법을 선택하는 것이 일반적이다.
- [0010] 최근, 부동산 시장의 불확실성이 높아짐에 따라 부동산 예측에 대한 사회적 관심이 확대되고 있다. 부동산 시장은 매우 다양한 요인에 의해 영향을 받으며, 특히 사람의 주관적인 심리 및 감성 등에 의해 영향을 많이 받으므로, 부동산 시장을 예측하기가 쉽지 않다.
- [0011] 이와 같은 부동산분야에서도 날로 증가할 것으로 예상되는 비정형데이터를 빅데이터 기술로 분석하여 부동산 활동의 효율성을 기하고 예측이 가능한 정보의 도출을 위한 기술들이 제안되고 있다.
- [0012] 특히, 뉴스 기사는 대중이 관심을 가지거나 대중에게 알려야 하는 정보를 정제된 표현으로 나타낸 글이다. 뉴스는 현실세계에 일어나는 각종 현상에 대한 설명과 미래의 정치, 경제, 사회, 기업 등과 관련하여 앞으로 어떠한 변화가 발생되고 진행되어 가는지에 대한 정보를 포함하고 있다.
- [0013] 따라서, 뉴스 기사 특히, 인터넷상에 업로드된 수많은 정보중 신뢰할 수 있고 중요하다고 여겨질 수 있는 정보중 하나인 인터넷 기사를 이용하여 부동산시장을 분석하고 예측하기 위한 연구 개발이 이루어지고 있다.
- [0014] 하지만, 실제 부동산과 주택(아파트)시장은 복잡한 매커니즘으로 인해 상호작용이 이루어진다. 많은 선행연구에서 부동산시장과 가격결정에 관하여 주택(아파트) 수요와 공급, 주택(아파트) 보급율, 경제 환경, 금리 등 정형화된 변수로 분석 예측하고 있으나 한계가 존재한다.
- [0015] 그리고 뉴스 기사의 수집 기간의 제한으로 인하여 감성분석을 통한 부동산의 흐름과 패턴의 검증하고 일반화하기에 충분하지 못한 문제가 있다.
- [0016] 또한, 지역별, 시기별 등 세밀한 분석을 하지 못하는 한계가 있고, 부동산에 특화된 감성사전 개발이 어려워 감성분석 결과 데이터의 유용성과 활용성이 떨어지는 문제가 있다.
- [0017] 따라서, 시차의 한계 극복과 외부의 이벤트의 영향 등 다양한 변수를 반영하고 샘플조사의 제한을 극복하여 정확도를 높일 수 있는 새로운 분석방법의 개발이 요구되고 있다.

선행기술문헌

특허문헌

- [0018] (특허문헌 0001) 대한민국 등록특허 제10-1849544호
(특허문헌 0002) 대한민국 등록특허 제10-1851788호
(특허문헌 0003) 대한민국 공개특허 제10-2018-0060871호

비특허문헌

- [0019] (비특허문헌 0001) Textmining에 의한 부동산 빅데이터 감성분석 모형 개발- 부동산 뉴스테이터에 의한 아파트 가격 예측 -한국부동산학회 2015년 정기학술대회

발명의 내용

해결하려는 과제

- [0020] 본 발명은 종래 기술의 텍스트 마이닝에 의한 시장 동향 분석 기술의 문제점을 해결하기 위한 것으로, 토픽별 감성사전 구축 및 감성사전을 기반으로 기사 극성 계산, 토픽별 월별 감성지수 구축을 통하여 시장의 거시적 동향을 나타내는 변수들과의 인과성을 정확하게 분석 및 예측할 수 있도록 한 감성사전 구축을 통한 주택시장 동향 분석을 위한 시스템 및 방법을 제공하는데 그 목적이 있다.
- [0021] 본 발명은 텍스트 마이닝 기법을 활용하여 신문기사를 토픽 별 논조를 분석하여 부동산 시장을 예측하기 위한 지표를 개발할 수 있도록 한 감성사전 구축을 통한 주택시장 동향 분석을 위한 시스템 및 방법을 제공하는데 그 목적이 있다.

- [0022] 본 발명은 언론의 부동산 시장 보도 논조가 주택소비심리 및 주택경기에 어떠한 영향을 미치는지 실증 분석을 하여 감성분석 결과 데이터의 유용성과 활용성을 높인 감성사전 구축을 통한 주택시장 동향 분석을 위한 시스템 및 방법을 제공하는데 그 목적이 있다.
- [0023] 본 발명은 신문기사 제목에 텍스트 마이닝 기법을 적용하여 본문 전체를 분석하는 것에 비해 형태소 분석 및 주택시장 맞춤형 감성사전 구축이 용이하도록 한 감성사전 구축을 통한 주택시장 동향 분석을 위한 시스템 및 방법을 제공하는데 그 목적이 있다.
- [0024] 본 발명은 개발된 심리지수를 주택가격, 거래량, 부동산심리지수와의 관계를 분석하고 시계열분석을 통해 주택 시장동향과의 인과관계를 검증하여 주택 시장 동향을 정확하게 분석 및 예측할 수 있도록 한 감성사전 구축을 통한 주택시장 동향 분석을 위한 시스템 및 방법을 제공하는데 그 목적이 있다.
- [0025] 본 발명은 단위근 검정과 시계열 그래프를 통해 분석 데이터의 시계열적 안정성을 검정하고, 비정상적 시계열은 1차 차분을 하여 변화량을 분석자료로 활용하여 시계열의 안정성을 확보할 수 있도록 한 감성사전 구축을 통한 주택시장 동향 분석을 위한 시스템 및 방법을 제공하는데 그 목적이 있다.
- [0026] 본 발명은 종속 변수가 자기상관 되어있을 경우, 설명변수에 시차(time lagged) 변수를 추가하여 화이트 노이즈 문제를 완화시키고, 예측모형 개발시에 인과성 및 시차확인으로 교차상관 분석 및 인과관계 검정을 하여 예측 정확성을 높인감성사전 구축을 통한 주택시장 동향 분석을 위한 시스템 및 방법을 제공하는데 그 목적이 있다.
- [0027] 본 발명의 다른 목적들은 이상에서 언급한 목적으로 제한되지 않으며, 언급되지 않은 또 다른 목적들은 아래의 기재로부터 당업자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

