

## 영양성분 프로파일링 기반 사료추천 알고리즘

송희석\*

### Nutrient Profiling-based Pet Food Recommendation Algorithm

Hee Seok Song\*

#### Abstract

This study proposes a content-based recommendation algorithm (NRA) for pet food. The proposed algorithm tries to recommend appropriate or inappropriate feed by using collective intelligence based on user experience and prior knowledge of experts. Based on the physical and health status of the dogs, this study suggests what kind of nutrients are necessary for the dogs and the most recommended pet food containing these nutrients. Performance evaluation was performed in terms of recall, precision, F1 and AUC. As a result of the performance evaluation, the AUC and F1 value of the proposed NRA was 15% and 42% higher than that of the baseline model, respectively. In addition, the performance of NRA is shown higher for recommendation of normal dogs than disease dogs.

Keywords : Recommender, Pet Food, Content-Based Recommendation Algorithm, Nutrient Profiling

Received : 2018. 11. 28.    Revised : 2018. 12. 15.    Final Acceptance : 2018. 12. 26.

※ This work was supported by 2018 Hannam University Research Fund.

\* Professor, Department of Global IT Business in Hannam University, 133, Ojungdong, Daedukgu, Daejeon City,  
e-mail: hssong@hnu.kr

## 1. 서 론

최근 반려인 가구가 급증하면서 반려견을 위한 사료 추천서비스에 대한 필요성이 강하게 제기되고 있다. 반려견 가구의 증가와 함께 사료의 종류도 급증하게 되면서 반려인들은 사료급여에 늘 민감하지만, 해당사료가 반려견에게 적합한 사료인지 판단하기가 어려워 브랜드 인지도만으로 사료를 구매하는 것이 일반적이다. 또한 반려견은 품종이 다양하며 연령이나 크기, 신체 상태, 질병 등에 따라 섭취해야 할 영양성분이 달라지기 때문에 반려견에게 필요한 영양성분을 포함한 사료를 선택하는 것은 복잡한 의사결정과정이라 할 수 있다. 한편, 추천시스템은 전자상거래를 위한 제품 및 서비스의 추천, 소셜 네트워크에서 친구 추천 및 태그 추천 등 다양한 분야에서 성공적으로 적용되고 있다. 특히 4차 산업혁명의 핵심기술인 빅데이터와 인공지능 기술에 관심이 집중되면서 모바일을 기반으로 한 공공서비스 추천에서부터 맛집이나 뉴스, 기호제품의 추천 등 추천서비스의 활용범위가 보다 광범위해지고 있다. 이에 본 연구에서는 반려견의 신체 상태나 건강상태를 토대로 해당 반려견에게 필요한 영양성분이 무엇인지 판단하여 이들 영양성분을 가장 충실히 포함하는 사료를 추천하는 알고리즘을 제안하고 성능평가를 통해 실제 적용가능성을 확인해 보기로 한다. 현재 인터넷 상에 많은 사료 쇼핑몰이 존재하고 있으나 대부분의 쇼핑몰은 반려견 특성에 맞춤형된 사료를 추천하는 기능을 제공하지 못하고 있으며 사료특성별 조건 검색 기능을 제공하거나 인기 사료를 추천하는 수준에 그치고 있다. 조건검색에 의한 사료정보 제공이나 인기사료를 추천하는 방식은 반려견의 신체 상태나 질병상태를 고려하지 못하고 있기 때문에 맞춤형 추천이 이루어지지 못한다는 한계점이 존재한다.

반려견을 위한 사료추천은 기존 추천시스템과 비교하여 다음과 같은 몇 가지 설계 이슈를 갖는다. 첫째, 기존의 추천시스템이 주로 사용자의 선호도에 기반하여 추천이 이루어지는데 비해 사료추천시스템은 특정 사료가 반려견의 신체 및 건강상태에 적합한 영양성분을 갖추고 있는지와 같이 적합도에 기반한 추천이 이루어져야 한다. 또한 사용자가 추천의 단위가 되는 기존 추천시스템과 달리 사료추천시스템에서 추천의 단위는 품종, 색상, 연령, 질병상태 등이 유사한 반려견 그룹이 되기 때문에 반려견 특성정보가 추천시스템을 구성하는 추가적인 요소가 된다. 둘째, 사료 급여 시 특정 사료가 해당 반려견의 신체 상태나 건강상태에 도움이 되는 경우가 있는가하면 오히려 설사와 같은 부작용을 일으키기도 한다. 따라서 사료 추천 시 적합도 외에도 부적합도를 고려하여 추천이 이루어져야 한다. 셋째, 사료추천시스템에 사용되는 학습데이터는 기존 추천시스템과 마찬가지로 사료 급여에 대한 후기평점 정보가 될 수 있다. 그러나 사용자 후기 외에도 수의사와 같은 반려견 전문가의 지식이 사료추천에 반영된다면 보다 정확한 추천이 이루어질 수 있다. 따라서 사용자 후기정보와 전문가의 선형적 지식을 어떻게 결합하여 사료추천시스템을 구성할 것인가가 또 다른 설계의 이슈가 된다. 상기의 특성들로 인해 반려견을 위한 사료추천시스템은 협업필터링과 같은 기존 추천 알고리즘을 직접 적용하여 개발하는 것은 곤란하다.

