



빅데이터에 의한 부동산 활동 의사결정 모형에 관한 연구

경정익(Ph.D)



발표 순서

- ▣ 연구 배경 / 목적
- ▣ 이론 및 선행 연구 고찰
- ▣ 빅데이터에 의한 부동산 의사결정 모형
- ▣ 실증 분석
- ▣ 연구 결과



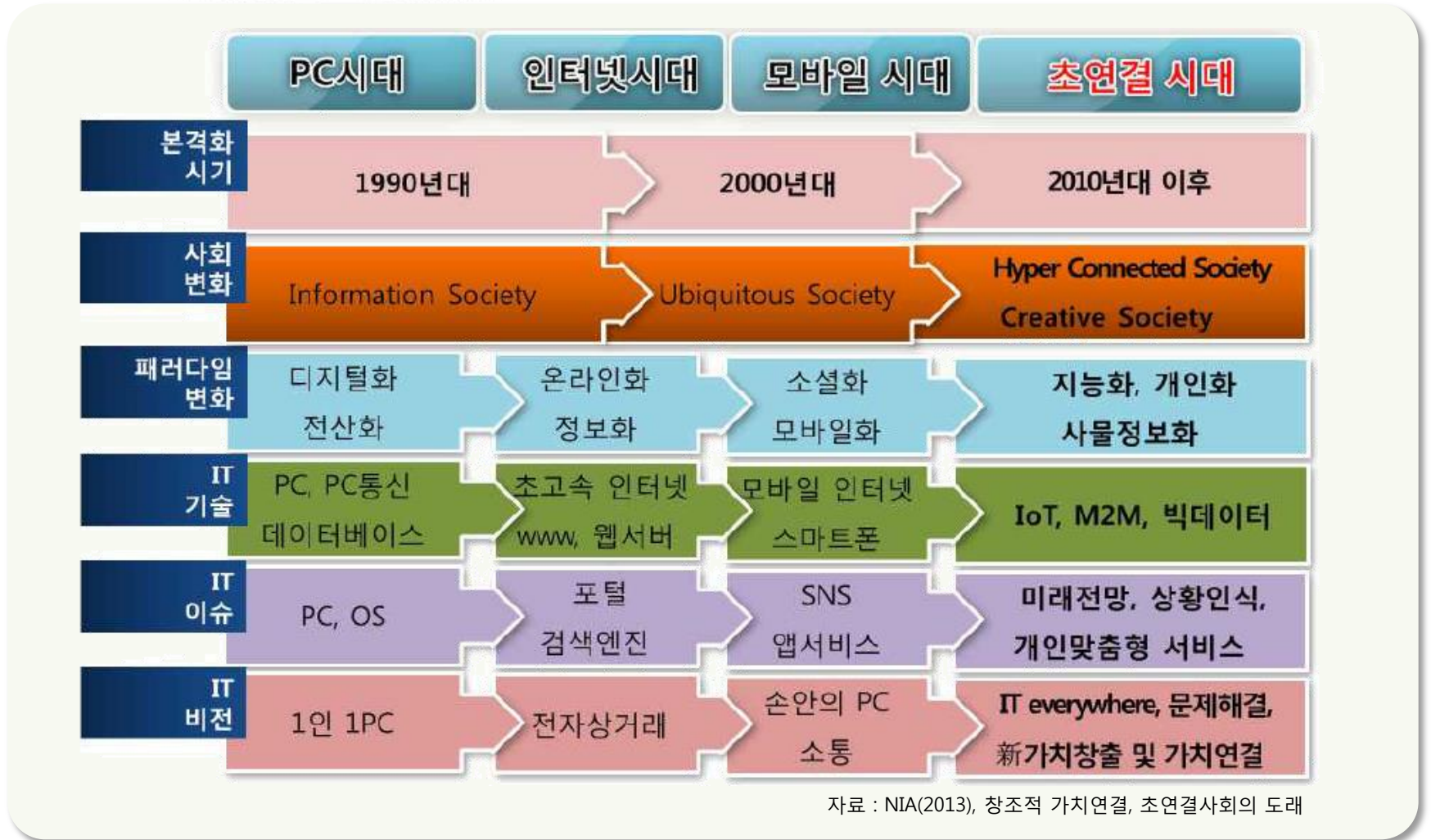
연구 배경 / 목적

연구 배경 및 필요성

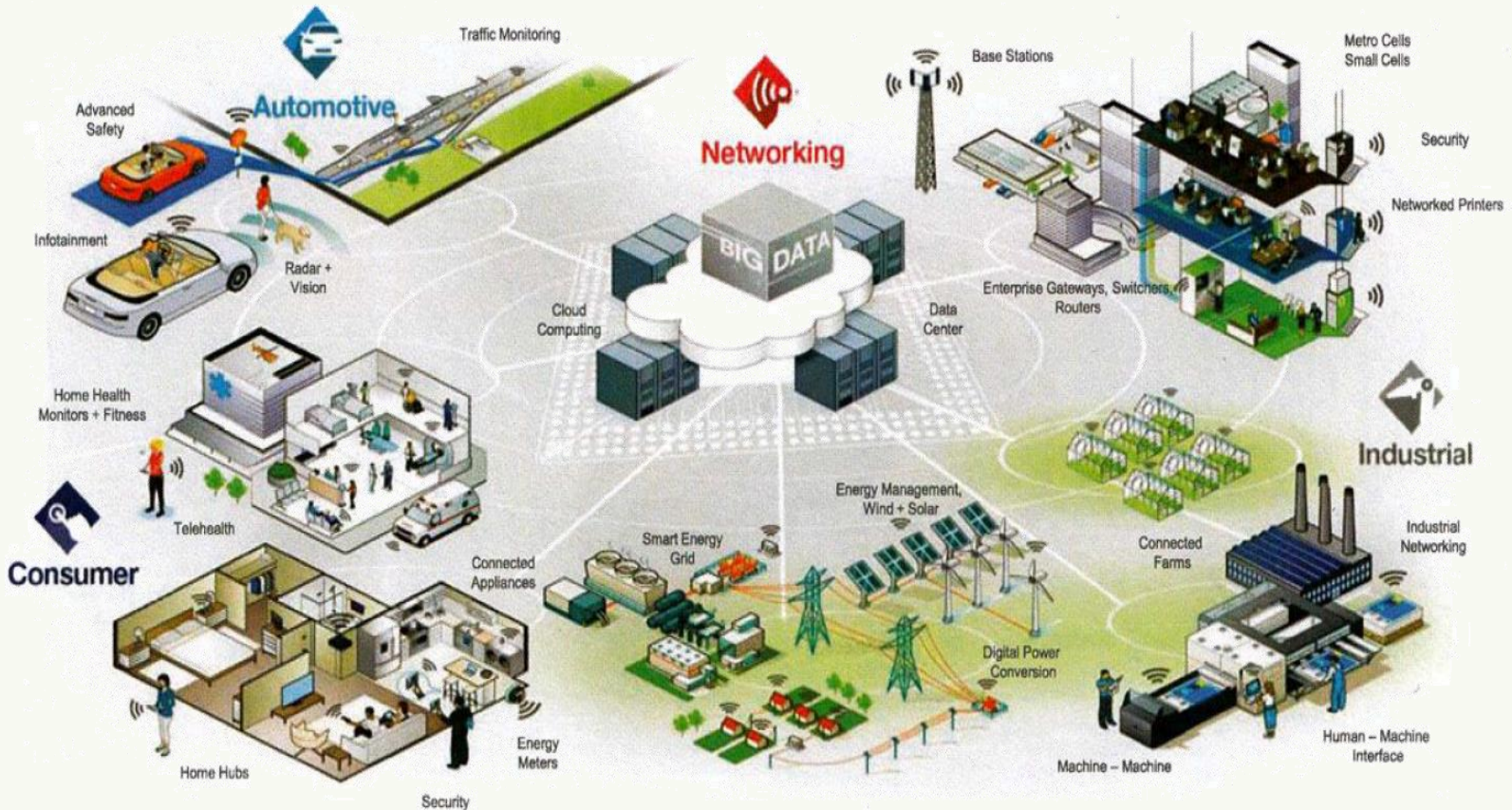
- ▶ 현대사회는 Internet과 Globalization으로 인해 불확실성과 리스크가 상상 이상으로 증가. 실제 세계는 금융위기 속에 경제 침체가 본격화되고 계층간 갈등, 고용불안, 양극화 심화 등 사회불안이 가중되고 난제 발생 현상.
- N. N. Taleb 「The Black Swan, 2007)」
- ▶ 기술은 발전할수록 수확가속법칙에 따라 가속도가 붙어 2045년에는 특이점 (Singularity) 통과, 향후 사회 및 기술의 폭발적인 발전 예측.
- Ray Kurzweil 「The Singularity is Near, 2005)」
- ▶ “미래 예측이란 미래를 예측하는 것이 아니라 현재의 미래성을 이해하고 현재를 분명히 認知하는 것이 핵심”
- Peter F. Drucker, 「The Next Society, 2002」
- 현재 드러나는 또는 드러나지 않은 현상을 설명하는 변인을 파악, 의미있는 가설을 통해 인과관계 규명이 중요, 그 출발은 “데이터”