- [0028] 상기와 같은 목적을 달성하기 위한 본 발명에 따른 감성사전 구축을 통한 주택시장 동향 분석을 위한 시스템은 주택시장 관련 데이터를 수집하는 데이터 수집부;상기 데이터 수집부에서 수집된 데이터에서 기사 제목의 형태소 분석 및 품사 부착을 하는 텍스트 전처리부;LDA(Latent Dirichlet Allocation) 토픽 모델링을 통한 토픽 분류를 하는 토픽 분류부;상기 토픽 분류부의 토픽 분류 결과를 이용하여 토픽별 감성사전 구축하고 감성사전을 기반으로 기사 극성 계산 및 토픽별 월별 감성지수 구축을 하는 감성 분석부;주택시장의 동향을 나타내는 변수들과의 인과성을 분석하는 시계열 분석부;를 포함하는 것을 특징으로 한다.
- [0029] 여기서, 감성 분석부는, 토픽 분류부의 토픽 분류 결과를 이용하여 토픽별 감성사전 구축하는 감성 사전 구축부와,감성 사전 구축부에서 구축된 감성사전을 기반으로 기사 극성 계산을 하는 극성 계산부와,극성 계산부에서 계산된 극성값을 이용하여 토픽별 월별 감성지수 구축을 하는 감성 지수 산출부를 포함하는 것을 특징으로 한다.
- [0030] 그리고 감성 분석부는 감성 분석을 통한 감성 사전 구축 및 지수 산출을 위하여, 신문 기사 제목을 기반으로 긍정, 부정, 중립으로 논조를 분류하여 논조를 결정짓는 단어 조합을 하여 감성 사전을 구축하고, 감성 사전 구축 시에 동시 출현 매트릭스를 기반으로 하고, 잘못된 분류가 있는 경우에는 해당 기사를 제외하고, 매트릭스로 분류가 안된 기사는 직접 단어 조합을 결정하여 신문 기사별 감성 지수를 정량화하는 것을 특징으로 한다.
- [0031] 그리고 감성 지수 산출부에서 산출된 감성 지수는,

$$\text{기사의 논조 (감성, sentiment)} = \frac{\text{긍정(excitement) 패턴} - \text{부정(anxiety) 패턴}}{\text{신문기사 제목 (r)의 형태소수}}$$

- [0032] 으로 정의되는 것을 특징으로 한다.

- [0033] 그리고 시계열 분석부는 주택시장의 동향을 나타내는 변수들과의 인과성을 분석하기 위한 예측 모델(Autoregressive and Distributed Lags;ARDL)을,

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_1 x_t + \dots + \beta_q x_{t-q} + e_t$$

[0034] 으로 정의하고,

- [0035] 여기서, y는 주택동향 변수, x는 신문기사 논조의 감성 지수, q는 교차상관 분석 및 SIC(Schwarz Information Criterions)의 결과값이 최소화되는 시차이고, α_0 는 상수항(intercept)이고, α_1 은 종속변수의 시차변수(1개월 시차)의 추정 계수값, $\beta_1 \sim \beta_q$ 는 독립변수(보도논조 지수)의 시차변수(1~q개월 시차)의 추정 계수값이고,

- [0036] 신문기사 논조의 모든 과거값이 y 에 장기적으로 미치는 영향($\frac{\beta_1 + \dots + \beta_q}{1 - \alpha_1}$)을 고려하는 것을 특징으로 한다.
- [0037] 다른 목적을 달성하기 위한 본 발명에 따른 감성사전 구축을 통한 주택시장 동향 분석을 위한 방법은 주택시장 관련 데이터를 수집하는 데이터 수집 단계; 기사 제목의 형태소 분석 및 품사 부착을 하는 텍스트 전처리 단계; LDA(Latent Dirichlet Allocation) 토픽 모델링을 통한 토픽 분류를 하는 토픽 분류 단계; 토픽 분류 단계를 통한 토픽 분류 결과를 이용하여 토픽별 감성사전 구축하고 감성사전을 기반으로 기사 극성 계산 및 토픽별 월별 감성지수 구축을 하는 감성 분석 단계; 주택시장의 동향을 나타내는 변수들과의 인과성을 분석하는 시계열 분석 단계;를 포함하는 것을 특징으로 한다.
- [0038] 여기서, 토픽 분류 단계는 신문기사의 토픽이 단어의 집합으로 이루어져있다고 가정하고, 신문기사 제목을 형태소 단위로 분리하여 설정한 토픽수 만큼 토픽을 확률적으로 구하는 머신러닝 기법으로 토픽 모델링을 하는 것을 특징으로 한다.
- [0039] 그리고 감성 분석 단계는, 토픽 분류 단계의 토픽 분류 결과를 이용하여 토픽별 감성사전 구축하는 감성 사전 구축 단계와, 감성 사전 구축 단계에서 구축된 감성사전을 기반으로 기사 극성 계산을 하는 극성 계산 단계와, 극성 계산 단계에서 계산된 극성값을 이용하여 토픽별 월별 감성지수 구축을 하는 감성 지수 산출 단계를 포함하는 것을 특징으로 한다.
- [0040] 그리고 감성 분석 단계에서 감성 분석을 통한 감성 사전 구축 및 지수 산출을 위하여, 신문 기사 제목을 기반으로 긍정, 부정, 중립으로 논조를 분류하여 논조를 결정짓는 단어 조합을 하여 감성 사전을 구축하고, 감성 사전 구축시에 동시 출현 매트릭스를 기반으로 하고, 잘못된 분류가 있는 경우에는 해당 기사를 제외하고, 매트릭스로 분류가 안된 기사는 직접 단어 조합을 결정하여 신문 기사별 감성 지수를 정량화하는 것을 특징으로 한다.
- [0041] 그리고 감성 지수 산출 단계에서 산출된 감성 지수는,
- [0042]
$$\text{기사의 논조 (감성, sentiment)} = \frac{\text{긍정(excitement) 패턴} - \text{부정(anxiety) 패턴}}{\text{신문기사 제목}(r)\text{의 형태소 수}}$$
 으로 정의되는 것을 특징으로 한다.
- [0043] 그리고 시계열 분석 단계에서 주택시장의 동향을 나타내는 변수들과의 인과성을 분석하기 위한 예측 모델(Autoregressive and Distributed Lags; ARDL)을,
- [0044]
$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_1 x_t + \dots + \beta_q x_{t-q} + e_t$$
 으로 정의하고, 여기서, y 는 주택동향 변수, x 는 신문기사 논조의 감성 지수, q 는 교차상관 분석 및 SIC(Schwarz Information Criteria)의 결과값이 최소화되는 시차이고, α_0 는 상수항(intercept)이고, α_1 은 종속변수의 시차변수(1개월 시차)의 추정 계수값, $\beta_1 \sim \beta_q$ 는 독립변수(보도논조 지수)의 시차변수(1~ q 개월 시차)의 추정 계수값이고,
- [0045] 신문기사 논조의 모든 과거값이 y 에 장기적으로 미치는 영향($\frac{\beta_1 + \dots + \beta_q}{1 - \alpha_1}$)을 고려하는 것을 특징으로 한다.
- [0046] 그리고 감성 분석 단계에서 단어의 출현 빈도로 긍정 또는 부정을 판단하지 않고, 동시출현 단어의 관계를 분석하여 긍정 또는 부정을 판단하는 것을 특징으로 한다.
- [0047] 그리고 동시출현 단어의 관계를 분석하는 과정에서, 시,도 단위 이상의 지역명, 검색어 키워드, 화폐 단위, 시간 단위에 관한 용어를 제외시키는 것을 특징으로 한다.
- 발명의 효과**
- [0048] 이상에서 설명한 바와 같은 본 발명에 따른 감성사전 구축을 통한 주택시장 동향 분석을 위한 시스템 및 방법은 다음과 같은 효과가 있다.
- [0049] 첫째, 토픽별 감성사전 구축 및 감성사전을 기반으로 기사 극성 계산, 토픽별 월별 감성지수 구축을 통하여 시장의 거시적 동향을 나타내는 변수들과의 인과성을 정확하게 분석 및 예측할 수 있도록 한다.
- [0050] 둘째, 텍스트 마이닝 기법을 활용하여 신문기사를 토픽 별 논조를 분석하여 부동산 시장을 예측하기 위한 유용

