

## 빅데이터 기반 소비자 유형별 농식품 추천시스템 구축 사례

문정훈\*, 장익훈\*, 최영찬°, 김진교\*\*, 박진\*\*\*

## Case Study of Big Data-Based Agri-food Recommendation System According to Types of Customers

Junghoon Moon\*, Ikhoon Jang\*, Young Chan Choe°, Jin Gyo Kim\*\*, Gene Bock\*\*\*

## 요약

농림수산식품교육문화정보원에서는 2015년 1월부터 공공데이터 포털 서비스를 시작하였으며 포털 내에 구축된 빅데이터 기반 농식품 추천 시스템을 이용한 맞춤형정보를 제공하고 있다. 추천시스템의 특징은 첫째, SNS오피니언마이닝, 소비자패널의 모든 구매내역 정보, 기후데이터, 도매가격 데이터와 같은 빅데이터의 성격을 가진 농식품분야의 다양한 데이터들을 이용하기 때문에 데이터 양의 관점에서 추천의 정확도를 높일 수 있다. 둘째, 추천시스템 구축 초기에는 사용자 정보 기반 추천이 어려운 한계를 극복할 수 있는 방법으로 식생활 라이프스타일과 메가트렌드 요인을 이용한 소비자 세분화방법을 사용한다. 이는 사용자 개인정보가 없는 상황에서도 다양한 식품 선호도를 반영할 수 있도록 하여 추천실패율을 낮춘다. 셋째, 디리슐레-다항분포를 이용하는 추천 알고리즘을 적용하여 다양한 상황적 요인들의 선호가 반영된 농식품 추천이 가능하도록 하였다. 이 외에도 추천 농식품에 대한 SNS 맛집정보와 버즈량, 관련 식재료를 판매하는 주변 소매점 위치 및 가격정보 등 다양한 정보를 제공하여 농식품 분야 정보에 관심을 높일 수 있도록 시스템을 구현하였다.

**Key Words** : Agri-food, Recommendation System, Types of Consumer, Food-Related Lifestyle, Big data Analysis

## ABSTRACT

The Korea Agency of Education, Promotion and Information Service in Food, Agriculture, Forestry and Fisheries launched a public data portal service in January 2015. The service provides customized information for consumers through an agri-food recommendation system built-in portal service. The recommendation system has following characteristics. First, the system can increase recommendation accuracy by using a wide variety of agri-food related data, including SNS opinion mining, consumer's purchase data, climate data, and wholesale price data. Second, the system uses segmentation method based on consumer's lifestyle and megatrends factors to overcome the cold start problem. Third, the system recommends agri-foods to users reflecting various preference contextual factors by using recommendation algorithm, dirichlet-multinomial distribution. In addition, the system provides diverse information related to recommended agri-foods to increase interest in agri-food of service users.

※ 본 연구는 미래창조과학부 및 정보통신기술진흥센터의 ICT융합고급인력과정지원사업의 연구결과로 수행되었음(IITP-2015-H8601-15-1007)

◆ First Author : Seoul National University Department of Agricultural Economics and Rural Development, moonj@snu.ac.kr, 정희원  
° Corresponding Author : Seoul National University Department of Agricultural Economics and Rural Development, aggi@snu.ac.kr, 정희원

\* Seoul National University Department of Agricultural Economics and Rural Development, iks0404@snu.ac.kr, 학생회원

\*\* Seoul National University Business School, jingkim@snu.ac.kr

\*\*\* Korea Agency of Education, Promotion & Information Service in Food, Agriculture, Forestry & Fisheries, cio@epis.or.kr, 정희원  
논문번호 : KICS2015-02-034, Received February 13, 2015; Revised May 6, 2015; Accepted May 6, 2015

## I. 서론

최근 빅데이터와 관련한 이슈가 대두되면서 2013년부터 안행부, 미래부 등 주요부처에서 빅데이터 관련 계획 수립과 시범사업을 통한 활용모델을 발굴하기 위해 노력하고 있다. 농식품 분야에서도 ICT융합 기술을 활용한 생산 및 유통 정보들, 개인 단위의 농식품 실구매 데이터 등 빅데이터의 속성을 가진 데이터들이 축적되면서 정부 3.0과 함께 농식품 공공 데이터의 수집, 관리, 개방을 통한 사회적 가치 창출이 요구되고 있는 시점이다. 농식품 분야의 공공데이터는 개방 완료된 120종 외에도 데이터베이스에 저장되어 있는 데이터도 290개에 달하는 것으로 조사되었다<sup>[1]</sup>. 그러나 데이터의 양에 비해 실효성 있는 양질의 데이터는 아직까지 많지 않은 것으로 평가되고 있는 만큼 기존에 보유 중인 데이터의 고도화도 중요한 이슈이다.

농림수산식품교육문화정보원(이하 농정원)에서는 2014년, “농식품 공공데이터 개방 및 빅데이터 모델 개발” 사업을 통해 “농림축산식품부 공공데이터 포털”을 구축하였는데 기존 농식품분야 공공데이터의 고도화와 함께 기보유 중인 공공데이터를 이용한 빅데이터 기반 농식품 추천 시스템을 공공데이터 포털 내의 하위 서비스로 개발하여 2015년 1월부터 제공하고 있다. 본 연구에서 소개하게 될 이 추천 시스템의 개발 목적은 농식품분야의 빅데이터를 이용하여 소비자의 니즈에 맞는 농식품 정보를 제공함으로써 추천시스템 사용자를 확대하는 것이다. 장기적으로 추천시스템 사용자가 많아진다면 생산자들을 연결하는 농식품 직거래 플랫폼과 연계하여 지역 농산물 소비를 촉진하는 것도 기대할 수 있다. 그러나 구축 초기인 현 시점에는 사용자 수가 적기 때문에 기존의 연구들에서 제시하는 협업필터링과 같은 사용자 정보 기반 추천은 부적절하다. 이러한 한계점을 극복하기 위해 본 추천 시스템은 설문조사를 통해 소비자 유형을 사전에 세분화하여 각 유형별로 선호하는 농식품을 추천할 수 있도록 하였고, 과거의 실제 구매내역과 현재의 선호의향 정보를 모두 고려한 선호 농식품 리스트를 추출하고 날씨별, 계절별, 가격변동과 같은 메가트렌드에 의한 선호체계 변화도 일 단위로 감지하여 반영할 수 있도록 하여 소비자 개인 프로파일 및 과거 이용 내역이 없는 상황에서도 추천 정확도를 극대화할 수 있도록 시스템을 설계하였다.