데이터기반의 예측과 선제적 대응을 위한 실시간 의사결정이 중요

연구 배경 및 필요성



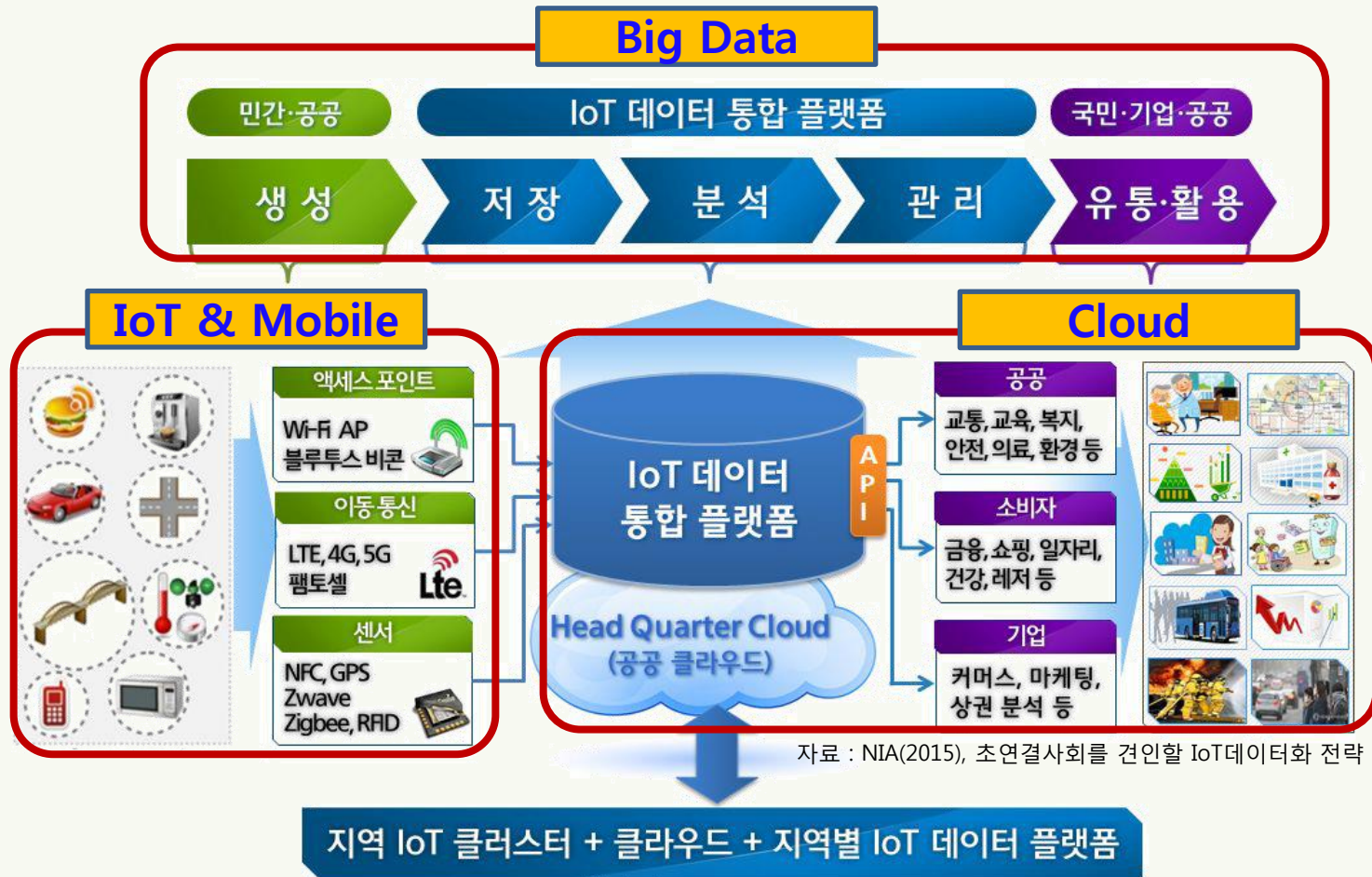
Hyper Connectivity Society 도래



자료 : freescale(<http://www.freescale.com/ja/webapp/sps/site/homepage.jsp?code=IOT-INTERNET-OF-THINGS&cid=ps39vr7h9j>)

데이터의 폭발적 증가 : 1.4Z(5000) ⇒ 1.8Z(2011) ⇒ 400Z(2018)

IT 핵심기술 발전(ICBM)



연구 수행 목적 / 절차



연구 목적

- ▶ 웹 클로링(Web Crawling)에 의해 수집된 부동산 관련 뉴스기사의 감성분석을 통해 부동산 활동의 의사결정 모형 제시
- ▶ 실거래가격, 주택거래량 등 정형데이터와 비교를 통해 모델의 적절성을 실증 검증하여 향후 모형의 활용 방안을 제시

빅데이터에 의한 부동산 활동 의사결정 모형 개발

관련이론/선행 연구고찰

기초데이터 수집

텍스트마이닝
오피니언 마이닝

모델제시/
실증분석

부동산 의사결정 모형제시

빅데이터 연구동향

부동산과 심리,
뉴스 관련

부동산 빅데이터
활용/추진

문헌/학술DB

주택관련 뉴스
Crawling, Parsing

주택 관련
데이터 수집

포털(네이버)
R-ONE시스템

형태소/극성
판단 단어 도출

긍정/부정 단어
극성 태깅

스코어를 설정
/극성판단

감성사전 개발
Komoran

부동산 의사결정
모형 제시

뉴스와 주택가격
변동의 패턴 검토

뉴스와 주택가격
등락 비교

SPSS(PASW)

주택가격 변동의
뉴스 변수 도출

모델 활용방안
모색

지능형 모델 개발



이론 및 선행 연구 고찰

부동산시장 vs. 심리적 태도



- ▶ 주택가격과 거래량의 양(+)의 관계와 같은 '이례현상(anomaly)'은 기존 경제이론으로 설명 곤란한 발생
 - 인간의 심리적 특성을 경제이론에 접목한 행동경제학(Behavioral Economics) 으로 설명
 - 전망이론(Prospect Theory)의 '손실회피성'이라는 인간의 심리적 편의(Bias)로 설명
- ▶ 시장 참여 주체들의 심리적 태도가 주택시장에 영향을 미침을 밝힘(Adams, 1964)
- ▶ 신규주택시장에서 ICS(Index of Consumer Sentiment)와 HIS(Index of Housing Sentiment)가 중요한 예측 자료로 활용 가능, 유용성 검증(Weber and Devancy, 1996)
- ▶ 소비자의 심리적 태도와 주택가격관 높은 관련성이 있음을 검증(Johnson, 2010)
- ▶ 인간의 심리적 태도가 부동산시장에 중요한 요소로 작용(김대원·유정석, 2013)
- ▶ 주택시장에서 시장 참여자의 직관과 투자기대심리도 어느 정도 영향을 미침(진창하, 2012)
- ▶ 투자심리가 전반적인 주택가격 변동에 영향을 미침을 실증(정의철, 2010; 이상준 외, 2013)

부동산시장 vs. 뉴스 기사



- ▶ 뉴스는 대중이 관심을 가지거나 알려야 하는 정보의 정제된 표현으로 인터넷 기사는 인터넷 상 신뢰할 수 있고 중요한 정보중 하나(Dimmick et al., 2004)
- ▶ 소비자는 직접적인 관찰이나 경험 또는 비공식적인 사회적 커뮤니케이션으로 경제상황을 판단하게 됨(Weatherford, 1983; MacKuen, Erikson & Stimson, 1992)
- ▶ 경제뉴스는 잠재적으로 소비자 심리의 장기적인 경제상황에 널리 영향을 미침
(Blood & Philips, 1997; 심재철, 2005)
- ▶ 소비주체는 상당부분 미디어가 보도하는 사실에 기초하여 경제현실을 평가하며, 신문이나 방송에 등장한 전문가의 의견을 통해 경제상황이나 정치적 환경을 이해
(Nedeau, Niemi, Fan & Amato, 1999; Sanders et al., 1993; Mutz, 1992; Hetherington, 1996)
- ▶ 미디어의 보도가 소비자의 인식에 영향을 주고 또 소비자의 인식이 보도에 영향을 주어 상호 영향관계가 순환적으로 움직임(Stevenson, Gonzenbach & David, 1994), 특히 부정적인 내용의 경제기사가 경제현실에 대한 소비자의 인식에 더욱 영향을 미친다고 함(Goidel & Langley, 1995)
- ▶ 국내에서는 부정적인 경제기사 논조가 미래 경제상황에 대한 소비자 인식에 영향을 미친다는 사실을 밝힘(이완수, 2007)