본 연구에서는 이상의 설계 이슈를 해결할 수 있는 사료추천 알고리즘을 제안하는 것을 목표로 한다. 이를 위해 Ueda et al.[2011]이 조리법 추천에서 사용한 콘텐츠기반 추천방식을 확장하여 영양성분 프로파일링 기반의 사료추천 알고리즘을 제시한다. 보다 구체적으로 사용자가 자신의 반려견에 대한 특성정보를 입력하면 해당 반려견과 유사한 반려견들에게 급여된 사료급여 후기와 전문가의 지식을 결합하여 해당 반려견

에게 필요한 영양성분을 발견하며, 이들 영양성분을 가장 적절하게 포함하고 있는 사료를 추천하는 방법을 제안한다. 반려견을 위한 사료추천은 반려견 시장의 급성장과 다수 반려인들의 요구에도 불구하고 아직 시도된 적이 없다. 본 연구의 구성은 다음과 같다. 먼저 제2장에서는 반려견 사료추천 도메인과 가장 유사한 요리 조리법 도메인의 추천시스템에 대한 문헌들을 조사한다. 제3장에서는 반려견을 위한 사료추천 알고리즘을 제시하고, 제4장에서는 실험에 사용한 데이터 셋과 성능평가 방법 및 성능평가 결과를 요약한다. 마지막으로 제5장에서는 제시한 알고리즘을 실제 상용화하기 위한 전략을 소개하고 향후 연구방향을 제시한다.

## 2. 기존연구

추천서비스는 빅데이터 및 인공지능 기술의 보급과 함께 제품 및 서비스의 추천, 친구 추천 및 태그 추천 뿐 아니라 공공서비스 추천에서부터 맛집이나 뉴스, 기호제품의 추천 등 다양한 분야로 확산되고 있으나[Rendle et al., 2010; Sato et al., 2015; Wang et al., 2011] 반려견을 위한 사료추천서비스는 아직 상용화된 사례를 찾기 어렵다. 따라서 여기서는 사료추천과 가장 유사한 분야인 조리법(Recipe) 도메인에서 추천시스템 관련 연구들을 살펴보고 사료추천시스템 설계에 필요한 아이디어를 도출하기로 한다 [Almeida, 2015].

추천시스템은 추천방식에 따라 일반적으로 지식 기반 추천시스템, 콘텐츠 기반 추천시스템, 협업 필터링 기반 추천시스템과 이들을 결합한 하이브리드 추천시스템으로 구분된다. 협업 필터링 기반 추천시스템은 가장 많은 분야에서 적용되는 방식으로 유사 사용자의 선호도 평가내역에 기반하여 추천이 이루어지는 방식이다. 즉 선호하는 아이템이 유사한 사용자를 발견하여 유사사용자가 구매

한 아이템 중 해당 사용자가 구매하지 않은 아이템을 추천하는 방식이다. 콘텐츠 기반 추천방식은 고전적인 정보검색(Information Retrieval)에 기반한 방법으로 사용자가 선호하는 아이템을 설명하는 문서로부터 키워드를 추출하여 사용자 프로파일을 구성하고 이 프로파일과 유사한 문서(아이템)를 추천하는 방식이다. 지식기반 추천시스템은 사용자의 니즈와 선호 체계에 대한 추론을 바탕으로 제품 또는 서비스를 추천하는 방식이다. 또한 하이브리드 추천시스템은 콘텐츠 기반 추천과 협업 필터링 기반 추천의 장점을 결합하여 구성한 시스템이다[Almeida, 2015]. 각 방법은 도메인 특성에 따라 다르게 선택되지만, 512명 사용자로부터 43,000개의 평가 데이터를 기반으로 실험한 Freyne and Berkovsky[2010]의 연구에서 언급된 것처럼 조리법 추천분야에서는 콘텐츠 기반 추천방법의 추천정확도가 협업필터링 기반 추천방법에 비해 높은 것으로 보고되는 경우가 많다.