본 연구에서는 선행연구를 통해 추천시스템의 특징과 연구 범위, 식품 관련 추천시스템 연구의 흐름을 알아보고 적절한 추천알고리즘을 검토한다. 이를 통해

본 연구의 추천시스템이 기존 추천시스템과 보이는 차이점과 유사점을 고찰하고, 추천시스템의 구축 방향과 구현 내용을 구체적으로 소개한다. 결론에서는 본 연구의 추천시스템의 한계점과 향후 연구방향을 검토한 후 농식품 추천 시스템이 정보통신기술 분야에 미치는 기대효과에 대해서 논의한다.

## II. 선행 연구

### 2.1 추천 시스템의 특징과 현황

추천 시스템은 추천의 기법에 따라서 내용기반(content-based), 협력기반(collaborative), 그리고 이 둘을 결합한 하이브리드(hybrid) 형태의 세 가지 접근법으로 분류할 수 있다. 내용 기반의 추천 기법은 사용자가 특정 아이템을 선호하면 이와 관련성이 높은 아이템을 추천하는 방식으로 추천 대상 아이템이 자주 바뀌지 않을 때 적합한 방법이다<sup>[2]</sup>. 협력 기반 추천 기법은 사용자들의 과거 성향이 지속적으로 유지되는 것을 가정하여 취향이 비슷한 사용자들에게 아직 구매하지 않은 상품을 교차 추천하는 방법이다. 추천 대상 아이템이 자주 바뀔 때 유용한 방법이다<sup>[2]</sup>.

내용 기반 추천 기법은 인구 통계적 특성을 얻기 힘든 경우가 많고 사용자가 명시한 선호 특성들을 구하기 어려운 문제를 가진다<sup>[3]</sup>. 협력 기반 추천 기법은 많은 연구들에서 비교적 정확도가 높은 것으로 평가되고 있으나 이 역시 사용자들의 평가정보를 얻기 힘든 경우가 많고, 평가 정보의 양이 많지 않을 경우 추천성능이 크게 감소하는 문제가 있다<sup>[3]</sup>. 하이브리드 방식은 이들 추천 기법의 단점을 보완하기 위해 각각의 특징을 결합한 추천기법으로 가중(Weighted), 전환(Switching), 혼합(Mixed), 기능조합(Feature Combination), 캐스케이드(Cascade), 메타레벨(Meta-Level)과 같은 방법이 사용된다<sup>[36]</sup>.

추천 시스템 관련 최근 연구 동향은 정확도 향상을

표 1. 추천시스템 관련 선행연구 유형 및 연구대상  
Table 1. Previous research types and subjects

Type	Research subjects and reference
Algorithm improvement	User classification <sup>[4,5]</sup> , Preference/Probability calculation <sup>[6-8]</sup> Contents-based filtering <sup>[9-11]</sup> , Kernal function <sup>[12]</sup> , Weight method <sup>[13]</sup>
Additional information usage	Questionnaire <sup>[14-16]</sup> , Previous case <sup>[17-19]</sup> Twitter profile and twit of an user <sup>[2]</sup> , Context information based on ubiquitous <sup>[20-22]</sup>

위해 추천 알고리즘을 개선하는 방향과 추가정보를 수집하는 방향의 연구가 다양한 방법으로 진행되어 왔다. <표 1>는 선행연구들을 정리한 것으로 대부분 영화, 음악, 방송과 같은 콘텐츠 상품을 다루고 있는 것이 특징이다.

## 2.2 식품 관련 추천시스템

음식을 대상으로 하는 추천시스템 연구는 활발하지는 않았으나 최근에는 몇몇 사례들을 찾아볼 수 있다. 음식 추천 시스템은 추천의 목적에 따라 일반 사용자용과 환자용으로 구분된다. 환자용 추천시스템은 식단 관리가 필요한 당뇨병이나 신장병 환자를 대상으로 영양정보에 기반한 음식을 추천한다<sup>[17,19]</sup>. 한편 일반 사용자용 추천시스템은 건강한 식단을 제공하기 위한 목적과<sup>[16,22]</sup> 즐겨 먹는 음식의 특징을 파악하여 소비자의 메뉴선택 의사결정을 도와 만족도를 높이는 목적으로<sup>[2,15,18]</sup> 다시 구분할 수 있다. 환자용 식단 추천 시스템의 경우 음식에 포함된 영양정보의 정확도가 매우 중요한 요인이기 때문에 영양정보가 포함된 방대한 양의 식품 데이터베이스로부터 특정한 조건의 음식을 분류하여 추천하는 것이 핵심이며 사례기반추론<sup>[17]</sup>이나 K-means, Self-Organizing Map(SOM)과 같은 군집분석 기법을 활용하여<sup>[18]</sup> 추천의 정확도를 높이는 것에 초점을 맞추고 있다. 일반인을 대상으로 하는 건강식단 추천시스템 연구로는 사용자의 키, 몸무게, 근육량 정보를 수집한 후 퍼지추론으로 비만도를 도출한 후 해당 비만도별로 고려되어야 하는 칼로리와 영양정보를 기준으로 식품을 선별하여 추천하는 방법<sup>[16]</sup>, 생체신호 센서와 주변환경 인식 센서와 같은 유비쿼터스 기술을 이용하여 사용자의 건강과 식사맥락을 파악하여 최적의 음식을 추천하는 방법<sup>[22]</sup>이 소개되었다. 사용자의 메뉴선택 의사결정을 지원하는 추천시스템으로는 과거의 식품 주문정보를 학습하여 선호도가 높은 식품을 추천하는 방법<sup>[18]</sup>, 사용자에게 직접 요구사항을 질문한 뒤 텍스트마이닝으로 사용자 니즈를 추출하여 영양정보를 기반으로 해당 질문과 가장 관계가 큰 식품을 추천하는 방법<sup>[15]</sup>, 그리고 사용자의 트위터 계정과 연동하여 작성된 트윗을 텍스트 마이닝하여 선호하는 식품 정보를 추출하여 추천에 활용하는 방법<sup>[2]</sup>이 소개되었다.