빅데이터 기술



Gartner 제시 10대 전략기술(2011-2016)

2011년	2012년	2013년	2014년	2015년	2016년
클라우드컴퓨팅	미디어태블릿 그 이후	모바일대전	다양한 모바일 기기관리	컴퓨팅 에브리웨어	디바이스에서
모바일앱과 미디어태블릿	모바일중심 애플리케이션과 인터페이스	모바일 앱 & HTML5	모바일앱과 애플리케이션	사물인터넷	애플리케이션 사용자경험
소셜커뮤니케이션 및 협업	상황인식과 소셜이 결합된 사용자경험	퍼스널 클라우드	인물인터넷	3D프린팅	3D프린팅재료
비디오	사물인터넷	사물인터넷	하이브리드 클라우드와 서비스브로커로서의 IT	보편화된 첨단분석	인물정보
차세대 분석	앱스토어와 마켓플레이스	하이브리드 IT & 클라우드 컴퓨팅	클라우드/클라이언트 아키텍처	콘텐츠 리치시스템	진보한 기계학습
소셜 분석	차세대 분석	전략적 빅데이터	퍼스널 클라우드의 시대	스마트머신	지능형기기
상황인식 컴퓨팅	빅데이터	실용분석	소프트웨어 정의	클라우드/클라이언트 컴퓨팅	상황에 따라 적응하는 보안구조
스토리지급 메모리	인메모리 컴퓨팅	인메모리 컴퓨팅	웹스케일IT	소프트웨어정의 애플리케이션과 인프라	진보된 시스템 아키텍처
유비쿼터스 컴퓨팅	저전력서버	통합생태계	스마트머신	웹스케일IT	메시앱과 서비스 아키텍처
패브릭기반 컴퓨팅 및 인프라스트럭처	클라우드 컴퓨팅	엔터프라이즈 앱스토어	3D프린팅	위험 기반 보안과 자가 방어	사물인터넷 플랫폼

■ 클라우드 ■ 모바일과 사물인터넷 ■ 데이터분석

Gartner(2011-2016) 재구성

사물인터넷(IoT) 환경으로 인해 데이터의 폭발적인 증가로 빅데이터에 의한 활용의 중요성 강조

빅데이터 기술

빅데이터 분석 기술

분석기법	주요 내용
텍스트 마이닝	<ul style="list-style-type: none"> · 비정형 또는 반정형 텍스트 데이터에서 자연어 처리 기술에 기반해 유용한 정보를 추출, 가공하는 기법 · 대용량 언어자원과 통계적, 규칙적 알고리즘 사용 · 문서분류, 문서 군집, 정보추출, 문서요약 등에 응용
오피니언 마이닝	<ul style="list-style-type: none"> · 텍스트에 나타난 감정, 뉴앙스, 저자의 태도 등을 판별하여 의미있는 정보로 변환하고 이를 의사결정에 활용하고자 하는 기술
평판분석	<ul style="list-style-type: none"> · 웹사이트와 소셜 미디어에 나타난 여론과 의견(긍정 및 부정, 중립 등)을 분석하여 실질적으로 유용한 정보로 재가공하는 기술 · 특정 서비스 및 상품에 대한 반응 및 입소문 분석 등에 활용
소셜분석	<ul style="list-style-type: none"> · 일명 소셜네트워크분석(SNA)으로 소셜네트워크의 연결구조 및 강도 등을 기반으로 소셜 네트워크상의 위치를 규명 · 소셜 미디어의 글과 사용자를 분석해 소비자 흐름이나 패턴 등을 분석하고 판매나 홍보에 적용
클러스터분석	<ul style="list-style-type: none"> · 통계기법에 의해 비슷한 특성을 가진 개체를 클러스터로 나누는 방법을 통해 유사성을 판단하는 기술로서 일명 군집분석이라 함
현실마이닝	<ul style="list-style-type: none"> · 사람들의 일상 또는 특정한 시기 행동패턴 예측을 위한 분석 · 사회적 행동과 관련된 정보를 휴대폰이나 GPS등의 기기를 통해 수집 · 현실에서 발생하는 정보를 기반으로 인간관계와 정보추구 또는 이용행태 등을 추론

이정미(2013)에서 발췌, 재구성

빅데이터 기술

Increased Deployment, Decreased Hesitation

Since 2014, the number of organizations that have already deployed/implemented data-driven projects has **INCREASED BY 125%**.

42% still planning data implementations



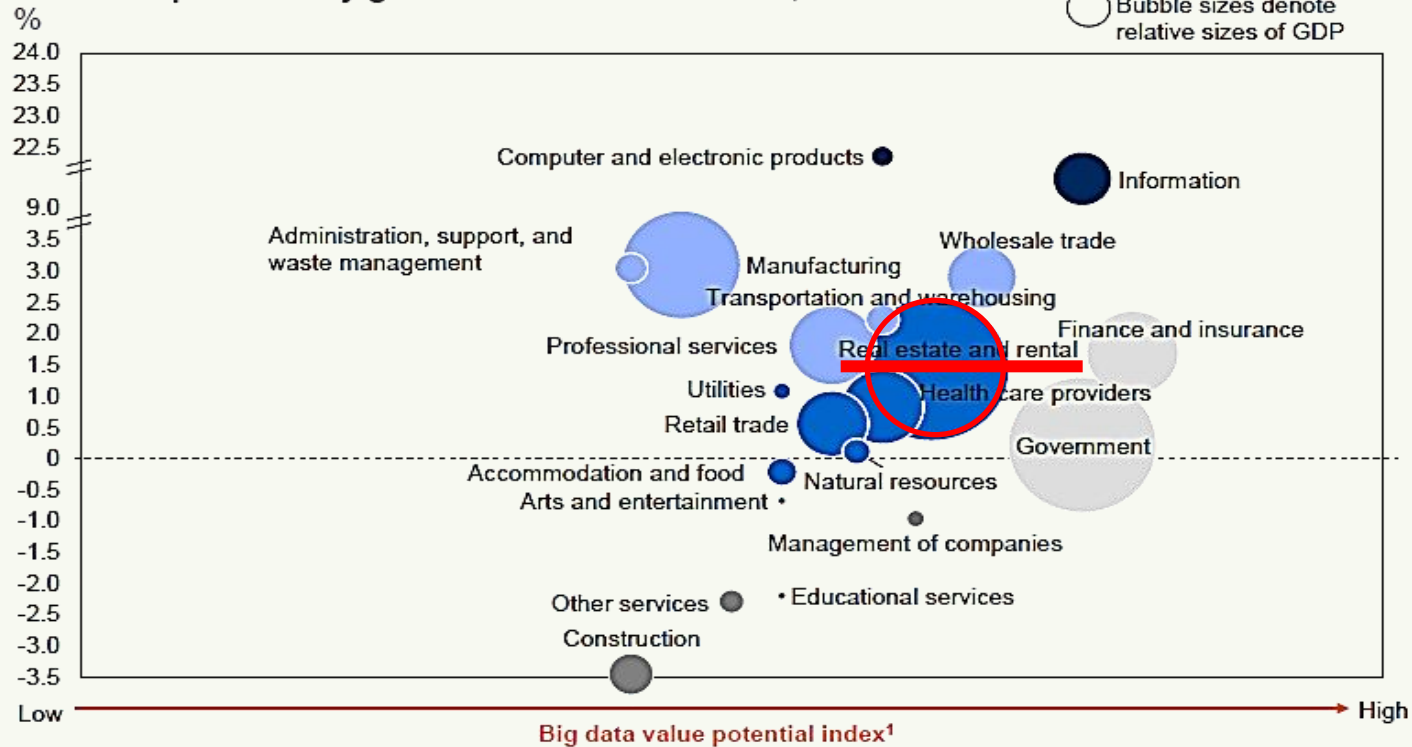
Q. Is your company currently implementing, planning or considering data-driven projects (i.e. devising strategies and projects to generate more value from existing data)?