한 지표를 개발할 수 있도록 한다.

- [0051] 셋째, 언론의 부동산 시장 보도 논조가 주택소비심리 및 주택경기에 어떠한 영향을 미치는지 실증 분석을 하여 감성분석 결과 데이터의 유용성과 활용성을 높일 수 있다.
- [0052] 넷째, 신문기사 제목에 텍스트 마이닝 기법을 적용하여 본문 전체를 분석하는 것에 비해 형태소 분석 및 주택시장 맞춤형 감성사전 구축이 용이하도록 한다.
- [0053] 다섯째, 개발된 심리지수를 주택가격, 거래량, 부동산심리지수와와의 관계를 분석하고 시계열분석을 통해 주택시장동향과의 인과관계를 검증하여 주택 시장 동향을 정확하게 분석 및 예측할 수 있도록 한다.
- [0054] 여섯째, 단위근 검정과 시계열 그래프를 통해 분석 데이터의 시계열적 안정성을 검정하고, 비정상적 시계열은 1차 차분을 하여 변화량을 분석자료로 활용하여 시계열의 안정성을 확보할 수 있도록 한다.
- [0055] 일곱째, 종속 변수가 자기상관 되어있을 경우, 설명변수에 시차(time lagged) 변수를 추가하여 화이트 노이즈 문제를 완화시키고, 예측모형 개발시에 인과성 및 시차확인으로 교차상관 분석 및 인과관계 검정을 하여 예측 정확성을 높인다.

도면의 간단한 설명

- [0056] 도 1은 본 발명에 따른 감성사전 구축을 통한 주택시장 동향 분석을 위한 시스템의 구성도
- 도 2는 감성 분석부의 상세 구성도
- 도 3은 본 발명에 따른 감성사전 구축을 통한 주택시장 동향 분석을 위한 방법을 나타낸 플로우 차트
- 도 4는 본 발명에 따른 감성 사전의 긍정 및 부정 패턴의 일 예를 나타낸 구성도
- 도 5는 본 발명의 일 실시 예에 따른 감성 분석을 통한 지수 산출 과정을 나타낸 플로우 차트
- 도 6a 내지 도 6c는 주택시장 관련 데이터의 토픽 분류에 따른 분석 결과를 나타낸 구성도
- 도 7은 긍정 뉴스의 동시 출현 단어 관계도
- 도 8은 부정 뉴스의 동시 출현 단어 관계도
- 도 9a와 도 9b는 본 발명에 따른 아파트 가격 감성 지수에 따른 매매 거래량 및 매매 가격지수의 시계열 패턴
- 도 10은 본 발명에 따른 감성 지수와 주택매매량의 교차상관분석에 따른 결과 그래프
- 도 11은 본 발명에 따른 감성지수와 미분양건수의 교차상관분석에 따른 결과 그래프
- 도 12는 본 발명에 따른 감성지수와 아파트매매가격 지수의 교차상관분석에 따른 결과 그래프
- 도 13은 본 발명에 따른 감성지수(주택시장)와 주택미분양의 교차상관분석에 따른 결과 그래프

발명을 실시하기 위한 구체적인 내용

- [0057] 이하, 본 발명에 따른 감성사전 구축을 통한 주택시장 동향 분석을 위한 시스템 및 방법의 바람직한 실시 예에 관하여 상세히 설명하면 다음과 같다.
- [0058] 본 발명에 따른 감성사전 구축을 통한 주택시장 동향 분석을 위한 시스템 및 방법의 특징 및 이점들은 이하에서의 각 실시 예에 대한 상세한 설명을 통해 명백해질 것이다.
- [0059] 도 1은 본 발명에 따른 감성사전 구축을 통한 주택시장 동향 분석을 위한 시스템의 구성도이고, 도 2는 감성 분석부의 상세 구성도이다.
- [0060] 본 발명에 따른 감성사전 구축을 통한 주택시장 동향 분석을 위한 시스템 및 방법은 토픽별 감성사전 구축 및 감성사전을 기반으로 기사 극성 계산, 토픽별 월별 감성지수 구축을 통하여 시장의 거시적 동향을 나타내는 변수들과의 인과성을 정확하게 분석 및 예측할 수 있도록 한 것이다.
- [0061] 이를 위하여 본 발명은 텍스트 마이닝 기법을 활용하여 신문기사를 토픽 별 논조를 분석하여 부동산 시장을 예측하기 위한 지표를 개발하여 이를 활용하는 구성을 포함할 수 있다.
- [0062] 본 발명은 감성분석 결과 데이터의 유용성과 활용성을 높이기 위하여 언론의 부동산 시장 보도 논조가 주택소비

심리 및 주택경기에 어떠한 영향을 미치는지 실증 분석을 하는 구성을 포함할 수 있다.

[0063] 본 발명은 신문기사 제목에 텍스트 마이닝 기법을 적용하여 형태소 분석 및 주택시장 맞춤형 감성사전 구축이 용이하도록 하고, 개발된 심리지수를 주택가격, 거래량, 부동산심리지수와 관계 분석하고 시계열분석을 통해 주택시장동향과의 인과관계를 검증하는 구성을 포함할 수 있다.