## 2.3 Latent Dirichlet Allocation 기반 추천 알고리즘

추천시스템에서 가장 많이 사용되는 협업 필터링 방법의 경우 평가 정보가 부족한 상황에서는 추천의

정확도가 매우 떨어지는 단점이 있다<sup>[3]</sup>. 이러한 평가 데이터 부족의 문제를 보완할 수 있는 방법으로 과거의 행동정보와 사용자(아이템)의 숨겨진 특성을 이용하여 평가정보를 대체 하는 방법으로 토픽 모델링이 사용될 수 있다. 이 기법은 문서 내에서 은닉 주제들을 찾아내기 위해 개발된 통계 추론 방법으로 LDA(Latent Dirichlet Allocation) 알고리즘이 가장 많이 사용된다. LDA 알고리즘은 Dirichlet 분포와 다항 분포 사이에 공액사전분포함수(conjugate prior) 관계가 성립하는 이점을 활용하여 문서의 주제 분포와 각 주제에 해당되는 단어의 출현 확률정보를 이용하여 주어진 문서가 어떤 주제인지를 통계적으로 계산할 수 있다<sup>[23]</sup>. 또한 LDA는 문서뿐만 아니라 비슷한 형태의 모든 데이터 분석에 활용될 수 있는 장점을 가진다<sup>[25]</sup>. 예를들면 LDA를 적용하기 위한 데이터 구성 방법으로 LDA에서의 문서를 소비자 유형으로, LDA에서의 단어들을 특정 유형 소비자가 언급했거나 구매한 음식들로 정의할 수 있다. 실제로 문서의 토픽 분류 분야 뿐 아니라 TV프로그램 추천<sup>[24,26,27]</sup>, 영화 추천<sup>[28]</sup>에서도 사용되어 추천의 정확도를 높이는 것이 확인되었다.

## III. 빅데이터 기반 소비자유형별 농식품 추천시스템 분석

### 3.1 추천 시스템의 구축 목적

본 연구에서 소개하는 빅데이터 기반의 농식품 추천 시스템은 일반적인 추천시스템에서 개인별 정보를 이용하는 것과는 달리 사전에 정의된 사용자 유형별 선호 정보를 이용하여 추천을 실행한다. 따라서 사용자 개인의 프로필 정보나 이용내역이 없는 상황에서도 사용자 유형별로 맞춤형 정보를 제공하는 것이 가능하며, 추천 시스템의 사용자 수가 추천 정확도를 확보할 수 있는 일정 수준 이상 도달하기 이전 단계에서 유용하게 활용될 수 있다. 이와 같이 유형별 선호도를 파악할 수 있게 된 것은 과거에는 다루기 어려웠거나 존재하지 않았던 소셜네트워크 서비스에서의 소비자 오피니언 정보, 소비자 패널의 식료품 구매 이력 정보들이 빅데이터의 영역에서 접근 가능해졌기 때문이다. 요약하면 본 추천 시스템의 목적은 개인별 정보를 이용한 맞춤형 추천의 이전 단계에서 활용될 수 있는 단순화된 형태의 맞춤형 추천 시스템의 개념을 제시하는 것을 목적으로 하며, <그림 1>과 같은 구조로 빅데이터 기반의 소비자 유형별 농식품 추천이 이루어진다.

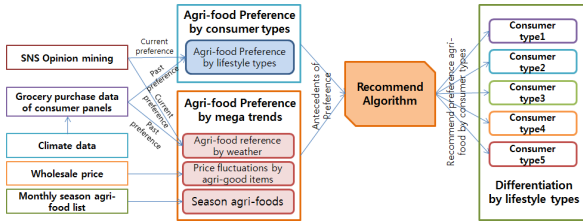


그림 1. 목표 농식품 추천 시스템 개념도  
Fig. 1. The conceptual diagram of target agri-food recommendation system

### 3.2 식생활 라이프스타일과 메가트렌드 고려

식생활 라이프스타일은 [29]가 처음으로 제시한 이론으로 소비자들의 식생활 행동패턴을 이해하기 위해 라이프스타일 차원의 측정도구를 개발해 새로운 관점에서의 라이프스타일 이론을 발전시켰다. [30]은 식생활 라이프 스타일의 요인을 식품구매 동기, 구매 제품의 특성, 구매 의사 결정의 목적, 조리 방법, 소비 상황을 기준으로 식품 소비와 관련된 라이프 스타일을 제시하고 이 후의 연구에서는 라이프스타일 변수를 이용한 소비자 유형 세분화가 진행되었다. 이와 같이 농식품 소비자의 유형을 분류하는 것은 유형별 니즈를 세분화할 수 있기 때문에 각각의 소비자 유형이 원하는 정보를 차별화하여 제공하는 것이 가능해진다.

메가트렌드는 농식품 수요에 영향을 미치는 거시요인들을 의미하는 것으로 가격의 변화, 계절별 제철 식재료 변화 및 날씨 변화에 의한 농식품 선택이 영향을 받으므로 이러한 요인들을 고려하여 추천을 하였을 때 정확도를 더 높일 수 있다. 실제로 음악 추천 시스템에서는 계절, 날씨, 요일, 기온을 고려한 음악 추천으로 사용자 만족도를 높이고 있다<sup>[20]</sup>.

### 3.3 농식품 리스트와 외부 판매처 연계

소비자에게 유형별 맞춤형 농식품 추천 정보를 제공하는 것과는 별개로 농식품 추천시스템이 추구하는 또다른 목표는 추천 메뉴 및 식재료 리스트의 하이퍼링크를 통해 지역 농식품을 취급하는 외부 판매처들과 연계를 하는 것이다. 즉, 추천시스템이 사용자가 선호할만한 농식품을 추천하는 것에서 그치지 않고 링크를 통해 해당 농식품의 소비와 관련한 판매처 정보까지도 연결시켜주는 것이다. <그림 2>는 이러한 연계를 ERD(Entity Relationship Diagram) 형태로 나타낸 것으로 메뉴와 식재료 리스트를 중심으로 식당 및 온·오프라인 판매점 정보가 연결되는 것을 보여준다.