IDG(2015), Enterprise Big Data and Analytics Sueuey

빅데이터와 부동산

Some sectors are positioned for greater gains from the use of big data

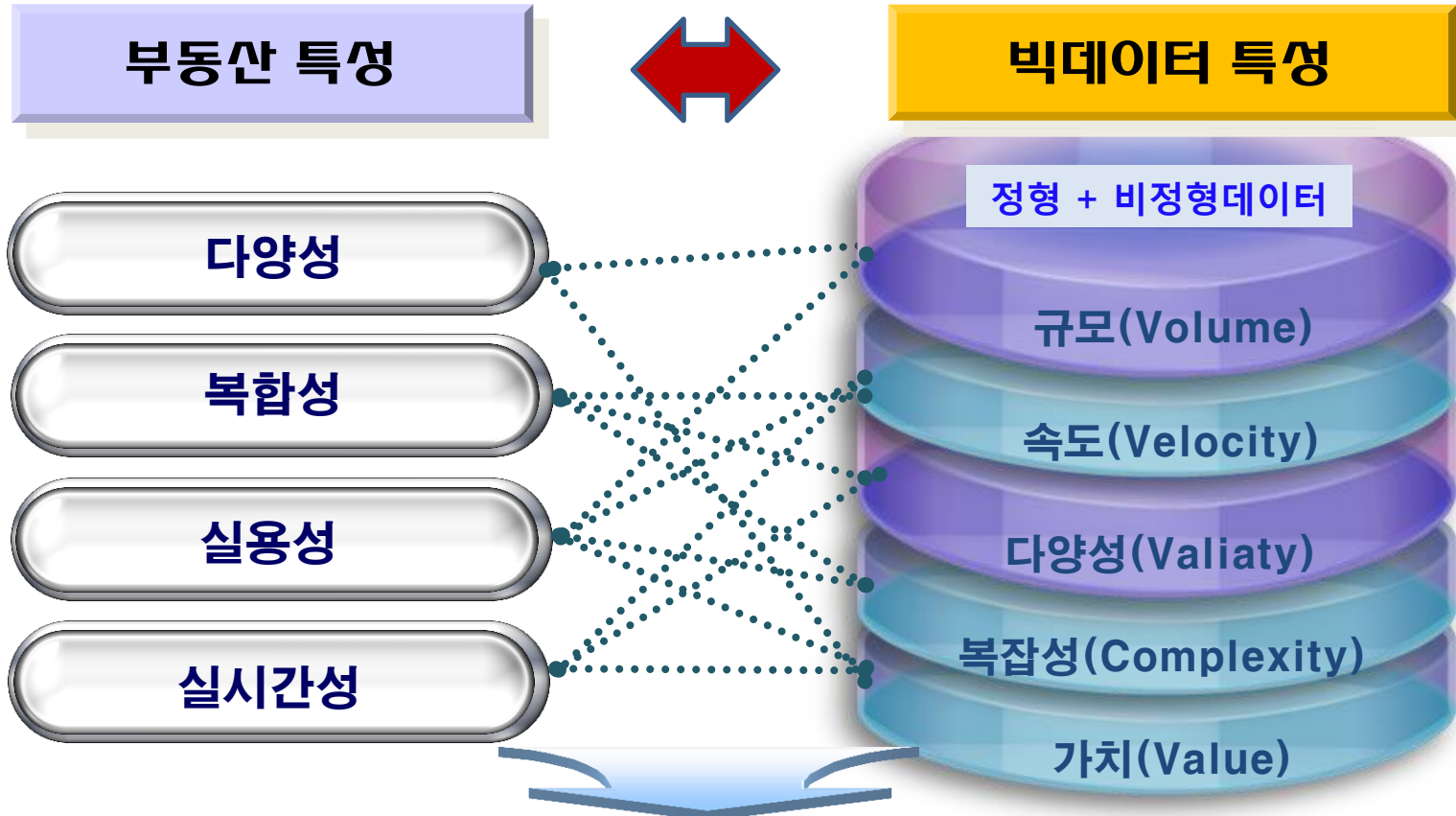
Historical productivity growth in the United States, 2000–08



1 See appendix for detailed definitions and metrics used for value potential index.
SOURCE: US Bureau of Labor Statistics; McKinsey Global Institute analysis

자료 : Mckinsey(2012)

빅데이터와 부동산



맞춤형 부동산정보 도출로 예측가능한 부동산활동 가능
(정책, 매각/매입, 교환, 이용/개발, 중개, 컨설팅 등)

자료 : 경정익(2015), 부동산 정보화의 이해, 부연사

빅데이터와 부동산



분 야	부동산 빅데이터 활용 트렌드
S(사회)	<ul style="list-style-type: none"> · 빅데이터 관심 고조 및 활용 인식 확산(정부/공공기관, 기업, 국민) · 정보화(ICBM)발달로 대량의 데이터 생산 및 유통 · 빅데이터가 IT분야의 새로운 패러다임이자 신 성장동력으로 급부상 · 데이터를 기반한 예측가능한 부동산정책의 선제적 대응 요구 증대
T(기술)	<ul style="list-style-type: none"> · 지식과 기술의 융·복합으로 새로운 기회 등장 · 국가 및 사회발전을 위한 ICBM 분야 중점 추진 · 수요자 중심의 정보 서비스 제공 필요성 증가 · 시스템 연계·통합으로 개인화된 맞춤형 부동산정보 제공 요구 증가
E(경제)	<ul style="list-style-type: none"> · 빅데이터는 경영혁신과 현안해결 등 경제·사회 발전의 원동력 · 주요국 및 글로벌 기업의 빅데이터 사업 육성 및 활용 주력 · 빅데이터를 통한 신규 비즈니스 모델 개발 활발 · 공공 및 민간 부동산 데이터 통합을 통한 부동산산업 확장 및 일자리 창출
E(환경)	<ul style="list-style-type: none"> · 공공데이터 개방 확대 및 민간데이터 융합에 의한 인사이트 창출 · 선진국 수준의 국민 신뢰도 제고 요구 증가 · 데이터 공개·공유 확대에 따른 개인정보 보호 제한 및 정보이용 역작용 · 부동산분야는 빅데이터 활용에 최적의 여건(데이터 확보, 특성 고려)
P(정책)	<ul style="list-style-type: none"> · 새로운 패러다임에 맞는 범정부 혁신 요구 증대 · 다양한 분야의 데이터 융·복합을 통한 창조경제 실현 · 빅데이터는 국가경쟁력 확보 및 미래를 대비하는 성장 동력이라는 공감대 형성 · 정부 3.0구현으로 빅데이터를 활용한 과학적 부동산 정책 추진

자료 : 경정익(2013), 빅데이터 부동산분야 활용방안 연구

빅데이터와 부동산

해외 주요국 빅데이터 활용 추진

구 분		주 요 내 용
공공 부 문	미국	주택도시개발부(Department of Housing and Urban Development)은 보고서, 자료, 부동산통계 등을 API 방식에 의해 테이블(엑셀), 지도, 인포그래픽스 등 다양한 형태의 자료를 HUD USER 포털사이트를 개설하여 제공
	영국	지방자치부(DCLG: Department of Communities and Local Government)는 Interform 사이트를 통해 부동산의 각종 자료 및 통계를 취합하여 제공
	캐나다	온타리오주의 MPAC(Municipal Property Assessment Corporation)은 빅데이터를 통해 부동산의 가치를 감정하여 재산세를 부과, 매년 감정시가를 통보하여 세금 부과 금액을 사전 공지
	일본	국토교통성은 전국적으로 운영중에 있는 부동산거래정보망인 REINS에 공공정보와 민간정보를 취합한 빅데이터를 연계한 '부동산 정보센터(Real Estate Information Center)' 구상 추진
민 간 부 문	SmartZip (미국)	부동산 매물의 2,000개의 속성정보를 데이터마이닝에 의한 빅데이터 분석으로 향후 6-12개월 이내 매물이 예상되는 대상을 제공하여 주는 매물 예상 분석서비스를 제공
	Zillow사 (미국)	미 전역의 47백만 가구 이상의 정보를 축적하여 주택소유자나 부동산 전문가가 'Zestimates'라는 툴을 활용하여 입력한 부동산 정보, GIS 위치정보, 인구 및 통계정보, 학군 정보 등 부동산 매매에 필요한 정보를 통합하여 주택가격 지수를 산정하여 활용
	Teranet사 (캐나다)	Geowarehouse' 사이트를 통해 토지정보, 가격, 소유주 등 다양하고 상세한 정보와 각 지역의 가구수, 평균소득 등 지역데이터와 물건에 대해 인공위성으로 찍은 3D사진, 그리고 다양한 각도에서 찍은 사진정보 등 방대한 부동산 등기정보를 데이터웨어 하우스 (www.geowarehouse.ca)를 구축하여 서비스

자료 : 김정익(2015), 빅데이터 부동산분야 활용방안 연구

빅데이터와 부동산

국내 빅데이터 활용 연구/추진

구분	추진 및 연구	비고
국토교통부	<ul style="list-style-type: none"> · 공간 빅데이터 구축 및 활용방안'에 대한 연구 · 공간빅데이터 체계구축 사업 추진1* 	2013 2014
국토연구원	<ul style="list-style-type: none"> · 국토정책 선진화를 위한 빅데이터 활용에 관한 기초연구 	2013
국토교통과학기술진흥원	<ul style="list-style-type: none"> · 빅데이터를 활용한 부동산시장 분석 및 예측 모형 개발 기획 	2014
대한지적공사	<ul style="list-style-type: none"> · LX 빅데이터 추진전략 수립 · LX 빅데이터 기반 구축 	2014

* 부동산 투기 감지 및 실수요 거래를 분석하여 수요 기반의 부동산 정책 지원, 공공 및 민간정보 융합 분석을 통한 정책 의사결정을 지원

경정익(2013), "부동산, 빅데이터로 효율성 높여라", 디지털타임즈, 2013. 8. 19.

