

# 빅 데이터를 활용한 애완동물 상품 추천 시스템 구현

김삼택  
우송대학교 IT융합학부 교수

## Implementation of a pet product recommendation system using big data

Sam-Taek Kim  
Professor, School of Information Technology Convergence, Woosong University

**요약** 최근, 애완동물의 급격한 증가로 애완동물의 건강상태 체크와 다양하게 수집된 데이터를 활용하여 사료 추천 등 통합적인 애완동물관련 개인화 상품 추천 서비스가 요구된다. 본 논문은 빅 데이터 기술을 활용하여 애완동물관련 데이터 수집, 전처리, 분석, 관리등 다양한 개인화서비스를 할 수 있는 상품 추천시스템을 구현한다. 먼저, 애완동물이 착용하고 있는 센서 정보와 고객의 구매 패턴, SNS 정보를 수집해 데이터베이스에 저장하고 통계적 분석을 활용하여 사료제작, 애완동물 건강관리 등 맞춤형 개인화 추천 서비스가 가능한 플랫폼을 구현한다. 본 플랫폼은 유사도가 분석 될 상품과 상품정보에 대한 유사도 상품 정보를 출력하고 최종적으로 추천 분석한 결과를 출력하여 고객에게 정보를 제공 할 수 있다.

**주제어** : 사물인터넷, 추천 시스템, 빅 데이터, 데이터 분석, 데이터 전처리

**Abstract** Recently, due to the rapid increase of pets, there is a need for an integrated pet-related personalized product recommendation service such as feed recommendation using a health status check of pets and various collected data. This paper implements a product recommendation system that can perform various personalized services such as collection, pre-processing, analysis, and management of pet-related data using big data. First, the sensor information worn by pets, customer purchase patterns, and SNS information are collected and stored in a database, and a platform capable of customized personalized recommendation services such as feed production and pet health management is implemented using statistical analysis. The platform can provide information to customers by outputting similarity product information about the product to be analyzed and information, and finally outputting the result of recommendation analysis.

**Key Words** : IoT, Recommendation System, Big Data, Data Analysis, Data Preprocessing

### 1. 서론

사물인터넷기술 및 빅 데이터기술을 활용하여 애완동물의 사료 추천 및 건강관리를 실시간으로 수행하는 기

기가 구현되고 있으며 수요 또한 증가하고 있지만, IoT 기기들은 간단한 센서 정보만을 활용하여 복잡한 처리와 애완동물의 건강상태를 파악하기 어렵고 빅 데이터 기술을 활용하여 맞춤형 추천 서비스 등의 고급기능을 처리

\*This research is based support of 2020 Woosong University Academic Research Funding.

\*Corresponding Author : Sam-Taek Kim(stkim@wsu.ac.kr)

Received September 25, 2020

Accepted November 20, 2020

Revised October 22, 2020

Published November 28, 2020

하는데 어려움이 있다.

이러한 문제를 해결하기 위해서는 다양하게 수집된 정보들을 빅 데이터 기술을 활용해 여러 방향에서 분석하여 개인화 추천서비스를 할 수 있는 애완동물 관련 데이터분석 플랫폼의 구현이 시급하다[1]. 따라서 본 논문에서는 사물인터넷기기의 센서 정보와 고객의 구매패턴 정보, 소셜미디어 정보 등을 수집하여 체계적으로 데이터베이스화하고, 데이터의 연관관계를 복합적으로 분석하여 정보 이용자 개인별로 맞춤 서비스를 수행 할 수 있는 플랫폼을 구현하여 애완동물의 개인화 추천 서비스에 활용하고자 한다.

## 2. 애완동물 개인화 추천 시스템 개념

수집된 애완동물의 다양한 데이터를 기반으로 통계적인 사례를 통한 특성과 비슷한 성질을 분석하여 개인에게 특화된 맞춤서비스를 제공하는 통합 플랫폼을 구현하였다. 애완동물 개인화 맞춤서비스 통합 플랫폼의 구성은 다음 Fig. 1과 같이 데이터 수집하여, 미리 처리하고, 분석과, 관리를 수행하는, 데이터 수집과 분석 및 개인화맞춤 서비스, 데이터 공유, 시각화 모듈로 구성된다.