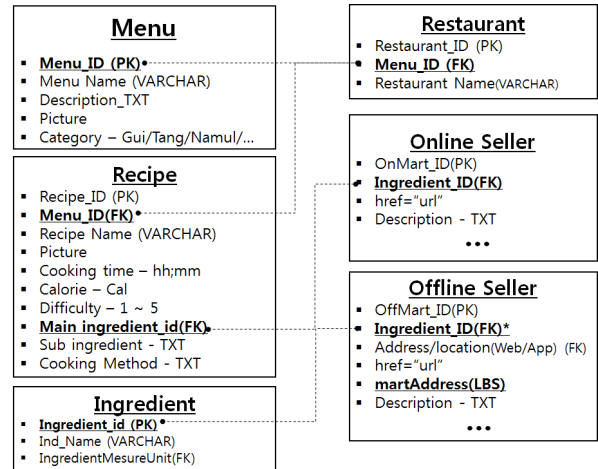


그림 2. 추천 농식품 연관 ERD  
Fig. 2. The ERD of recommended agri-food

## IV. 빅데이터 기반 소비자유형별 농식품 추천시스템 설계

### 4.1 활용 데이터 범위

빅데이터 기반 농식품 추천 시스템에 사용되는 데이터는 <표 2>와 같다. 농식품 추천시스템은 기본적으로 소비자 유형을 분류하기 위한 데이터, 소비자 선호도 분석을 위한 데이터, 그리고 메가트렌드를 감지하기 위한 데이터가 필요하다. 소비자 유형 분석은 농촌진흥청에서 보유하고 있는 수도권 소비자패널을 대상으로 식생활 라이프스타일 관련 설문을 수집하여 유형분류 분석에 활용하였다. 농촌진흥청 소비자패널의 경우 모든 농식품 구매내역 정보를 확인할 수 있기 때문에 소비자 인식에 대한 설문결과와 실제 소비패턴 정보를 연결시킬 수 있어 활용 가치가 높다. 소비자 농식품 선호도 분석은 SNS에서의 오피니언마이닝 정보, 수도권 소비자패널의 농식품 구매내역 정보를 이용하였으며, 소비자 유형별로 선호 농식품 리스트와 빈도를 계산하여 선호도 정보를 추출하였다. 이와 같이 SNS와 구매내역 정보를 동시에 활용하는 것은 소비자 유형별 농식품 선호 정보를 구매내역과 같은 과거의 선호 뿐만 아니라 SNS를 통해 현재의 선호까지 동시에 반영하기 위해서이다. 메가트렌드의 감지는 기상청의 기후 데이터로부터는 과거의 특정날씨의 시점을 도출하여 소비자패널이 해당 날씨에 실제 구매한 농식품 리스트를 추출하는데 사용하였고, 가격변화 정보는 도매시장 가격정보를 수집하여 품목별 시세하락을 감지하는데 사용하였다. 제철 농식품 리스트는 주요 포털서비스에서 수집하였으며 월별 제철 농식품의 추천을 위해 사용하였다. 메가트렌드는 소비자 유형에

표 2. 빅데이터 기반 농식품 추천시스템 사용 데이터  
Table 2. Data sources of agri-food recommendation system based on bigdata

Types	Purpose	Source
Survey data of metropolitan consumer panels	Classification of food-related lifestyle types	Rural Development Agency(2014)
SNS opinion mining data	Analyzing Agri-food preference of SNS users	Daumsoft (past 1 week)
Agri-food purchase data of metropolitan consumer panels	Analyzing Agri-food preference of metropolitan households	Rural Development Agency (2010~2013)
Weather data	Extracting the day with Particular weather in the past	Korea Meteorological Administration (2010~2013)
Wholesale price	Extracting price decline pattern	Seoul Agro-Fisheries & Food Corporation (2010~2014)
Season agri-food list	Extracting Monthly season agri-food	Naver, Daum

관계없이 선호도에 영향을 줄 수 있는 요인들을 반영하기 위해 고려하였다.

4.2 식생활 라이프스타일 유형에 따른 소비자 분류  
식생활 라이프스타일(Food-Related Lifestyle: FRL)은 많은 학자들에 의해 연구된 이론으로 FRL 분류를 위한 측정문항은 꾸준히 개선되어 왔다. 본 연구에서

표 3. 식생활 라이프스타일 유형별 특징  
Table 3. Characteristics by Food-related lifestyle type

Types	Characteristics
Health menu seek group	Prefer low-calorie, low-salt, low-sugar food
Food security oriented group	Prefer organic food, Confirm place of origin and certification(GAP, pesticide-free)
Economic oriented group	Prefer inexpensive grocery, Check price information, Consider price when purchasing
Food adventurer group	Prefer a new experience with foods, unusual restaurant, Try to cook novel menu
Local Food seek group	Prefer traditional foods, Interest in local restaurants and agri-foods

는 [31]이 제시한 FRL 측정지표를 참고하여 설문문항을 개발하였으며 조사대상은 농촌진흥청 소비자패널 800명을 대상으로 설문지를 배포한 후 643부를 회수하여 분석에 사용하였다. 분석방법은 먼저 요인분석을 실시하여 측정된 38개 라이프스타일 지표들로부터 10개의 요인을 추출하였다. 다음으로 분류된 10개의 라이프스타일 요인별 상위집단의 인구통계변수, 식생활 패턴, 품목별구입액의 차이를 비교하여 요인별로 유의한 차이를 보이는 5개 요인을 선별하여 본 추천시스템에서의 식생활 라이프스타일 유형으로 정의하였다. <표 3>는 각 유형의 특징을 정리한 것이다.

### 4.3 농식품 메뉴 및 식자재 리스트 표준화

농식품 추천 시스템에서 제공하는 메뉴와 식자재 리스트의 범위는 ‘피자’와 같이 포함된 식재료에 따라 다수의 명칭을 가지는 경우 ‘피자’ 대신 ‘고르곤졸라 피자’로 분류되어 피자에 대한 선호도가 아닌 피자 세부 종류별 선호도로 계산되게 된다. 선호도는 선택확률에 기반하기 때문에 하위 명칭을 사용할 경우 선호도가 분산되는 문제가 발생한다. 따라서 모든 농식품 리스트를 사용하는 것은 바람직하지 않기 때문에 소비자가 유사한 종류로 인식하는 농식품은 동일한 카테고리 맵핑하여 메뉴 및 식자재 리스트를 표준화하였다. 이러한 과정을 통해 소비자패널 데이터에 기록되어 있는 1만6000여 개의 서로다른 메뉴들은 651개의 표준화 메뉴로 변환하였고, 식자재의 경우 채소, 과일, 축산, 수산물 품목 전체에서 소비금액 기준으로 상위 99개만 표준화 식재료로 선정하였다.