경정익(2014), "부동산분야의 빅데이터 활용 방안과 정책적 제언", 「부동산경영」, 한국부동산경영학회, 제10호

경정익(2015), 「스마트 빅데이터 시대 부동산 정보화의 이해」, 두남출판사

경정익(2015), "부동산정책 빅데이터로 신뢰 높여라", 디지털타임즈, 2015. 5. 26

경정익 외(2015), "부동산분야에서 빅데이터 전략적 활용의 영향요인에 관한 연구", 「부동산산업연구」, 제1권 1호

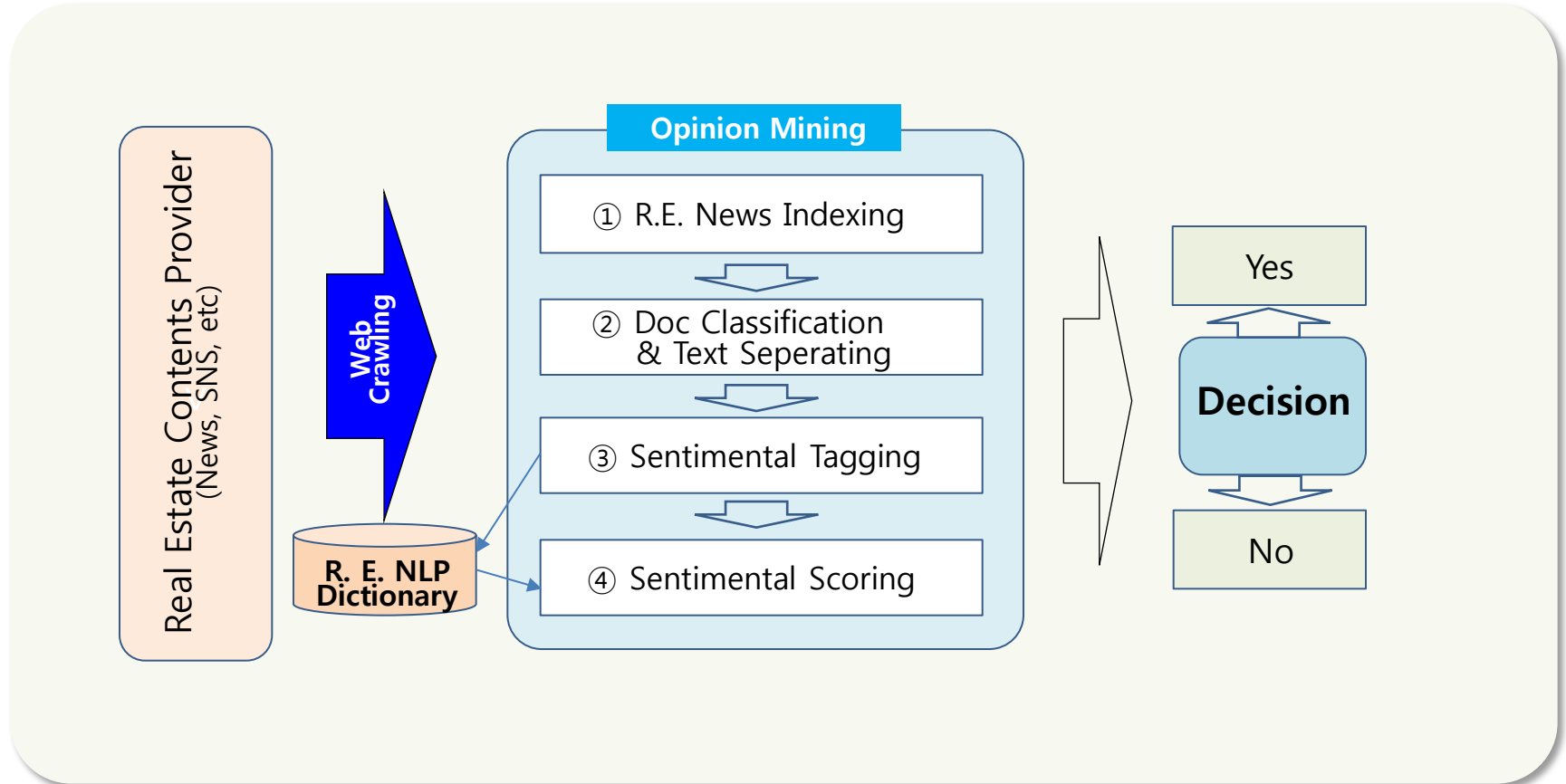
경정익 외(2015), "부동산분야의 빅데이터 도입의도에 미치는 영향요인에 관한 연구", 「부동산분석」, 제1권 1호

김대중(2013), "국토정책 선진화를 위한 빅데이터 활용에 관한 기초연구", 국토연구원,



빅데이터에 의한 부동산 의사결정 모형

부동산 빅데이터 의사결정 모형 구조



부동산 빅데이터 감성분석



1) 부동산 뉴스별 긍정/부정값(NwsDn)을 계산

$$NwsDn = \sum_{i=1}^n Sen(i) \quad \text{<식 1>}$$

- ① 개별 뉴스에는 여러 문장(Sen)이 존재하며 이는 긍정과 부정의 극성을 가짐
- ② 수집된 뉴스와 관련된 부동산 도메인의 감성사전을 개발
- ③ 각 문장의 속성과 표현을 기준으로 긍정과 부정 극성을 감성사전과 비교하여 태깅.
D(n)=문장별 긍정값 (긍정 : 1, 부정 : -1)
- ④ 각 문장의 긍정/부정값을 뉴스의 극성값으로 합산

2) 일일 발생한 모든 뉴스의 긍정과 부정 극성값을 합산하여 일일 감성점수 산출

$$DayNws = (NwsDn1 + NwsDn2 + NwsDn3 + NwsDn4 + NwsDn5, \dots + NwsDni)$$

$$DayNws = \sum_{j=1}^n NwsDn(j) \quad \text{<식 2>}$$

3) 분석하고자 하는 기간에 따라 감성점수 합산(월단위 또는 10일단위)

- ① 월단위
- ② 10일간격(1~10일)

$$MonNws = \sum_{j=1}^{31} DayNws(n), \quad DtnNws = \sum_{n=1}^{10} DayNws(n) \quad \text{<식 3>}$$



실 증 분 석

데이터 산출 및 수집

부동산 뉴스 데이터(비정형데이터)

구 분	내 용	비 고
데이터 수집원	네이버>부동산>뉴스>분야별 뉴스	
수 집 기 간	2013. 1. 1 ~ 2015. 5. 31 (978일)	
수집 뉴스량	35,082건	
수 집 방 법	Web Crawling	Script 개발

감성분석 결과 데이터

구 분	수집 기사	감성사전 반응 기사				감성사전 무반응 기사
		계	긍정 반응	부정 반응	긍정+부정 반응	
기 사 수	35,082	16,236	8,531	3,033	4,672	18,846
비 고	100 %	46.3%	24.4 %	8.6 %	13.3 %	53.7 %

데이터 산출 및 수집

부동산 정형데이터

데이터 명	공표기관	형태	데이터 출처	비 고
월별 아파트 실거래가	국토교통부	엑셀	실거래가 공개시스템	10일 단위
아파트 실거래가격지수	한국감정원	엑셀	R-ONE 부동산통계정보시스템	월 단위
주택시장 소비심리지수	국토연구원	엑셀	R-ONE 부동산통계정보시스템	월 단위
월별 아파트 거래현황	한국감정원	엑셀	R-ONE 부동산통계정보시스템	월 단위
월간 KS 주택가격 동향	국민은행	PDF	국민은행>KB부동산>통계	월 단위

아파트 실거래가격 데이터

구 분	단지수	세대수	비 고
계	75개 단지	95,980	
서울 지역	31개 단지	48,841	강남개포, 송파신천, 강동상일, 강서등촌, 양천목동, 노원중개
수도권 지역	44개 단지	47,139	광명하안/철산, 남양주별내, 일산동구, 성남정자, 수원영통