[0064] 본 발명에 따른 감성사전 구축을 통한 주택시장 동향 분석을 위한 시스템은 도 1에서와 같이, 주택시장 관련 데이터를 수집하는 데이터 수집부(10)와, 기사 제목의 형태소 분석 및 품사 부착을 하는 텍스트 전처리부(20)와, LDA(Latent Dirichlet Allocation) 토픽 모델링을 통한 토픽 분류를 하는 토픽 분류부(30)와, 토픽 분류부(30)의 토픽 분류 결과를 이용하여 토픽별 감성사전 구축하고 감성사전을 기반으로 기사 극성 계산 및 토픽별 월별 감성지수 구축을 하는 감성 분석부(40)와, 주택시장의 동향을 나타내는 변수들과의 인과성을 분석하는 시계열 분석부(50)를 포함한다.

[0065] 여기서, 감성 분석부(40)의 상세 구성은 도 2에서와 같이, 토픽 분류부(30)의 토픽 분류 결과를 이용하여 토픽별 감성사전 구축하는 감성 사전 구축부(41)와, 감성 사전 구축부(41)에서 구축된 감성사전을 기반으로 기사 극성 계산을 하는 극성 계산부(42)와, 극성 계산부(42)에서 계산된 극성값을 이용하여 토픽별 월별 감성지수 구축을 하는 감성 지수 산출부(43)를 포함한다.

[0066] 감성 지수 산출부(43)에서 산출된 감성 지수는 수학적 1에서와 같이 정의될 수 있다.

수학적 1

$$\text{기사의 논조 (감성, sentiment)} = \frac{\text{긍정(excitement) 패턴} - \text{부정(anxiety) 패턴}}{\text{신문기사 제목}(r)\text{의 형태소 수}}$$

[0067]

[0068] 그리고 시계열 분석부(50)는 주택시장의 동향을 나타내는 변수들과의 인과성을 분석하기 위하여 그랜저(Granger) 인과관계 분석을 한다.

[0069] 원인과 결과가 불분명한 경우에는 함수관계에 대해 명확한 결정을 내릴 수 없기 때문에 시차분포모형(distributed lag model)을 이용하여 원인과 결과를 확인할 수 있는 인과관계 검정(causality test)으로 그랜저(CWJ Granger, 1969)에 의해 개발된 방법을 적용한다.

[0070] 시차분포모형이란 회귀모형이 설명변수의 현재 관측값은 물론 과거의 관측값까지 포함하고 있는 설명변수들로 정의된 경우의 모형을 의미한다.

[0071] 본 발명에 따른 감성사전 구축을 통한 주택시장 동향 분석을 위한 방법을 구체적으로 설명하면 다음과 같다.

[0072] 도 3은 본 발명에 따른 감성사전 구축을 통한 주택시장 동향 분석을 위한 방법을 나타낸 플로우 차트이고, 도 4는 본 발명에 따른 감성 사전의 긍정 및 부정 패턴의 일 예를 나타낸 구성도이다.

[0073] 본 발명에 따른 감성사전 구축을 통한 주택시장 동향 분석을 위한 방법은 도 3에서와 같이, 주택시장 관련 데이터를 수집하는 데이터 수집 단계(S301)와, 기사 제목의 형태소 분석 및 품사 부착을 하는 텍스트 전처리 단계(S302)와, LDA(Latent Dirichlet Allocation) 토픽 모델링을 통한 토픽 분류를 하는 토픽 분류 단계(S303)와, 토픽 분류 단계를 통한 토픽 분류 결과를 이용하여 토픽별 감성사전 구축하고 감성사전을 기반으로 기사 극성 계산 및 토픽별 월별 감성지수 구축을 하는 감성 분석 단계(S304)와, 주택시장의 동향을 나타내는 변수들과의 인과성을 분석하는 시계열 분석 단계(S305)를 포함한다.

[0074] 여기서, 토픽 분류 단계(S303)는 신문기사의 토픽이 단어의 집합으로 이루어져있다고 가정하고, 신문기사 제목을 형태소 단위로 분리하여 설정한 토픽수 만큼 토픽을 확률적으로 구하는 머신러닝 기법으로 토픽 모델링을 한다.

[0075] 그리고 감성 분석 단계(S304)는, 토픽 분류 단계의 토픽 분류 결과를 이용하여 토픽별 감성사전 구축하는 감성사전 구축 단계와, 감성 사전 구축 단계에서 구축된 감성사전을 기반으로 기사 극성 계산을 하는 극성 계산 단계와, 극성 계산 단계에서 계산된 극성값을 이용하여 토픽별 월별 감성지수 구축을 하는 감성 지수 산출 단계를 포함할 수 있다.

[0076] 그리고 감성 사전 구축 단계에서 구축된 감성 사전은 도 4에서와 같이, 토픽을 '아파트 가격'으로 하는 경우에,

[경충, 훈풍, 활기, 돌파, 뛰어, 고공행진, 뿔박질, 무섭게 상승, 주춤했다 상승, 다시 상승세, 만에 상승, 상승 + 큰폭, 상승 + 커, 상승 + 최대, 많이 + 올] 등의 단어가 긍정 패턴으로 분류된다.

[0077] 그리고 [제자리, 과열, 꺾, 절반, 축소, 위축, 반토막, 경고, 거품, 다시 하락세, 정부규제 칼, 대책 주시, 눈치 + 보기, 상승 + 정체, 거래량 + 감소, 추가 + 규제] 등의 단어가 부정 패턴으로 분류된다.

[0078] 표 1은 감성 사전 구축시에 긍정 패턴으로 분류되는 단어 리스트를 나타낸 것이다.

표 1

Positive Sentiment Dictionary			
↑	뛰었	수혜	최고치
[^가구] 육박	똥	숨통 트이나	최대폭 상승
가치 상승	뿔박질	시대	큰.*상승
가팔라	뜨거	쑥쑥	투기 재연
강세	띄우기	앞질러	튀네
갠	만에 상승	악발 끝났나	필필
거래.*활발	만에 오름	악발은	풍부한
거품 아니다	많이 올	에도 상승	하락세.*멈
高	무섭게 상승	역대.*최고	하락하던
고개드는	바닥쳤다	역세권 불패	호재
고공행진	반등	연속 상승	활기
과열.*해제	부글부글	연속 오름	활황기
급등	부터 상승	오르나	후끈
급증	불붙은	오를 것	훈풍
경충	불확실성 해소	오를것	The number of patterns: 91
꿈틀	사상.*최고	오름.*커	
넘어	상승 기대감에	오름세 전환	
높은 주거비	상승 기대에	온기	
누적 상승률	상승 이어져	올라	
다시 상승세	상승.*진입	올랐다	
돌파	상승.*최대	육박	
동반 상승	상승.*커	잡으려	
들썩	상승.*큰폭	재건축	
떨어지던	상승.*확산	주춤했다 상승	
뛰네	상승세 번	지금이 바닥	
뛰는	상승폭 더	천장 뚫	
뛰어	상승폭.*확대	최고.*상승	

[0079]

[0080] 표 2는 감성 사전 구축시에 부정 패턴으로 분류되는 단어 리스트를 나타낸 것이다.