### 4.4 과학적 방법에 의한 농식품 추천

본 연구에서의 농식품 추천 시스템은 소비자 유형별 선호도뿐만 아니라, 메가트렌드(날씨, 가격, 계절)에 의한 맥락적 요인이 동시에 고려되어야 한다. 또한 소비자 유형별 선호도의 경우 과거의 구매행동에 의한 결과와 현재의 소비성향을 구분하여 현재의 선택에 서로 다른 경로로 영향을 주는 것을 고려해야 한다. 따라서 이와같은 여러 가지 고려사항들의 조합별로 적절한 추천이 이루어지게 하기 위해서는 다양한 고려사항의 조합에 대한 선호도가 소비자 유형별로 다르게 나타나는 것을 반영하여 추천이 이루어지는 것이 바람직하다. 따라서 본 추천시스템에서는 고려사항의 선호도를 과거의 선택빈도를 통해 산출하기 위해 확률이론에 기반하는 디리슬레-다항분포(Dirichlet-Multinomial Distribution) 기반의 추천알고리즘을 적용하여 설계하였다. 디리슬레-다항분포 모형은 시간

단위를 특질 변수들의 추가로 주별, 월별 혹은 특정 시기(예를 들어 명절이나 각종 기념일 기준)로 확장 가능하며, 예측값을 구하는 기간을 조정하는 것도 가능하다. 따라서 추가적인 피드백을 쉽게 수용할 수 있는 장점과 지속적 개선 가능성을 모두 갖추고 있다.

### V. 빅데이터 기반 소비자유형별 농식품 추천시스템 구현

#### 5.1 소비자 유형별 선호 농식품 리스트 도출

소비자 유형별 선호 농식품 리스트는 과거 시점의 선호를 확인할 수 있는 소비자패널 구매내역 데이터와 현재 시점의 선호를 확인할 수 있는 SNS 오피니언 마이닝 데이터로부터 도출하였다. 소비자패널의 경우 앞서 정의된 5가지 소비자유형의 집단에서 과거 4년간 구매한 농식품을 표준화 리스트에 맞춰 주별 빈도값을 구한 후 순위를 구하였다. 순위값은 다시 유형별로 비교하여 타 유형대비 순위가 높은 농식품을 해당 유형의 과거 선호 리스트로 추출하였고, 빈도값은 선호도 산출에 필요한 사전확률 계산의 입력값으로 사용하였다.

SNS 오피니언 마이닝 데이터는 다음소프트에서 제공하는 소셜메트릭스 서비스에서 제공받았으며, 추천이 이루어지는 시점으로부터 최근 1주일 동안의 트위터, 블로그, 뉴스 등의 SNS에 작성된 게시물을 대상으로 하였다. 유형별 선호 농식품 도출은 <표 3>에서와 같이 5가지 소비자 유형별 특징에 해당되는 탐색 키워드와 먹는 상황에 해당되는 공통 탐색어를 검색 조건으로 하여 연관 탐색어를 수집하고, 750개(메뉴 651개, 식재료 99개)의 표준화 리스트를 필터로 하여 식품과 무관한 탐색어는 제외되도록 하였다. 과거 선호 리스트와 동일한 방법으로 빈도값 기준의 순위를 구한 후 타 유형 대비 순위가 높은 농식품을 해당 유형의 현재 선호 리스트로 추출하였다. 빈도값 역시 사전확률 계산의 입력값으로 사용하였다.

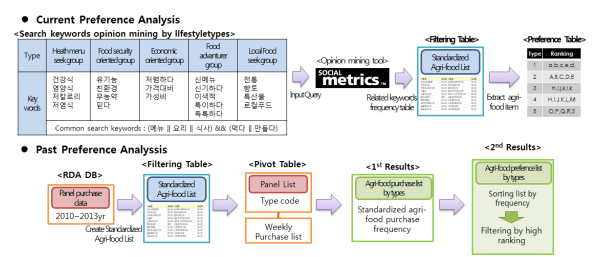


그림 3. 소비자 유형별 선호 농식품 리스트 도출 절차  
Fig. 3. The procedure of creating an agri-food preference list according to consumer types

#### 5.2 메가트렌드에 의한 선호 농식품 리스트 도출

##### 5.2.1 날씨 유형별 선호 농식품 리스트 도출

농식품 선택 상황에서 메가트렌드는 소비자유형에 의한 영향에 앞서 모든 소비자유형에 동일하게 영향을 주는 요인을 의미한다. 날씨도 그 중 하나로 날씨 유형별 선호 농식품 리스트는 소비자 유형별 선호 리스트와 동일하게 과거 시점의 선호를 확인할 수 있는 소비자패널 구매내역 데이터와 현재 시점의 선호를 확인할 수 있는 SNS 오피니언마이닝 데이터로부터 도출하였다. 소비자패널의 경우 기상청 기후데이터로부터 특정 날씨의 시점을 구하여 과거 4년간의 해당 날씨에 구매한 농식품을 표준화 리스트에 맞춰 빈도값을 구한 후 순위를 구하였다. 날씨 유형별 선호의 경우 유형 간 선호 농식품의 차이가 뚜렷한편이어서 유형간 순위비교 과정은 제외하였으나 빈도값은 사전확률 계산의 입력값으로 사용하였다.

SNS 오피니언 마이닝 데이터는 추천이 이루어지는 시점으로부터 최근 1주일 동안의 트위터, 블로그, 뉴스 등의 SNS에 작성된 게시물을 대상으로 하였다. 날씨별 선호 농식품 도출은 <그림 4>에서와 같이 5가지 날씨 유형별 특징에 해당되는 탐색키워드와 먹는 상황에 해당되는 공통 탐색어를 검색조건으로 하여 연관 탐색어를 수집하고, 동일한 방법으로 750개(메뉴 651개, 식재료 99개)의 표준화 리스트를 필터로 하여 식품과 무관한 탐색어는 제외되도록 하였다. 순위비교는 제외하였고 빈도값은 사전확률 계산의 입력값으로 사용하였다.

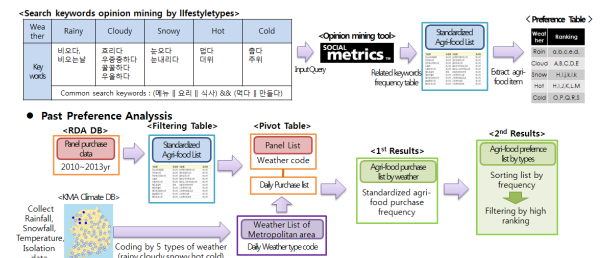


그림 4. 날씨 유형별 선호 농식품 리스트 도출 절차  
Fig. 4. The procedure of creating an agri-food preference list according to weather information

##### 5.2.2 제철농식품 및 시세하락 품목 도출

메가트렌드 중 계절적인 요인은 농식품 선택에서 주요한 영향을 주는 요인이다. 따라서 제철농식품에 대한 적절한 추천을 하는 것은 중요하다. 제철농식품 리스트는 주요 포털서비스가 보유하고 있는 콘텐츠를 참고하여 도출하였다. 본 추천시스템에서는 월별 제철

농식품 리스트에서 각 농식품의 가중치는 모두 동일한 것으로 가정하였기 때문에 현재의 시스템에서는 사전확률이 동일하게 적용되게 된다. 시세하락 품목의 도출은 소비자의 관점에서는 가격이 저렴하게 형성된 농산물을 사기 때문에 편익이 발생하며, 생산자의 관점에서는 가격이 급락하는 상황에서 농식품 추천을 통한 수요 확대를 기대할 수 있게 된다. 가격 하락에 대한 정보는 과거 5년간 도매시장 경락가격에서 최대치, 최저치를 제외한 3년치의 평균값 대비 오늘의 도매경락가격의 높고 낮음을 비율 값으로 하여 사전확률 계산의 입력값으로 사용하였다. 과거 대비 하락폭이 크면 클수록 확률값이 커지게 되어 추천될 가능성이 높아지게 된다.