감성분석 데이터 타당성

상관분석 결과

Senti_M	Pearson 상관계수 유의확률 (양쪽) N	1 .000 32	.618** .000 32	.439** .012 32	.329 .066 32	.434** .013 32	.671** .000 32	.678** .000 32	.673** .000 32	.612** .000 32	.453** .009 32	.508** .003 32	.515** .003 32	.637** .000 32	.600** .000 32	.870** .000 32	.869** .000 32	.871** .000 32
Senti_10_N	Pearson 상관계수 유의확률 (양쪽) N	.618** .000 32	1 .000 96	.597** .000 96	.571** .000 96	.612** .000 96	.544** .001 32	.523** .002 32	.527** .002 32	.456** .007 34	.273 .118 34	.339** .050 34	.354** .047 32	.399** .024 32	.377** .033 32	.606** .000 32	.628** .000 32	.629** .000 32
실거래가_서울_10	Pearson 상관계수 유의확률 (양쪽) N	.439** .012 32	.597** .000 96	1 .000 96	.390** .000 96	.395** .000 96	.612** .000 32	.535** .000 32	.607** .000 32	.624** .000 34	.430** .010 34	.489** .003 32	.590** .001 32	.547** .001 32	.517** .002 32	.382** .031 32	.388** .028 32	.399** .024 32
실거래가_경기_10	Pearson 상관계수 N	.329 .066 32	.390** .000 96	.395** .000 96	1 .000 905	.458** .000 458	.433** .000 433	.450** .000 450	.480** .000 480	.289 .130 338	.458** .001 445	.445** .001 390	.390** .002 355	.255 .031 244	.255 .031 244	.270** .002 270	.270** .002 270	.270** .002 270

구 분	실거래 가격 (서울)	실거래 가격 (경기)	실거래 가격지수 (서울·경기)	실거래 가격지수 (서울)	실거래 가격지수 (경기)
감성분석 결과	.571**	.612**	.597**	.523**	.612**
구 분	아파트 거래량 (서울)	아파트 거래량 (경기)	주택심리지수 (전국)	주택심리지수 (서울)	주택심리지수 (경기)
감성분석 결과	.399*	.377*	.606**	.628**	.629**

** 상관계수는 0.01수준에서 유의. * 상관계수는 0.05수준에서 유의.

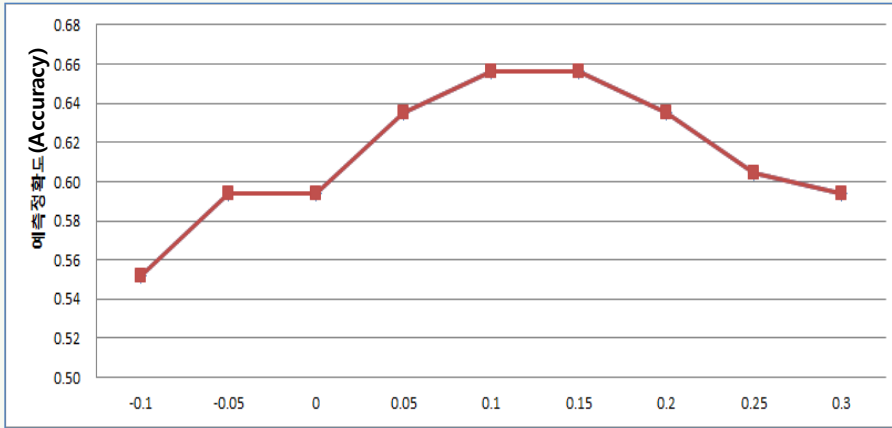
유의확률 (양쪽) N	.000 32	.033 32	.002 32	.027 32	.003 32	.000 32	.000 32	.000 32	.000 32	.007 32	.002 32	.000 32	.000 32	.001 32	.001 32	.001 32	.001 32	
소비심리지수_전국	.870** .000 32	.606** .000 32	.382** .031 32	.255 .159 32	.449** .010 32	.676** .000 32	.685** .000 32	.678** .000 32	.614** .000 32	.450** .010 32	.506** .003 32	.445** .011 32	.614** .000 32	.578** .001 32	.578** .001 32	.976** .000 32	.976** .000 32	.992** .000 32
소비심리지수_서울	.869** .000 32	.628** .000 32	.388** .028 32	.244 .179 32	.492** .004 32	.734** .000 32	.744** .000 32	.738** .000 32	.695** .000 32	.577** .001 32	.626** .000 32	.420** .017 32	.584** .000 32	.553** .001 32	.976** .000 32	.976** .000 32	.985** .000 32	.985** .000 32
소비심리지수_경기	.871** .000 32	.629** .000 32	.399** .024 32	.270 .134 32	.461** .008 32	.728** .000 32	.732** .000 32	.727** .000 32	.667** .000 32	.489** .004 32	.553** .001 32	.453** .009 32	.616** .000 32	.578** .001 32	.992** .000 32	.985** .000 32	.985** .000 32	.985** .000 32

** 상관계수는 0.01 수준(양쪽)에서 유의합니다.

* 상관계수는 0.05 수준(양쪽)에서 유의합니다.

감성분석에 의한 예측 타당성

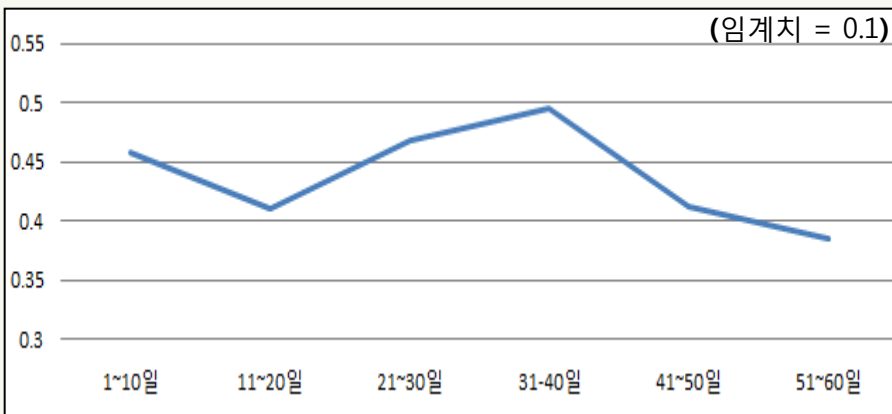
감성분석 데이터 임계치 산출



임계치	대상	등락일치	정확도
-0.1	96	53	0.55
-0.05	96	57	0.59
0	96	57	0.59
0.05	96	61	0.64
0.1	96	63	0.66
0.15	96	63	0.66
0.2	96	61	0.64
0.25	96	58	0.60
0.3	96	57	0.59

감성분석 예측정확도 (Accuracy)

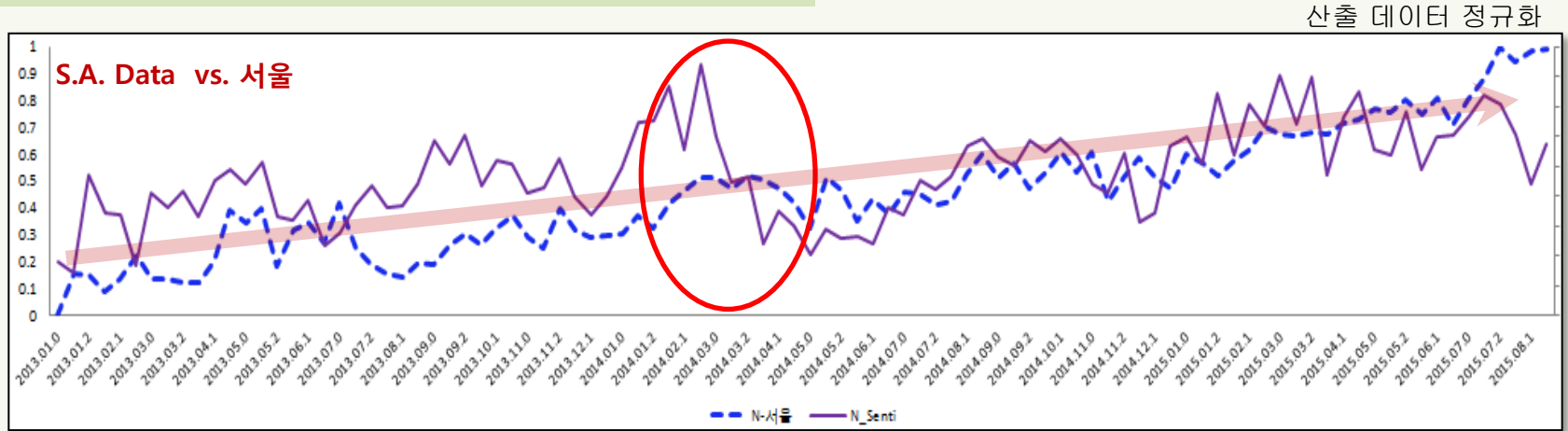
$$Accuracy = \frac{True\ Pos + True\ Nnn}{True\ Pos + True\ Nnn + False\ Pos + False\ Nnn}$$