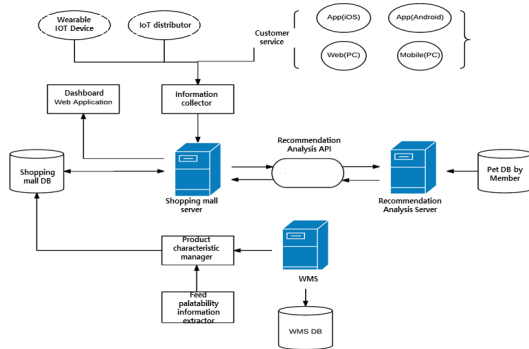


Fig. 1. Integrated platform of Pet customized service

고객 애완동물 정보 수집기는 다양한 플랫폼내에서 회원에 가입할 때, 무게 알려지 종 품종 나이 등 애완동물 관련 정보를 입력하게 한다.

고객 행동 데이터 수집기는 서비스 플랫폼 내에서 애완동물 보호자를 대상으로 어떤 행동을 하는지 분석하기 위해, 카테고리 및 상품 클릭정보와 검색정보 등을 수집한다.

고객 성향 추천 분석기는 분석을 효과적으로 처리하기

위해 협업필터링 알고리즘을 채택하였으며, 사용자의 피드백 상황과 데이터 수집 등을 고려할 때 명시적 피드백 수집이 어렵기에 암시적 피드백을 이용하였다[2-4].

## 3. 애완동물 개인화 추천 서비스 통합 플랫폼

### 3.1 애완동물 데이터 수집기(미들웨어)

애완동물 수집기는 사용자 정보 또는 활동로그 등을 수집할 수 있는 쇼핑물 사용자 정보 수집 모듈, SNS로부터 사용자에게 해당하는 비정형 데이터를 수집하는 로봇인 크롤러를 이용한 사용자 SNS 정보 수집 모듈, 사용자 착용 또는 사물인터넷기기로부터 센서 정보를 수집할 수 있는 IoT 센서 정보 수집 모듈이 존재한다. 소비자 패턴 분석 기반마련을 위한 정형, 비정형 데이터 수집기를 구현 했고 플랫폼 안에서 회원 가입 시, 애완동물 관련 정보를 입력하게 하고 정확한 의학적 사실을 바탕으로 기능을 구현하기 위해, 수의학과 교수 및 동물병원에 자문을 구하여 구현하였다.

### 3.2 데이터 전 처리기

입력된 데이터 중 일부 결손된 데이터에 대해 전역상수, 속성 평균값, 표본값의 평균값 등을 적용하여 데이터를 복구하고 데이터에 포함된 오류 또는 잡음이 포함된 경우 구간화, 단순 또는 복합 회귀값, 군집화를 적용하여 데이터 정제하며 데이터의 일관성 결여, 입력자 오류 등을 확인하여 데이터 불일치를 정제 처리하는 데이터 정제 모듈을 구현했다[5-7]. 또한 다양하게 수집된 데이터들에 대해 메타데이터, 상관성 분석, 데이터 충돌 탐지 등의 통합작업을 통해 데이터들을 결합하는 데이터 통합 모듈을 구현했고, 중복 및 필요 없는 데이터가 포함되어 있을 경우 웨이블릿 변환 신호처리 기법 또는 주성분분석 알고리즘을 활용하여 데이터 축소 모듈을 구현했다 [8,9].

### 3.3 애완동물 데이터 분석기

추출된 의미 있는 데이터로부터 아이템과 연관성 및 분류, 소비자의 예상 행동 패턴 등을 예측하여 개인별 맞춤형 서비스를 제공하기 위한 데이터 분석기는, 데이터가 가지고 있는 상호 연관성을 추출하여 상관도를 측정하는 데이터 연관성 분석 모듈과, 분류규칙이나 의사결정트리, 수학공식의 학습모델을 선택하고, 감독과 무감독학습을

활용하여 예측 정확도를 측정할 수 있는 데이터 분류 모듈로 구성되어 있다. 또한, 입력된 과거 데이터로부터 설명변수와 예측변수 사이의 관계를 파악하여 사용자의 미래 행동예측과 경향을 분석 하고, 사용자를 분류하거나 상품배합을 분할 기법의 하나인 k-means 알고리즘을 활용하여 분할을 수행하는 데이터 군집화 모듈이 있다.[10].