### 5.3 소비자 유형별 농식품 추천 결과 표출

소비자 유형별 추천을 위한 입력 데이터는 앞서 설명한 것과 같이 수집 및 가공을 거쳐 추천알고리즘의 분석 데이터로 사용된다. 추천 알고리즘은 다리솔레다항분포 모형을 이용하여 소비자 유형과 메가트렌드별 모든 경우의 수에 대한 조합별로 해당 조건에 따른 추천 농식품 리스트를 확률값과 랭크값과 함께 출력한다. 즉, 해당 조건에서 가장 선호확률이 높은 농식품을 1순위라고 하고 가장 선호확률이 낮은 농식품을 최하위로 두는 리스트를 제공하는 것이다. 따라서 추천 결과의 표출은 리스트의 상위 몇위까지를 추출할 것인지를 결정한 후 웹페이지 UI에서 해당 순위까지의 리스트를 제공하였다. 본 추천시스템에서는 메뉴 5건, 식재료(채소, 과일, 수산물) 카테고리별 각 3건씩을 추천하였고, <그림 5>과 같이 각 유형별 추천 농식품 리스트를 상단 탭을 바꾸어가면서 확인이 가능하도록



그림 5. 소비자 유형별 농식품 추천결과  
Fig. 5. Agri-food recommendation results for each type of consumer

하였다. <그림 5>의 추천 농식품 리스트의 우측편에는 아이콘이 추가된 것을 확인할수 있는데, “℃”는 날씨추천, “↓”는 시세하락추천, 나뭇잎 아이콘은 제철추천을 의미한다.

맞춤정보 제공을 위해서는 사용자 로그인이 필요하다. 회원가입을 할 경우 식생활 관련한 간단한 설문에 응답해야 해야 하는데 이 설문의 응답결과를 이용하여 유형분류 분석에서 도출한 판별규칙에 의해 가입자의 성향이 가장 높게 나타나는 하나의 유형을 지정해주게 된다. 그리고 UI상에서도 본인에 해당되는 유형에 대한 농식품 추천정보를 확인할 수 있고 추가적으로 해당 유형과 관련성이 높은 SNS 및 공공영역의 정보를 제공하여 식생활과 관련한 다양하고 풍부한 지식을 사용자에게 전달하고 있다.

<그림 6>은 이번주의 추천 메뉴 중 삼계탕을 클릭하였을 때 추천정보 하단에 나타나는 화면으로 메뉴를 구성하는 주재료 및 부재료 정보와 SNS 상에서의 맛집관련 연관어, SNS 버즈량, 주요 오피니언 원문정보를 제공하고 있다. 이번주의 추천 메뉴 외에 식재료에 해당되는 농수축산물을 클릭하거나 추천 메뉴 하위의 주재료 및 부재료를 클릭하게 되면 내 위치 인근의 해당 식재료를 판매하는 소매점 위치와 가격 정보를 제공해주는 기능까지 구현하였다. 이와같이 농식품 추천을 시작으로 소비자들은 레시피, SNS여론, 소

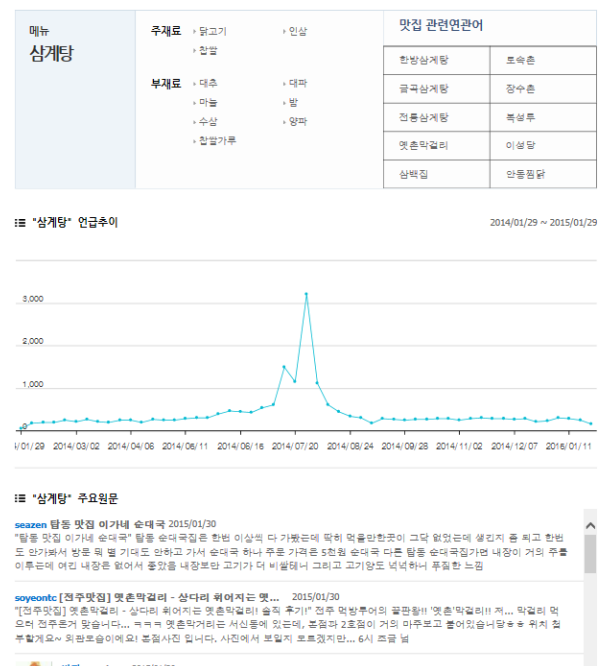


그림 6. 특정 메뉴에 대한 추가정보 제공 UI  
Fig. 6. UI reflected with additional information for a specific menu

매점 위치 및 가격정보 등 다양한 농식품 관련 정보를 접할 수 있도록 할 경우 소비자는 원하는 정보에 빠르게 접근할 수 있기 때문에 서비스에 대한 만족도가 높아지고 지속적인 서비스 이용 동기를 부여하게 된다.

## VI. 결론 및 향후 연구 방향

본 연구에서 소개하고 있는 빅데이터 기반 소비자 유형별 농식품 추천 시스템이 가지는 의미는 다음과 같다. 첫째, SNS 오피니언 마이닝, 소비자패널 구매이력, 기상청 기후데이터, 도매시장 경락가격정보와 같은 빅데이터의 특징을 가지는 다양한 데이터를 연계하였으며 이를 활용하여 분석결과의 품질을 높였다. 둘째, 기존의 추천시스템들과는 차별화되는 방법으로 사용자층이 확보되지 않은 서비스 초기 단계에서도 비교적 정확한 추천이 가능하게 하였다. 셋째, 농식품 소비자들에게 가치있는 정보를 제공할 수 있는 비즈니스 모델 개발 사례를 보여주었다. 넷째, 본 추천시스템과 같은 정보제공 서비스가 안정적으로 자리잡게 된다면 향후 ICT융합 기술과의 접목을 통해 더 많은 가치를 제공할 수 있는 기회를 제공할 수 있다.