구분	정확도	등락일치
1~10일 후	0.46	44
11~20일 후	0.41	39
21~30일 후	0.47	44
31~40일 후	0.49	46
41~50일 후	0.41	38
51~60일 후	0.38	35

실증 분석 결과

감성분석 데이터 vs 실거래가격 추세(1)



구분	단지수	세대수	비고
계	75개 단지	95,980	
서울 지역	31개 단지	48,841	강남개포, 송파신천, 강동상일, 강서등촌, 양천목동, 노원중개
수도권 지역	44개 단지	47,139	광명하안/철산, 남양주별내, 일산동구, 성남정자, 수원영통

1) 거래된 해당 아파트 m당 실거래 가격

$$\text{AptPM} = \frac{\text{AptP}}{\text{전용면적 (m}^2\text{)}}$$

2) 기간별 거래된 동일 아파트 단지별 평균 실거래 가격

$$\text{AptGr} = \frac{\sum_{i=1}^n \text{AptPM}_i}{n}$$

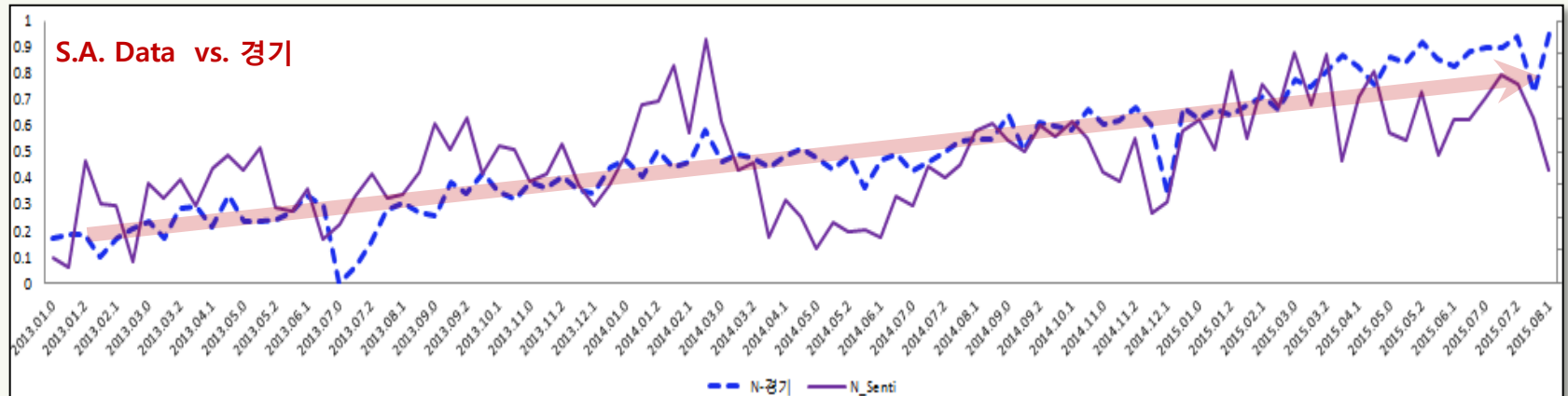
3) 기간중 거래된 전체 아파트 평균가격 산출

$$\text{AptGPr}(1-10) = \frac{\sum_{i=1}^n \text{AptGri}}{n} \quad \text{AptGPr}(11-20) = \frac{\sum_{i=1}^n \text{AptGri}}{n}$$

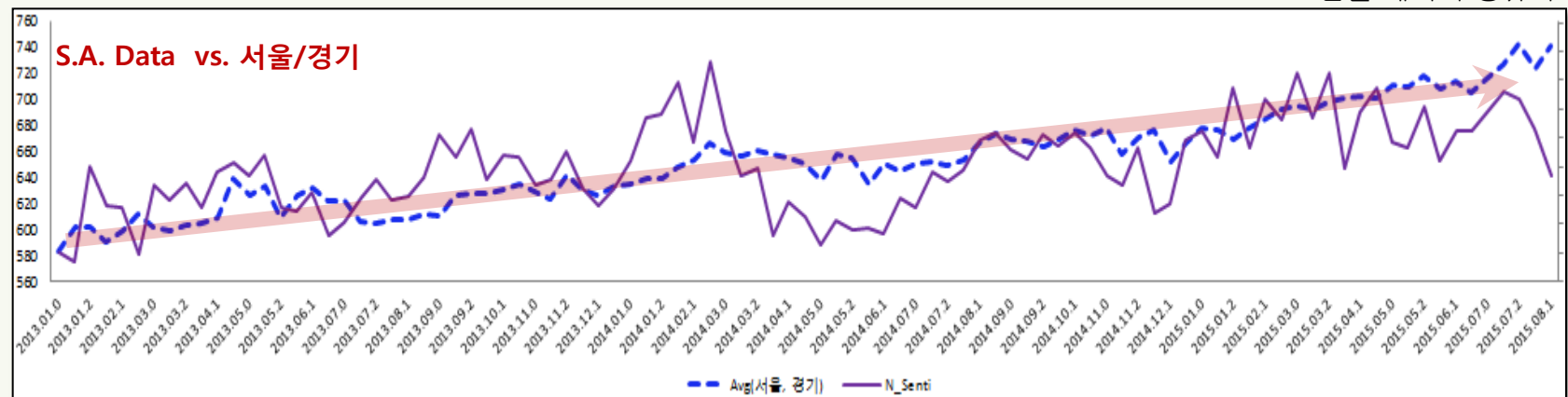
실증 분석 결과

감성분석 데이터 vs 실거래가격 추세(2)