표 2

Negative Sentiment Dictionary			
↓	다시 하락세	안정세	추락
8 2.*대책	대책 임박	악발 계속	축소
감감	대책 주시	악세	침체
강력 규제	대출 규제	엇갈려	폭락
거래 한산	둔화	움짤	풍선효과.*차단
거래량.*감소	떨군	위기	하락
거품	떨어	위축	하우스푸어
결국.*내리막	뚝뚝	잠잠	한숨
경고	마지노선	잡기	한파
곤두박질	머뭇	잡는다	The number of patterns: 90
과열	멈춰	전세난	
그쳐	멈췄	절반	
그칠	모니터링	정부규제 칼	
금리.*인상	무너	제자리	
급감	미분양 늘	좁	
급락	바닥	주춤	
깎	반토막	줄어	
강통	변수	줄었	
깨	부정사전	집값.*빠	
꺾	붕괴	질은	
공공	빙하기	쪼그라든	
낙폭	상승 멈	찢끔	
내려	상승.*정체	찬바람	
내렸	상승.*제한	천장.*찢나	
내림	속탄다	최저	
눈치 보기	숨고르기	추가.*규제	
눈치보기	악몽	추가.*대책	

[0081]

[0082]

도 5는 본 발명의 일 실시 예에 따른 감성 분석을 통한 지수 산출 과정을 나타낸 플로우 차트이다.

[0083]

그리고 도 6a 내지 도 6c는 주택시장 관련 데이터의 토픽 분류에 따른 분석 결과를 나타낸 구성도이고, 도 7은 긍정 뉴스의 동시 출현 단어 관계도이고, 도 8은 부정 뉴스의 동시 출현 단어 관계도이다.

[0084]

본 발명은 감성 분석을 통한 감성 사전 구축 및 지수 산출을 위하여, 신문 기사 제목을 기반으로 긍정, 부정, 중립으로 논조를 분류한다.

[0085]

그리고 논조를 결정짓는 단어 조합을 하여 감성 사전을 구축하고, 감성 사전 구축시에 동시 출현 매트릭스를 기반으로 하고, 잘못된 분류가 1건이라도 있는 경우에는 해당 기사를 제외하고, 매트릭스로 분류가 안된 기사는 직접 단어 조합을 결정한다.

[0086]

그리고 신문 기사별 감성 지수를 정량화하기 위하여 수학적 1에서와 같은 방법으로 감성 지수를 산출한다.

[0087]

도 5는 본 발명의 일 실시 예에 따른 감성 분석을 통한 지수 산출 과정으로, 불용어의 제거, 형태소 분석, 감성 분석을 하여 감성 지수를 산출하는 것을 나타낸 것이다.

[0088]

본 발명에 따른 감성사전 구축을 통한 주택시장 동향 분석을 위한 시스템 및 방법에서는 시계열의 안정성을 위하여 단위근 검정과 시계열 그래프를 통해 분석 데이터의 시계열적 안정성을 검정하고, 비정상적 시계열은 1차 차분을 하여 변화량을 분석자료로 활용한다.

[0089]

그리고 종속 변수가 자기상관 되어있을 경우, 설명변수에 시차(time lagged) 변수를 추가하여 화이트 노이즈 문제를 완화시킨다.

[0090] 그리고 예측모형 개발시에 인과성 및 시차확인으로 교차상관 분석 및 인과관계 검정을 하여 예측 정확성을 높인다.

[0091] 본 발명에 따른 예측 모형(Autoregressive and Distributed Lags;ARDL)은 수학적 2에서와 같이 정의될 수 있다.

수학적 2

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_1 x_t + \dots + \beta_q x_{t-q} + e_t$$

[0093] 여기서, y 는 주택동향 변수(매매량,가격지수,미분양 등), x 는 신문기사 논조의 감성 지수, q 는 교차상관 분석 및 SIC(Schwarz Information Criterions)의 결과값이 최소화되는 시차이고, α_0 는 상수항(intercept)이고, α_1 은 종속변수의 시차변수(1개월 시차)의 추정 계수값, $\beta_1 \sim \beta_q$ 는 독립변수(보도논조 지수)의 시차변수(1~ q 개월 시차)의 추정 계수값이고, 신문기사 논조의 모든 과거값이 y 에 장기적으로 미치는 영향($\frac{\beta_1 + \dots + \beta_q}{1 - \alpha_1}$)을 고려한다.

[0094] 이와 같이 α_1 은 종속변수의 시차변수(1개월 시차)의 추정 계수값이고, 종속변수는 보도논조지수가 아닌 각종 주택동향 변수이다.

[0095] 주택가격과 같은 주택 동향 변수는 과거값이 현재에 영향을 미치기 때문에 포함된 값으로, 즉, 1달 전 주택동향이 현재 주택동향에 얼마나 자기 상관 되었는지 (얼마나 영향을 미치는 지)를 의미한다.

[0096] 그리고 $\beta_1 \sim \beta_q$ 는 독립변수 (보도논조 지수)의 시차변수 (1~ q 개월 시차)의 추정 계수값으로, 본 발명에 따른 예측 모형(Autoregressive and Distributed Lags;ARDL)을 해석 할 때는 독립변수의 모든 과거치를 함께 고려한다고 가정한다.

[0097] 즉, 독립변수의 모든 과거치가 현재 주택동향에 얼마나 영향을 미치는지 계산할 때 사용되는 것으로, 영향의 정도는 $\frac{\beta_1 + \dots + \beta_q}{1 - \alpha_1}$ 으로 계산될 수 있다.

[0098] 도 6a 내지 도 6c는 주택시장 관련 데이터의 토픽 분류에 따른 분석 결과를 나타낸 것으로, 명사 기반의 분석 결과를 보면 도 6a에서와 같이 관련 지역은 서울,강남,수도권 관련 기사가 큰 비중을 차지하고, 관련 내용은 가격의 상승 또는 하락 여부, 재건축 관련 기사, 부동산 대책 및 규제 관련 기사가 높은 비중을 차지하는 것을 알 수 있다.