그러나 본 연구는 몇 가지 한계점을 가지고 있는데, 첫째, 추천결과에 대한 사용자의 평가가 이루어지지 않아서 본 연구에서 주장하고 있는 빅데이터 기반의 추천방법이 어느정도 정확도를 가지는지 검증하지 못했다. 이는 정식서비스가 론칭된지 1개월밖에 되지 않았기 때문에 평가가 어려운 점도 있었지만 서비스 사용자의 평가를 수집할 수 있는 기능이 구현되지 않은 것이 직접적인 문제이므로 향후 개선이 필요하다. 둘째, 현재 5개로 정의된 소비자 유형으로는 다양한 소비자의 니즈를 모두 반영하는 것이 어렵다. 따라서 서비스 초기단계에서는 5개의 유형에 대한 맞춤형 정보를 제공할 수 있지만 사용자들이 지속적으로 흥미를 느끼고 사용하게 하기 위해서는 보다 개인화된 추천이 필요하다. 이를 위해서는 향후 클릭스트림 정보나 추천에 대한 평가결과를 이용하여 개인별로 선호도가 업데이트될 수 있도록 시스템 개선이 필요하다.

농식품 소비자들의 건강과 식생활에 대한 관심은 점차 높아지고 있다. 향후 ICT융합 기술의 확대로 산업차원에서는 생산-가공-유통-판매에 이르는 공급사슬에서의 이력정보가 투명해지고, 소비자 단계에서는 상황인식 및 생체신호 센서와 같은 기술이 보편화되면 농식품 소비자들에게 신뢰할 수 있는 농식품 정보와 정교한 맞춤형 추천이 가능해지게 될 것이다. 농식품 공급사슬의 이력정보 투명화를 위해서는 RFID/

USN 기술을 이용한 물류시스템의 구축이 필요하며<sup>[32]</sup>, 농식품 소비 단계에서의 상황인식이 가능하기 위해서는 [33]에서 소개하고 있는 유비쿼터스 기반의 홈네트워크 기술의 상용화가 선행되어야 한다. 홈네트워크 기술이 일반화 될 경우 음성인식을 기술<sup>[34]</sup>을 통해 사용자의 요구사항을 추천시스템과 연계할 수 있고, 생체정보 인식기술<sup>[35]</sup>이 적용될 경우 사용자의 건강상태가 실시간으로 반영된 식단 추천도 가능해진다.

미래의 ICT융합 기반 비즈니스 모델이 활성화되기 위해서는 이와 같은 유비쿼터스 요소 기술들이 발전도 중요하다. 하지만 무엇보다 중요한 것은 농식품 정보제공 서비스가 활성화되어 다수의 사용자 확보가 선행되어야 한다. 이를 위해 본 연구에서 소개하는 빅데이터 기반 농식품 추천 시스템이 농식품 분야의 대표적인 정보제공 서비스의 기틀을 마련할 수 있기를 기대해 본다.

## References

- [1] EPIS, *RFP for research on agri-food public open data and development bigdata business model*, pp. 3, 2014.
- [2] J. S. Jeong and S. J. Kang "Hybrid food recommendation system using auto-generated user profiles," *J. Korean Inst. Intell. Syst.*, vol. 21, no. 5, pp. 609-617, 2011.
- [3] H. J. Ahn and J. W. Kim, "Comparison of product and customer feature selection methods for content-based recommendation in internet storefronts," *J. KIPS*, vol. 13-D, no. 2, pp. 279-286, 2006.
- [4] S. I. Lee and S. Y. Lee, "A collaborative filtering-based recommendation system with relative classification and estimation revision based on Time," *J. KIIS*, vol. 20, no. 2, pp. 189-194, 2010.
- [5] Y. J. Lee and K. J. Kim, "Product recommender systems using multi-model ensemble technique," *J. KIIS*, vol. 19, no. 2, pp. 39-54, 2013.
- [6] J. S. Kim, Y. A. Do, J. W. Ryu, and M. W. Kim, "A collaborative recommendation system using neural networks for increment of performance," *J. KBS*, vol. 1, no. 2, pp. 233-244, 2001.



- [7] C. H. Shin, J. W. Lee, H. N. Yang, and I. Y. Choi, "The research on recommender for new customers using collaborative filtering and social network analysis," *J. KIISS*, vol. 18, no. 4, pp. 19-42, 2012.
- [8] S. J. Park, S. K. Kang, and Y. K. Kim, "A multimedia contents recommendation system using preference transition probability," *J. KIIS*, vol. 16, no. 2, pp. 164-171, 2006.
- [9] S. Y. Oh, Y. H. Oh, S. H. Han, and H. J. Kim, "Broadcast content recommender system based on user's viewing history," *J. Korean Soc. of Broadcast Engineers*, vol. 17, no. 1, pp. 129-139, 2012.
- [10] K. T. Han, M. K. Park, and Y. S. Choi, "Adaptive and collaborative recommendation using content type," *J. KIISE : Softw. and Appl.*, vol. 38, no. 1, pp. 50-56, 2011.
- [11] B. H. Oh, J. H. Yang, and H. J. Lee, "A hybrid recommender system based on collaborative filtering with selective utilization of content-based predicted ratings," *J. KIISE : Softw. and Appl.*, vol. 41, no. 4, pp. 289-294, 2014.
- [12] J. H. Lee, J. P. Hwang, and E. T. Kim, "A new kernelized approach to recommender system," *J. KIIS*, vol. 21, no. 5, pp. 624-629, 2011.
- [13] D. S. Park, "Improved movie recommendation system based-on personal propensity and collaborative filtering," *KIPS Trans. Comput. and Commun. Syst.*, vol. 2, no. 11, pp. 475-82, 2013.
- [14] N. K. Kim and S. Y. Lee, "Bayesian network based music recommendation system considering multi-criteria decision making," *The J. Digital Policy & Management*, vol. 11, no. 3, pp. 345-352, 2013.
- [15] T. Ueta, M. Iwakami, and T. Ito, "Implementation of a goal-oriented recipe recommendation system providing nutrition information," in *IEEE Int. Conf. TAAI*, pp. 183-188, Chung-Li, Nov. 2011.
- [16] H. M. Kim, S. M. Rho, and J. K. Hong, "Design and implementation of fuzzy-based menu recommendation system," *J. Korea Navig. Inst.*, vol. 16, no. 6, pp. 1109-1115, 2012.
- [17] A. S. Khan and A. Hoffmann, "Building a case-based diet recommendation system without a knowledge engineer," *Artificial Intell. Med.*, vol. 27, no. 2, pp. 155-179, 2003.
- [18] H. K. Han and U. H. Suh, "Learning based personalized foods recommendation agent," in *Proc. 32th KIPS Conf.*, vol. 16, no. 2, pp. 313-314, 2009.
- [19] M. Phanich, P. Pholkul, and S. Phimoltares. "Food recommendation system using clustering analysis for diabetic patients," in *IEEE.Int. Conf. Inf. Sci. and Appl.(ICISA)*, pp. 1-8, Seoul, 2010.
- [20] J. S. Lee and J. C. Lee. "A case based music recommendation system using context-awareness," *J. KIISS*, vol. 12, no. 3, pp. 111-126, 2006.
- [21] S. I. Lee and S. Y. Lee, "A recommendation system using context-based collaborative filtering," *J. KIIS*, vol. 21, no. 2, pp. 224-229, 2011.
- [22] Y. Oh, A. Choi, and W. Woo, "u-BabSang: A context-aware food recommendation system," *The J. Supercomputing*, vol. 54, no. 1, pp. 61-81, 2010.
- [23] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *J. Machine Learning Research*, vol. 3, pp. 993-1022, 2003.
- [24] S. J. Pyo, E. H. Kim, and M. C. Kim, "Topic modeling based similar user grouping and TV program recommendation for smart," in *Proc. Korean Soc. Broadcast Eng. Conf.*, pp. 116-119, 2012.
- [25] T. M. Cho and J. H. Lee, "Latent keyphrase extraction using LDA model," in *Proc. KIIS Fall Conf. 2014*, vol. 24, no. 2, pp. 125-126, 2014.
- [26] E. H. Kim, S. J. Pyo, and M. C. Kim, "Automatic TV program recommendation using LDA based latent topic inference," *J.*