### 3.4 개인화 서비스 대시보드

회원정보, 클릭정보, IoT착용형, IoT배급기에 대한 정보 수집 및 저장은 쇼핑몰 DB에 저장되고 대시보드 웹/애플리케이션의 경우, 내부 서버에 저장되어 있으며 추후에 활발한 분석활동을 할 예정이므로, 서버 리소스가 많이 필요하다.

#### 3.4.1 회원정보, 클릭정보 대시보드

회원정보 대시보드는 쇼핑몰 서비스에 대한 회원들의 통계 분석을 위한 대시보드로 주로 회원 수에 대한 정보를 모니터링하며 기간을 설정하여 특정 기간 동안 발생 하였던 데이터를 그래프로 확인하여, 현재 추세를 분석할 수 있다.

클릭정보 대시보드는 특정시기(또는 현재)의 클릭횟수 <Bar형 차트>와 모니터링특정시기(또는 현재)동안 상품에 대한 클릭순위 <Bar형 차트>모니터링, 실시간 클릭정보 전체클릭 비율, 회원클릭 비율 <Pie형 차트> 모니터링으로 구현했다.

#### 3.4.2 IoT 착용형, 배급형 대시보드

연동된 착용형 IoT기기의 각 센서에서 발생하는 값을 모니터링하는 기능이고 센서의 종류에 따라 발생하는 값의 종류도 다르므로, 센서의 종류를 선택하고 선택한 회원 센서(심전도, 가속도 등)의 실시간 센서값 데이터 그래프 모니터링 결과로 해당 센서의 통계데이터가 표시된다.

### 3.5 SNS 정보 수집기

트위터에서 제공하는 API를 통하여, 해당 키워드에 대한 글 내용, 일시, 등록번호 등을 수집한다. 회사에서 판매하고 있는 '브랜드'를 키워드로 해당 브랜드에 대한 정보를 수집하고 있으며, 담당자는 수집되는 글 내용을 분석하여, 브랜드에 대한 현재 고객 인식에 대해 평가를 하여 기록한다. 트위터에서 제공하는 여러 API중, 현재 상

황에 맞는 API는 <Search API>이며, API를 사용하기 위해 트위터 사이트에 구현자 계정을 등록해야한다. 또한 트위터 API는 대부분 OAuth 인증이 기본이므로, 해당 토큰을 확인하여, 통신패킷의 헤더에 토큰을 실어야 결과 값을 가져올 수 있고 쇼핑몰 서버에 <SNS 정보 수집기>를 Restful 방식 API로 구현하였으며, 해당 URL을 실행하면 지정된 기능이 실행되고, Search API에서 확인되는 글에 대한 정보를 모두 JSON형식으로 결과 데이터를 받아와, 파싱을 한 뒤 <쇼핑몰 DB>의 <collect\_sns\_data> 테이블에 저장되어 관리하게 된다.

트위터 SNS 수집처리에 대한 구조도는 다음 Fig. 2와 같다.

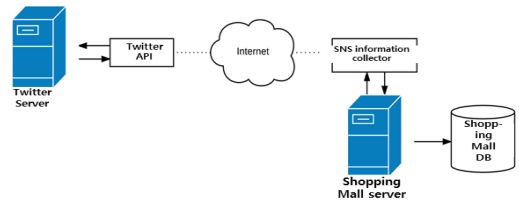


Fig. 2. Structure of Twitter SNS Collection Process

### 3.6 개인화 서비스 고객성향 추천 분석기

개인화 맞춤 서비스를 제공하기 위해 사용자별 아이템 추천 서비스와, 쇼핑몰에서 행동패턴을 분석하여 탈퇴 또는 활동 지속여부 등 위험요소를 분석하는 사용자 행동 예측 서비스, 쇼핑몰에서 분석된 특정 자료를 요청할 경우 자료를 제공할 수 있도록 분석정보 검색 서비스와 쇼핑몰 사용자에게 필요한 관계기관을 연계시켜 적시에 서비스를 지원받을 수 있도록 쇼핑몰에 관련 정보를 제공하는 관계기관 연계 서비스를 구현 했다[11].

추천분석을 위한 알고리즘은, 다음과 같이 암시적 피드백을 이용하였다.