[0099] 도 6b는 긍정 기사의 분석 결과를 나타낸 것이고, 도 6c는 부정 기사의 분석 결과를 나타낸 것이다.

[0100] 아파트 가격을 부정적인 논조로 쓴 기사의 경우 '상승'과 같은 긍정적인 단어의 출현 빈도도 상당히 높은 것을 나타내고 있어 단순히 상승/하락 같은 단어의 출현 빈도로 긍정/부정을 판단하는 것은 정확성을 저하시키게 된다.

[0101] 이 문제를 해결하기 위하여 본 발명에서는 동시출현 단어의 관계를 분석한다.

[0102] 도 7은 긍정 뉴스의 동시 출현 단어 관계를 나타낸 것이고, 도 8은 부정 뉴스의 동시 출현 단어 관계를 나타낸 것으로, 시,도 단위 이상의 지역명, 아파트 가격과 같은 검색어 키워드, 화폐 단위, 시간 단위 등의 용어를 제외한다.

[0103] 본 발명에 따른 감성사전 구축을 통한 주택시장 동향 분석을 위한 시스템 및 방법을 적용한 시계열 패턴 분석 결과에 관하여 설명하면 다음과 같다.

[0104] 도 9a와 도 9b는 본 발명에 따른 아파트 가격 감성 지수에 따른 매매 거래량 및 매매 가격지수의 시계열 패턴이다.

[0105] 그리고 도 10은 본 발명에 따른 감성 지수와 주택매매량의 교차상관분석에 따른 결과 그래프이다.

[0106] 전체적으로 감성 지수와 주택매매량의 두 변수가 동행하는 경향을 나타내고, 1달 전 감성지수와 이번 달 주택매매거래량(변화량)의 상관계수가 가장 높게 나타나는 것을 알 수 있다.

[0107] 도 11은 본 발명에 따른 감성지수와 미분양건수의 교차상관분석에 따른 결과 그래프이다.

[0108] 비교적 긴 시차(12달)이 지나도 감성지수와 미분양건수의 상관계수가 높으며 감성지수가 선행하는 경향을 보인다.

[0109] 그리고 미분양 건수의 변화량은 감성지수를 후행보다 선행하는 경향을 나타내는 것을 확인할 수 있다.

[0110] 도 12는 본 발명에 따른 감성지수와 아파트매매가격 지수의 교차상관분석에 따른 결과 그래프이다.

[0111] 미분양 건수와 유사하게 감성지수가 간 시차(12달)를 뒤도 상관계수가 크게 떨어지지 않으며 감성지수가 선행하는 경향을 나타낸다.

[0112] 반면, 가격지수가 감성지수에 선행할 경우에는 상관계수가 크게 떨어지는 것을 알 수 있고, 가격지수의 변화량과 한달 전 감성지수의 상관계수가 가장 높은 것을 알 수 있다.

[0113] 도 13은 본 발명에 따른 감성지수(주택시장)와 주택미분양의 교차상관분석에 따른 결과 그래프이다.

[0114] 신문기사 논조 기반의 감성지수와 국토연구원에서 설문을 통해 매달 발표하는 소비자 심리지수(주택시장)의 값은 시차에서 가장 상관계수가 높다.

[0115] 또한, 소비자 심리지수(주택시장)의 변화량도 감성지수와 같은 시차에서 가장 상관계수가 높다.

[0116] 이상에서 설명한 본 발명에 따른 감성사전 구축을 통한 주택시장 동향 분석을 위한 시스템 및 방법은 신문기사 제목에 텍스트 마이닝 기법을 적용하여 형태소 분석 및 주택시장 맞춤형 감성사전 구축이 용이하도록 하고, 개발된 심리지수를 주택가격, 거래량, 부동산심리지수와와의 관계를 분석하고 시계열분석을 통해 주택시장동향과의 인과관계를 검증하여 주택 시장 동향을 정확하게 분석 및 예측할 수 있도록 한 것이다.

[0117] 이상에서의 설명에서와 같이 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 본 발명이 구현되어 있음을 이해할 수 있을 것이다.

[0118] 그러므로 명시된 실시 예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 하고, 본 발명의 범위는 전술한 설명이 아니라 특허청구 범위에 나타나 있으며, 그와 동등한 범위 내에 있는 모든 차이점은 본 발명에 포함된 것으로 해석되어야 할 것이다.