*Broadcasting Eng.*, vol. 17, no. 2, pp. 270-283, 2012.

[27] C. Y. Park, J. D. Lee, J. H. Park, and J. H. Lee, "Contents recommendation method using LDA clustering based on program synopsis," in *Proc. 2013 Conf. HCI Soc. Korea*, pp. 618-621, 2013.

[28] K. S. Kim, H. K. Jeong, H. J. Lee, and H. J. Kim, "Collaborative filtering using topic models for rating based recommender systems," in *Proc. Korean Inf. Sci. Soc. Conf.*, vol. 39, no. 1B, pp. 381-383, 2012.

[29] K. G. Grunert, K. Brunsø, and S. Bisp, *Food-related life style: Development of a cross-culturally valid instrument for market surveillance*, MAPP Århus, Denmark, Oct. 1993.

[30] K. Brunsø and K. G. Grunert, "Cross-cultural similarities and differences in shopping for food," *J. Business Research*, vol. 42, no. 2, pp. 145-150, 1998.

[31] K. G. Grunert, T. Perrea, Y. Zhou, G. Huang, B. T. Sørensen, and A. Krystallis, "Is food-related lifestyle(FRL) able to reveal food consumption patterns in non-western cultural environments? its adaptation and application in urban China," *Appetite*, vol. 56, no. 2, pp. 357-367, 2011.

[32] B. D. Jeong and K. Y. Jang, "Model of u-distribution with use RFID/USN," *J. Korea Inst. of Maritime Inf. & Commun. Sci.*, vol. 11, no. 10, pp. 1814-1820, 2007.

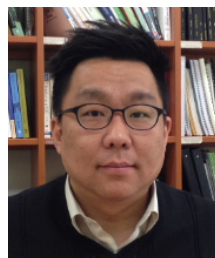
[33] D. Y. Kim, J. E. Lee, S. H. Cha, and K. H. Cho, "Design and implementation of a management framework for ubiquitous home networks," *J. KICS*, vol. 33, no. 5, pp. 275-284, 2008.

[34] H. W. Lee and H. S. Jeong, "Speech recognition system for intelligent home network system," *J. KICS*, vol. 33, no. 4, pp. 162-167, 2008.

[35] K. H. Chon and H. J. Choi, "A study on ubiquitous psychological state recognition model using bio-signals," *J. KICS*, vol. 35, no. 2, pp. 232-243, 2010.

[36] R. Burke, "Hybrid recommender systems: Survey and experiments," *User modeling and User-Adapted Interaction*, vol. 12, no. 4, pp. 331-370, 2002

문 정 훈 (Junghoon Moon)



1996년 2월 : 서울대학교 농학사 졸업  
 1999년 2월 : 서울대학교 교육학 석사 졸업  
 2006년 2월 : 뉴욕주립대학교 경영학 박사 졸업  
 2006년~2010년 : KAIST 경영과학과 조교수

2010년~현재 : 서울대학교 농경제사회학부 부교수  
 <관심분야> 농식품 분야 ICT융합, Internet of Things

장 익 훈 (Ikhoon Jang)



2006년 2월 : 연세대학교 기계전자공학부 졸업  
 2006년~2009년 : 삼성전자 무선사업부 연구원 재직  
 2012년 8월 : 서울대학교 농경제사회학부 석사 졸업  
 2012년 9월~현재 : 서울대학교 농경제사회학부 박사과정

<관심분야> 농식품 분야 ICT융합, 빅데이터분석

최 영 찬 (Young Chan Choe)



1982년 서울대학교 농업교육과 졸업  
 1998년 미시간 주립대학교 농경제학 석사 졸업  
 1991년 미시간 주립대학교 농경제학 박사 졸업  
 1993년 8월~현재 : 서울대학교 농경제학부 교수

<관심분야> 농식품정보화, 전자상거래, 빅데이터분석

김진교 (Jin Gyo Kim)



1991년 : 서강대학교 경영학과  
학사 졸업  
1993년 : 서강대학교 경영학과  
석사 졸업  
2002년 : University of Toronto  
마케팅 박사 졸업  
(前) M.I.T. 교수

(現) 서울대학교 경영학과 교수  
<관심분야> 베이지안 마케팅 모형

박진 (Gene Bock)



1984년 2월 : 연세대학교 기계  
공학과 졸업  
1987년 7월 : University of Iowa  
전산학 수학 학사 졸업  
1989년 5월 : USC 전산학과  
석사 졸업  
2013년 9월~현재 : 연세대학교  
기술정책협동과정 박사과정

2013년 9월~현재 : 농림수산물교육문화정보원, 정보  
보화사업본부장  
<관심분야> 정보화사업, 공공정책평가, 농업정보화