- $r_{ui}$  는 고객의 클릭의 수, 고객의 상품의 수
- $p_{ui}$  는 고객의 해당 상품을 누를 확률(신뢰)를 의미한다.

$$p_{ui} = \frac{1 \text{ if } r_{ui} > 0}{0 \text{ if } r_{ui} = 0}$$

사용자  $u$  가 항목  $i$  ( $r_{ui} > 0$ )를 소비하면, 사용자  $u$  가 항목  $i$   $p_{ui} = 1$  좋아한다고 표시한다.  $r_{ui}$ 가 커지면 사용자는 실제 항목을 좋아한다고 강한 표시를 나타나게 된다.  $c_{ui}$ 는  $p_{ui}$ 에 대한 신뢰  $c_{ui} = 1 + \alpha r_{ui}$  이다. 모든 사용자와 항목 간의  $p_{ui}$ 에 대해 최소한의 신뢰는 갖지만 좀 더 긍정 선호를 가지면  $p_{ui} = 1$ 에 대한 신뢰도는 그에 따라 증가한다.

증가율은 상수  $\alpha$ 에 의해 제어된다. 각 사용자  $u$ 에 대한 벡터  $x_u \in R^f$ 와 사용자 선호와 사용자 선호도를 고려할 각 항목  $i$ 에 대한 벡터  $y_i \in R^f$ 를 찾는다[12].

$$\min_{x_u, y_i} \sum_{u,i} c_{ui} (p_{ui} - x_u^T y_i)^2 + \lambda (\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2)$$

$\lambda (\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2)$  항은 모델로 정규화 하여 훈련 데이터에 과부하가 걸리지 않도록 하는데 필요하고, 매개 변수  $\lambda$ 의 정확한 값은 데이터 종속적이며 상호 유효성 검사에 의해 결정한다. 해당 공식으로 “애완동물 종 정보를 가진 회원들의 클릭정보”를 이용하여, 추천모델링을 만들어 관리한다. 추천분석 서버 안에, 추천모델링에 대한 정보를 관리하며, 고객성향 추천분석기는 해당 추천모델링 데이터를 이용하여 추천결과를 제시한다. 추천분석기는 입력으로는 다음 인자 값을 받는다[13-15].

① m\_no : 회원번호

② analysis\_cnt : 추천 상품리스트 개수

m\_no로 인자 값을 넣으면, 해당 회원에 맞는 추천 상품리스트를 analysis\_cnt만큼 출력으로 리턴 한다. 쇼핑몰 서버와 추천분석 서버는 물리적/논리적으로 떨어져있으므로, 쇼핑몰서버에서 추천분석에 대한 결과를 유연하게 처리하기 위해서는 API가 필요하다. API는 RestFul API로, 추천분석서버에서 오픈한 웹서비스의 접근방식으로 구현되었고 결과 값은 쉬운 파싱을 위해, JSON 형식으로 반환을 받는다.

{u'recommend':

{u'learning\_contents': u'  
12952\t541\t0.346042408619\t

..}

learning\_contents 안에 있는 내용이 실제 추천된 결과이며, 파싱 방법은 토큰 '\t'를 기준으로 한다. 12952\t541\t0.346042408619\t 에서

'\t'토큰을 기준으로, 파싱을 하였을 때,

① 12952 ② 541 ③ 0.346042408619로 데이터를 구분할 수 있으며, 1번은 추천결과 회원번호이며, 2번은 추천된 상품이며, 쇼핑몰 Server에 저장되어있는, 541번 상품이다. 3번의 결과는 해당 상품을 누를 확률에 대한 신뢰도이며, 최대 신뢰도는 1.0값을 가지므로 34.6%의 확률로 계산된다. 다음 그림 Fig. 3은 회원에 대한 추천 분석 과정에 대한 구조도이다.

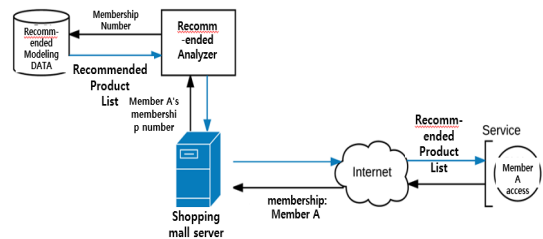


Fig. 3. Structure diagram of recommendation

### 3.7 추천 분석기의 군집분석

추천시스템에 활용할 수 있도록 추천 분석기의 군집분석을 위해 Table 1과 같이 애완동물의 사료 중 7대 중요 영양소를 분석하여 군집화, 정규화, 범위를 벗어난 데이터는 정제 하였다.