부호의 설명

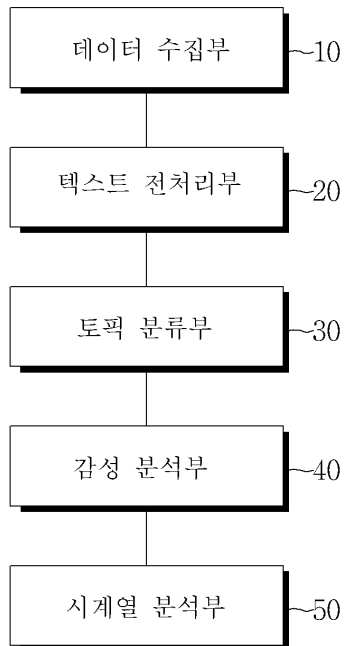
[0119] 10. 데이터 수집부 20. 텍스트 전처리부

30. 토목 분류부

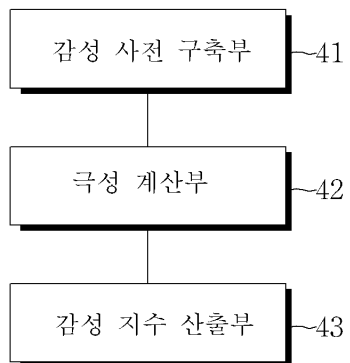
50. 시계열 분석부

도면

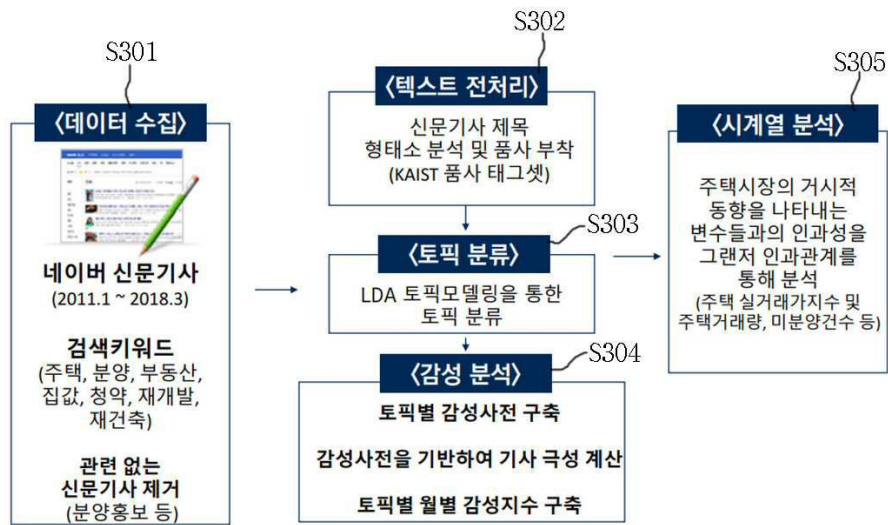
도면1



도면2



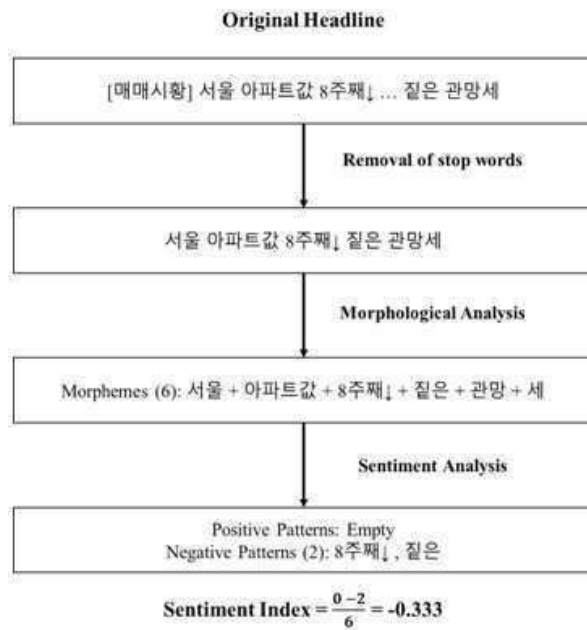
도면3



도면4

긍정 패턴	✓ 께충, 훈풍, 활기, 돌파, 뛰어, 고공행진, 뽐박질,... ✓ 무섭게 상승, 주째 상승, 다시 상승세, 만에 상승,... ✓ 상승 + 큰폭, 상승 + 커, 상승 + 최대, 많이 + 올...
부정 패턴	✓ 제자리, 과열, 꺾, 절반, 축소, 위축, 반토막, 경고, 거품,... ✓ 다시 하락세, 정부규제 칼, 대책 주시, 눈치+보기... ✓ 상승 + 정체, 거래량 + 감소, 추가 + 규제, ...