산출 데이터 정규화

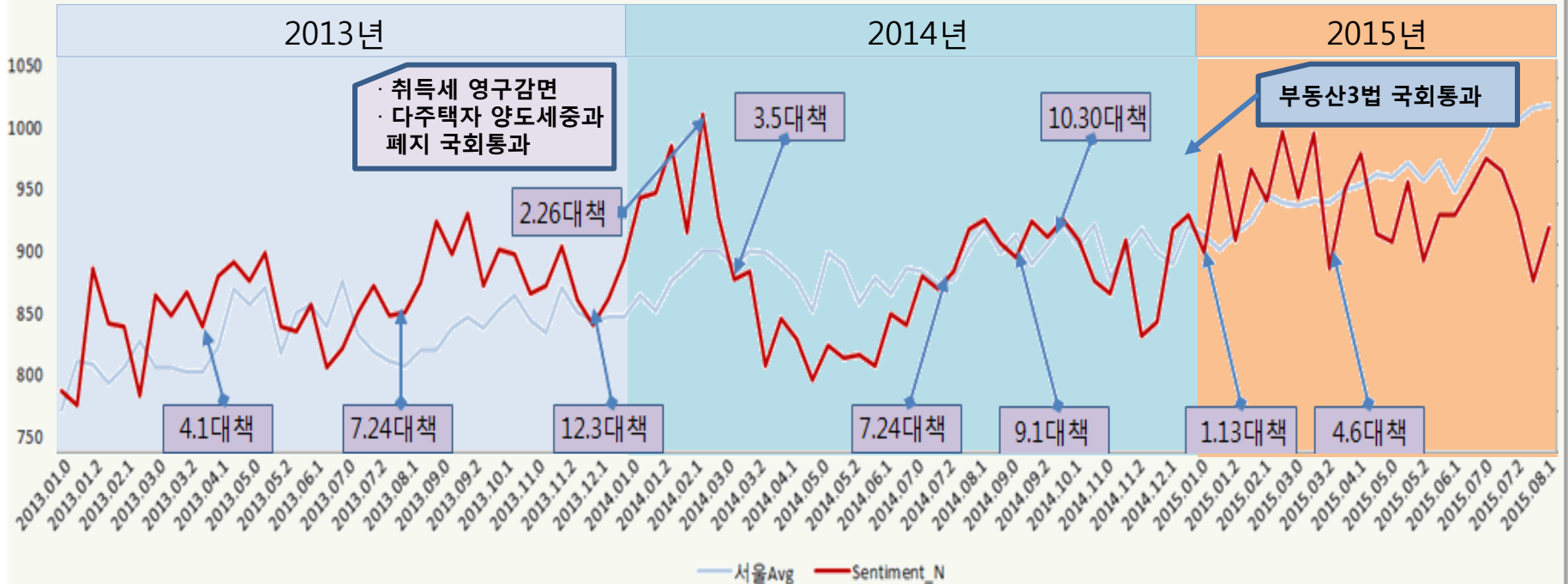


산출 데이터 정규화



실증 분석 결과

감성분석 데이터 vs 부동산정책 변화





연구 결과

연구 요약 / 안계

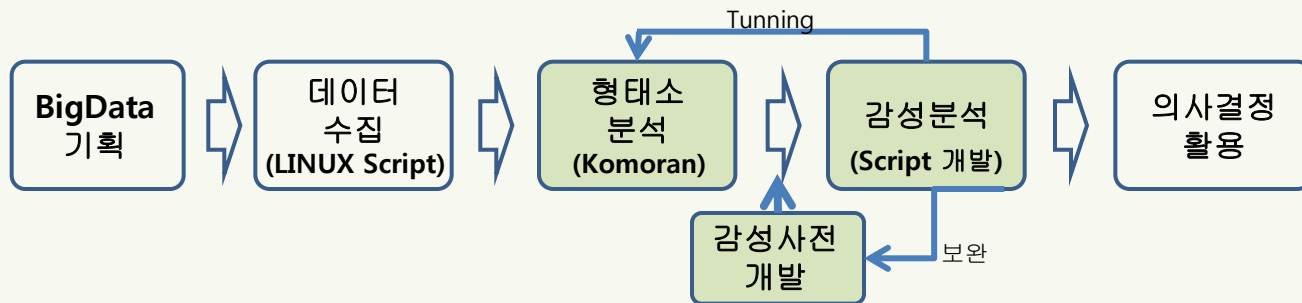
연구요약

- ▶ **비정형 데이터에 의한 부동산 상황인식, 문제해결, 예측 가능성에 대한 연구 시도**
 - 실용 수준의 연구 : 全量 수집, 감성분석의 최신 기술 적용
 - 분석 대상의 확대 : 정형데이터 + 비정형데이터(뉴스, SNS 등)
 - 부동산분야의 감성사전 개발
- ▶ **부동산 뉴스의 Text Mining과 Opinion Mining(감성분석)에 의한 부동산 의사결정 모형 제시**
 - 부동산 뉴스(비정형데이터) 의 감성분석에 의한 데이터 유의성 검증
- ▶ **특히 부동산 뉴스의 감성분석 결과와 실거래가격 데이터간 유사한 패턴 발견**
부동산 빅데이터에 의한 다양한 분야에서 본 연구의 제시 모형 활용 기대
 - 부동산 가격 예측
 - 정책 진단 및 평가 예측 등

연구 요약 / 안계

연구의 한계

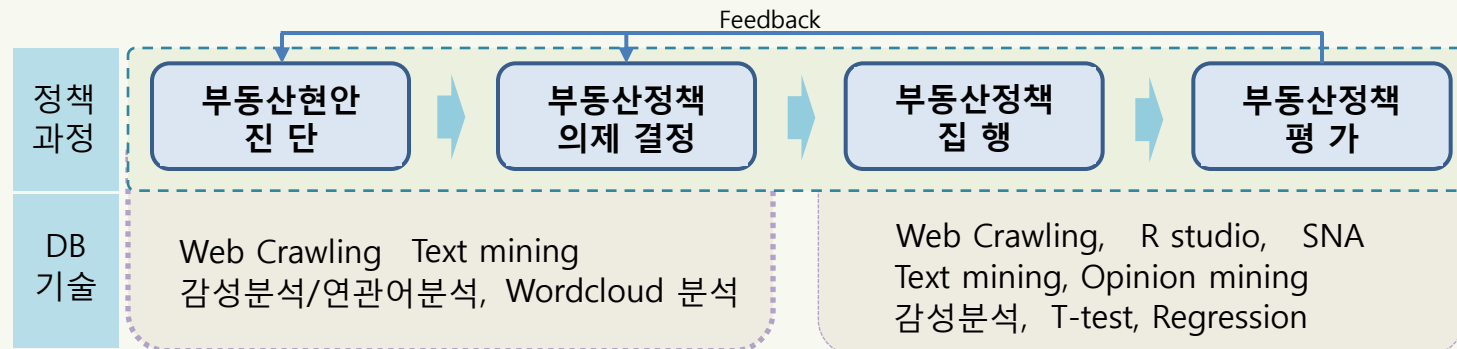
- ▶ 비교적 단기간 단선의 뉴스 데이터 수집으로 감성분석의 정확성과 일반화 제한
⇒ 뉴스 수집원의 다양화 및 수집기간 확대 필요
- ▶ 빅데이터 수집, 저장, 분석에 인력과 시간의 제한으로 감성분석의 정확성 미흡
⇒ 기획 : 수집 데이터 분류(형태, 지역, 시간 등)
⇒ 기술 : 형태소 분석과 감성사전 정교화



- ▶ 감성분석 결과 데이터와 가격데이터 비교를 통한 유효성 검증 미흡
⇒ 국토부 실거래가격 데이터 산출 대상 확대(일부에서 전국, 일일 단위))

양우 연구과제

- ▶ 부동산 뉴스의 감성분석 결과와 실거래 가격 데이터간 유사한 패턴 발견
부동산 빅데이터에 의한 다양한 분야에서 본 연구의 의사결정 모형 활용 기대
- ▶ **부동산 가격 예측 모형 개발**
 - 비정형데이터(뉴스, SNS 등) 감성분석에 의한 가격 예측
 - 다양한 정형데이터 + 비정형데이터 활용 가격 예측
- ▶ **부동산 정책 발굴 및 효과 예측**



참고문헌

1. 경정익(2013), "부동산, 빅데이터로 효율성 높여라", 디지털타임즈, 2013. 8. 19.
2. 경정익(2014), "부동산분야의 빅데이터 활용 방안과 정책적 제언", 「부동산경영」, 한국부동산경영학회, 제10호 pp.65-98
3. 경정익(2015), 「스마트 빅데이터 시대 부동산 정보화의 이해」, 두남출판사
4. 경정익(2015), "부동산정책 빅데이터로 신뢰 높여라", 디지털타임즈, 2015. 5. 26
5. 경정익 외(2015), "부동산분야에서 빅데이터 전략적 활용의 영향요인에 관한 연구", 「부동산산업연구」, 제1권 1호 pp.31-72
6. 경정익 외(2015), "부동산분야의 빅데이터 도입의도에 미치는 영향요인에 관한 연구", 「부동산분석」, 제1권 1호
7. 김대원·유정석(2013), "주택거래에 대한 심리적 태도가 주택매매거래량에 미치는 영향 분석", 「주택연구」, 제21권 2호, pp.73-92
8. 김대중(2013), "국토정책 선진화를 위한 빅데이터 활용에 관한 기초연구", 국토연구원
9. 김유신, 김남규, 정승렬(2012), "뉴스와 주가: 빅데이터 감성분석을 통한 지능형 투자 의사결정모형", 「지능정보연구」, 제18권 2호, pp.143-158
10. 김정선·송태민(2014), "빅데이터 기술수용의 초기 특성 연구", 한국콘텐츠학회 논문지 제14집 9호, pp.538-555
11. 송치영(2005), "뉴스가 금융시장에 미치는 영향에 관한 연구", 「국제경제연구」, 제8권 3호, pp.1-34.
12. 이완수·심재철·박양수(2007), "경제뉴스, 경제상황, 소비자 기대심리 그리고 소비행위의 상호 속성 의제설정 관계에 대한 시계열 분석", 「한국언론학보」, 제51권4호, pp.280-307
13. 조성우(2011), "Big Data 시대의 기술", KT종합기술원
14. 진창하(2012), "Newspaper content and home prices: perception, reasoning and affect", 「부동산학연구」, 제18집 2호, pp.126-142
15. Bing Liu(2012), "Sentiment Analysis and Opinion Mining".
16. J. Dimmick., Y. Chen, and Z. Li(2004), "Competition between the Internet and traditional news media: The gratification opportunities niche dimension," The Journal of Media Economic, Vol. 17. Issou 1, pp.19-33
17. LG경제연구원(2012), "빅데이터 시대의 한국, 갈라파고스 되지 않으려면"
18. Ashenfelter, Orley C, David Card(1985), "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs", The Review of Economics and Statistics, Vol. 67, Issue 4, pp.648-660
19. Blundell, Richard, Alan Duncan, Costas Meghir(1998), "Estimating Labor Supply Responses Using Tax Reforms", Econometrica, Vol. 66, Issue 4, pp.827-861
20. Card, David(1990), "The Impact of the Mariel Boatlift on the Miami Labor Market", Industrial and Labor Relations Review, Vol. 43, Issue 2, pp.245-257