Table 1. Feed for measuring the similarity of the recommendation system (example)

ID	Name of product	ID	Name of product
v1	[Pet Botanics] GRAIN-FREE Omega-Rich Chicken	v36	ANF LAMB 28 Premium15
v2	[Pet Botanics] GRAIN-FREE Omega-Rich Salmon	v37	ANF LAMB HOLISTIC MB 15
v3	[NUTRIO] JUNIOR DOG	v38	ANF Organic CHICKEN_SB_6
v12	---	v47	---

다음으로 군집분석을 위해 사료의 특징 간 유사도를 측정하였다. 유사도 분석은 Pet\_type 인자 값을 바탕으로 강아지 또는 고양이 DB의 1군 특성(주요 영양소), 2군 특성(특이 영양소)을 로드하고 해당 로드한 값으로, 유사도 공식으로 분석하여 1군 분석 결과 값과 2군 분석 결과 값을 DB에 삽입하고 최종 유사도 분석 결과 값을 유사도 결과 DB에 삽입하여 유사도 분석결과를 상품 추천 시 활용한다. 즉, 유사도 분석은 해당 상품리스트에서 각 상품별 주요영양소(feed\_spec1\_x), 특이영양소(feed\_spec2\_x)를 가져와, 유사도 공식에 의거하여 분석 한다 . 유사도 측정 알고리즘은 Cosine distance, Manhattan Distance를 이용하였고 군집 분석 결과의 계통도를 그리면 다음 그림 Fig. 4와 5 와 같다. 코사인 유사도는 평점을 벡터로 생각하여, 2개 벡터사이의 각도가 작을수록 근접해 있기 때문에 유사성이 크다고 판단 하는 방식이다.

$$sim(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| * |\vec{b}|} \text{ (Cosine Similarity)}$$

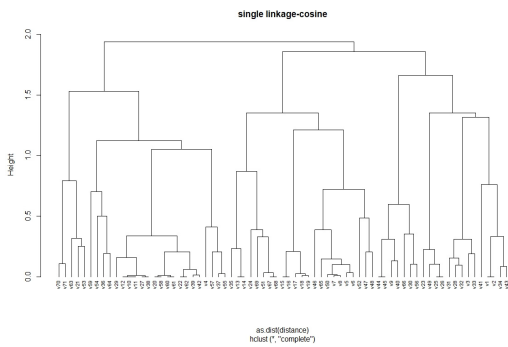


Fig. 4. Similarity measurement algorithm Cosine distance

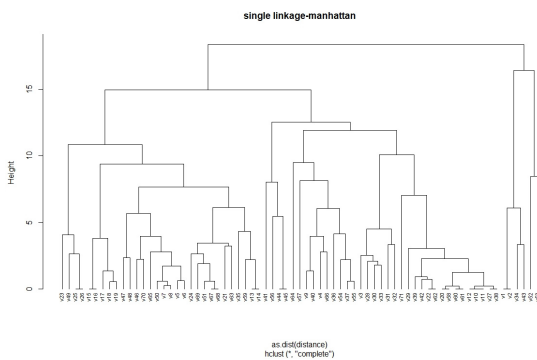


Fig. 5. Similarity measurement algorithm Manhattan Distance



Fig. 6. Similarity product information dashboard



Fig. 7. Dashboard that exposes recommendation analysis and recommendation analysis time

Fig. 6은 분석될 상품정보에 대한 유사도 상품 정보 대시보드이다. 성분 간 유사도 분석 값이 가장 높은 순으로 노출된다. Fig. 7은 최종적으로 추천 분석 및 추천 분석 시간을 노출한 대시보드 결과이다.