도면5



도면6a



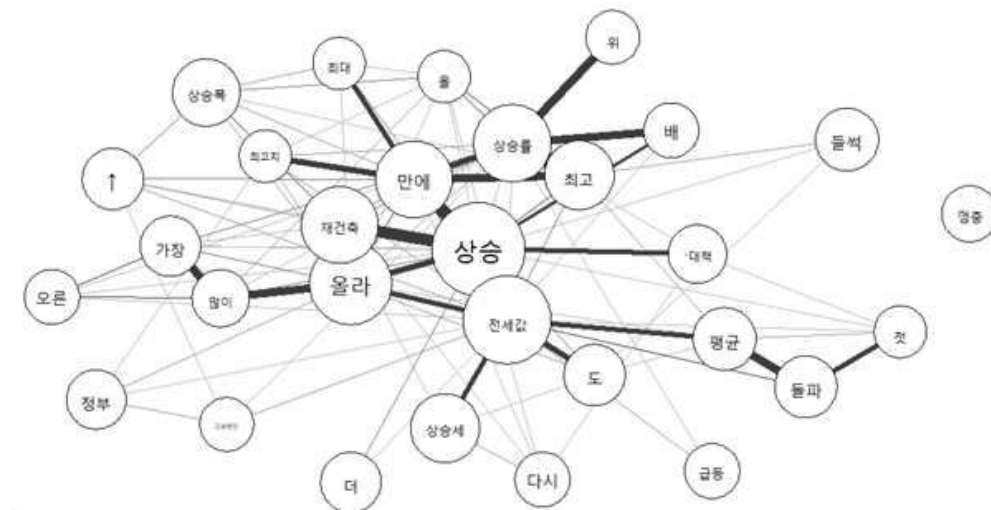
도면6b



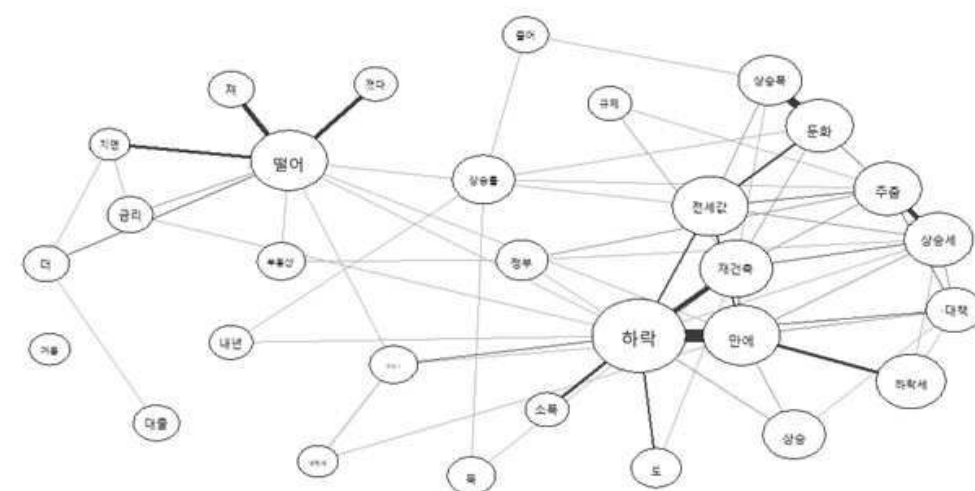
도면6c



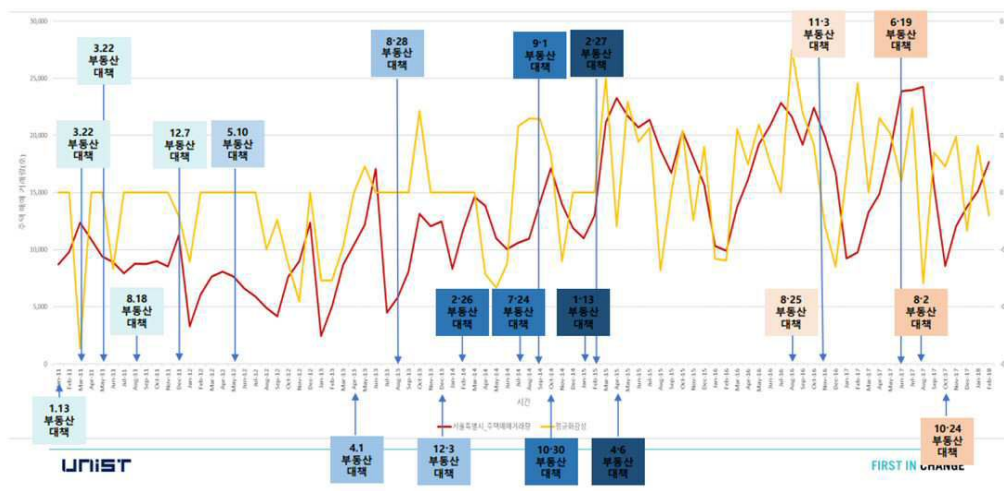
도면7



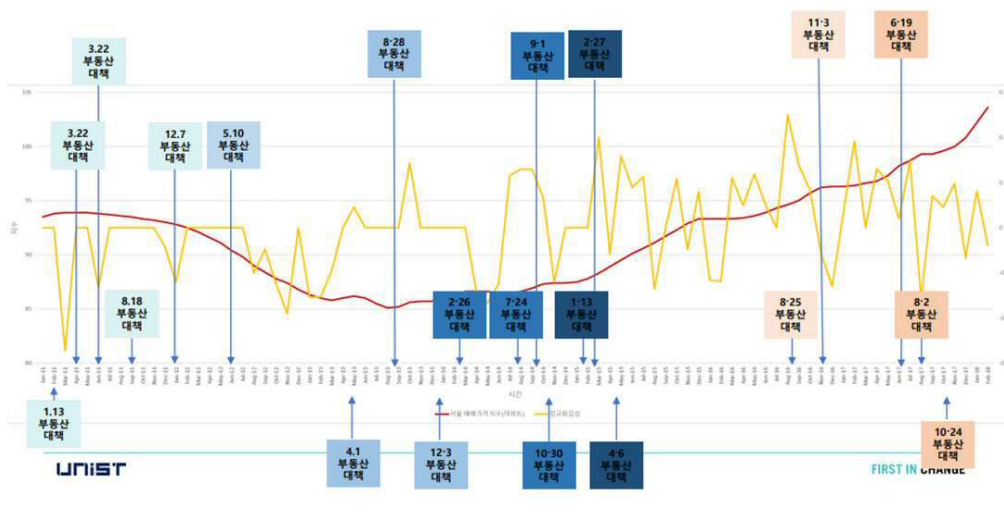
도면8



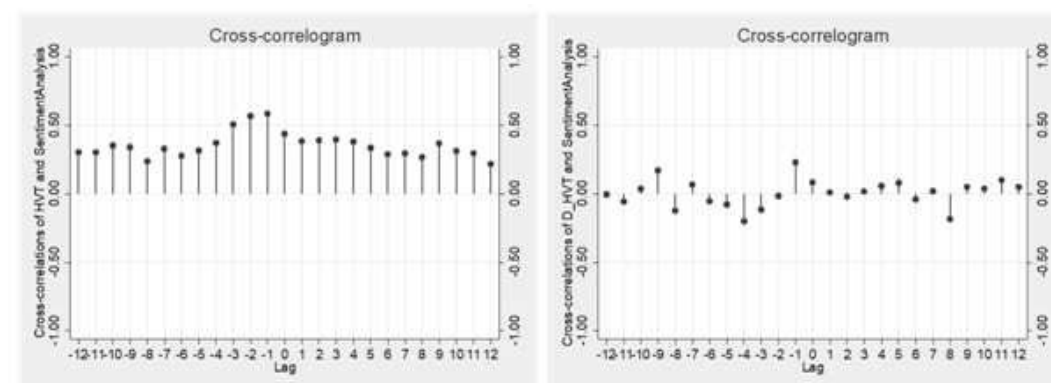
도면9a



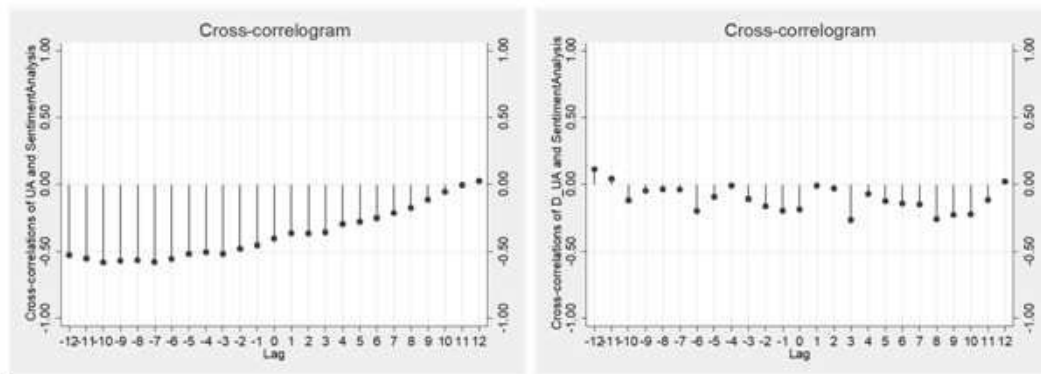
도면9b



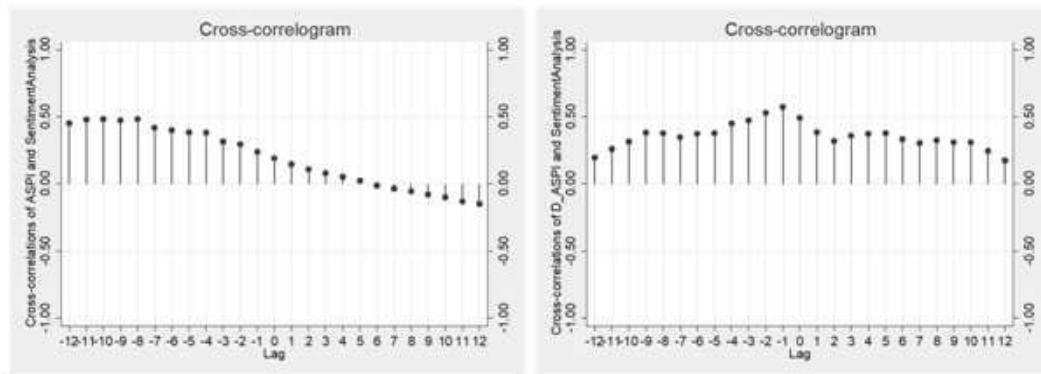
도면10



도면11



도면12



도면13