#### 4. 결론

본 논문은 빅 데이터 기술을 활용하여 다양한 IoT의 센서, 쇼핑몰 고객의 구매 패턴, SNS등의 정보를 수집할 수 있는 미들웨어와 수집된 정보를 정제하고 분석 및 추론 가능한 분석기 및 분석된 정보를 이용하여 고객에게 상품 추천 및 의약품 제작 등의 애완동물의 건강관리에 활용할 수 있는 맞춤형 개인화서비스를 구현했다. 본 플랫폼은 유사도가 분석될 상품과 상품정보에 대한 유사도 상품 정보를 출력하고 최종적으로 추천 분석한 결과를 보여 주게 된다. 본 논문의 결과를 통해 고객의 성향에 맞는 개인화 서비스를 추진하고, 애완동물 삶의 전주기적 건강관리를 통해 애완동물 시장을 활성화시킬 수 있으며, 애완동물에 대한 데이터 구축 및 분석을 통해 다양한 산업에서 활용 가능하다. 구현한 통합 플랫폼을 활용하여 반려동물 쇼핑몰을 운영하는 모 회사에서 반려동물에게 맞춤형 사료제작 및 의약품제작 등의 개인화서비스 추진하고 고객 니즈에 대한 선제적인 서비스 제공을 통해 중소기업의 경쟁력 확보했다. 향후 연구할 부분으로는 사용자의 구매정보와 애완동물의 사료 하루 섭취량을 병합하여 분석한다면 애완동물의 헬스케어시스템, 추천시스템 등에 고도화된 서비스가 가능할 것이다.

#### REFERENCES

- [1] C. Gurrin, A. F. Smeaton & A. R. Doherty. (2014). LifeLogging: Personal Big Data. *Foundations and Trends in Information Retrieval*, 8(2), 1-107.
- [2] S. Niwattanakul, J. Singthongchai, E. Naenudorn, & S. Wanapu. (2013). Using Jaccard Coefficient for Keywords Similarity. *Proceedings of the International MultiConference of Engineers and Computer Scientists*, 1(2202), 380-384.
- [3] G. Karypis. (2001). Evaluation of Item-Based Top-N Recommendation Algorithms. *Proc. of CIKM '01 Conference*, 247-254.
- [4] J. L Herlocker, J. A. Konstan, L. G. Terveen & J. Riedl. (2001). Evaluating Collaborative Filtering Recommender Systems. *ACM Transactions on Information Systems*,

22(1), 5-53.

- [5] J. Horey, E. Begoli, R. Gunasekaran, S. Lim & J. Nutaro (2012). Big Data Platforms as a Service: Challenges and Approach. *USENIX Workshop on Hot Topics in Cloud Computing (HotCloud)*. <https://www.usenix.org/system/files/conference/hotcloud12/hotcloud12-final61.pdf>
- [6] S. S. Weng & M. J. Liu. (2004). Feature-based recommendations for one-to-one marketing. *Expert Systems with Applications*,26(4),493-508.
- [7] E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen & X. Sun. (2010). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50, 559-569.
- [8] T. C. Fu. (2011). A Review on Time Series Data Mining. *Engineering Application of Artificial Intelligence*, 24(1), 164-181.
- [9] Y. Zhu, Y. Fu & H. Fu. (2010). A New Class of Attacks on Time Series Data Mining. *Intelligent Data Analysis*, 14(3), 405-418.
- [10] R. Agrawal & R. Srikant. (2000). Privacy-Preserving Data Mining. *In Proc. of Conf. on Management of Data, ACM SIGMOD, Dallas, TX*, 439-450.
- [11] S. T. Kim. (2016). Design of Convergence Platform for companion animal Personalized Services. *Journal of the Korea Convergence Society*, 7(6), 29-34. DOI : 10.15207/JKCS.2016.7.6.029
- [12] S. M. Bea & V. Torra. (2011). Trajectory Anonymization from a Time Series Perspective. *In Proc. of IEEE Int'l Conf. on Fuzzy Systems, Taipei, Taiwan*, 401-408.
- [13] S. H. Namn & K. S. Noh. (2015). A Study on the Effective Approaches to Big Data Planning. *Journal of digital Convergence*, 13(1), 227-235.
- [14] K. S. Noh. (2015). Convergence Analysis of Recognition and Influence on Big data in the e-Learning Field. *Journal of digital Convergence*, 13(10), 51-58.
- [15] H. Jung. (2015). The Analysis of Data on the basis of Software Test Data. *Journal of digital Convergence*, 13(10), 1-7.

김 삼 택(Sam-Taek Kim)

[정회원]



- 2005년 2월 : 중앙대학교 중앙대학원 컴퓨터공학과(공학박사)
- 1990년 5월 ~ 1995년 2월 : LG연구소 전임연구원
- 1995년 3월 ~ 현재 : 우송대학교 IT 융합학부 교수
- 관심분야 : 무선/유선 네트워킹, VoIP,

모바일 컴퓨팅, IoT, Big Data, Blockchain

· E-Mail : stkim@wsu.ac.kr