먼저 Ueda et al.[2011]은 사용자의 선호도를 이용한 콘텐츠 기반의 개인화된 조리법 추천시스템을 제안하였다. 이 시스템에서는 사용자가 인터넷에서 열람했던 조리법 또는 이미 요리를 한 이력을 바탕으로 사용자 선호재료 프로파일을 작성하고 이 프로파일과 가장 유사한 조리법을 추천하게 된다. 사용자 선호재료 프로파일은 조리법에 사용된 재료들을 키워드로 구성된 벡터모델을 사용하는 데, 각 재료에 대한 선호 가중치는 TF-IDF를 응용한 FF(성분빈도)-IRF(역조리법빈도)를 사용하고 있다. 또한 Almeida[2015]는 식당별 조리법 추천을 위해 YoLP 시스템을 제안하였다. YoLP는 콘텐츠 기반 추천시스템으로 사용자 프로파일과 조리법 특징벡터들을 비교하여 유사도 순으로 추천을 수행한다. 이 때 조리법 특징벡터는 단순히 어휘벡터(Word vector)가 아니라 요리를 구성하는 특징벡터로 요리 카테고리, 지역, 식당 ID, 재료성분 등 다양한 정보로 구성된다.

한편 Lin et al.[2014]는 조리법 추천이 요리절차와 성분에 대한 다양한 이질적인 정보를 포함하고 있기 때문에 기존에 많이 시도된 음악, 영화, 책 등의 아이템 추천과는 다른 특성을 가진다고 하였다. 이 연구에서는 조리법을 재료성분, 카테고리, 요리법, 영양성분 등으로 부터 추출된 각종 특징들의 집합체로 정의하고 행렬분해 방법(Matrix Factorization)을 이용한 추천방식을 제안하였는데, 행렬분해는 아이টে에 대한 사용자의 평가가 잠재된 특징에 기반 한다고 보고 이러한 잠재된 특징을 발견하여 미지의 아이টে에 대한 평가치를 예측하는 것이다. 이 연구에서는 시간의존적인 특성을 포함하여 추천모형을 구성했다는 점이 특징이다. Freyne and Berkovsky[2010]는 음식 및 식이요법 영역에서 사용자의 음식 선호도에 따라 다이어트와 라이프 스타일을 감안한 개인화된 콘텐츠 기반의 조리법 추천기를 제안하였다. 이 연구에서는 사용자로 하여금 조리법에 대해 평가하게 하고 조리법별 평가 내용을 성분별 평가로 변환한 후 사용자별 성분 선호도 평가치를 도출하였으며 간단한 인공지능 기법을 도입함으로써 추천의 정확도를 높일 수 있음을 보여주었다. 구체적인 추천방법은 사용자가 특정 조리법을 5점 만점에 5점만큼 좋아했다면 해당 조리법의 성분을 전개한 후 각 성분에 5점을 부여하는 방식으로, 모든 사용자가 평가한 조리법에 대해 성분평가점수를 합산한 후 평가횟수로 나누어준 것이 사용자가 특정성분을 얼마나 좋아하느냐의 값이 된다. 결국 사용자가 미지의 조리법을 얼마나 좋아할지는 이 조리법이 보유한 성분에 대한 평가치의 합산이 되는 셈이다. Forbes and Zhu[2011]는 조리법 추천과정에서 행렬분해 기반의 협업필터링방법이 초기사용문제(Cold Start)가 있으므로 조리법별 성분행렬과 같은 기지의 콘텐츠를 행렬분해과정에 통합함으로써 추천의 정확도를 높일 수 있다는 것을 보여주었다. 행렬분해에서 기지의 콘텐츠 정보를 통합하는 구체적

인 방법은 Lin et al.[2014]의 CTRMF 방법과 동일하다.

조리법 추천은 다양성과 복잡성으로 인해 추천 시스템 연구 중 가장 도전적인 도메인 영역중 하나이다. 조리법 추천 시 사용자 추천에 고려되는 정보는 음식의 내용물 또는 성분 뿐 아니라 요리 방법, 성분비 및 가용성, 조리의 복잡성, 준비 시간, 영양소, 성분 조합 효과 등이며 나아가서는 문화적 사회적 요인들도 고려대상이 된다. 반려견을 위한 사료 추천 또한 조리법 추천과 마찬가지로 반려견 신체 상태나 질병상태, 영양성분 등 다양한 이질적인 정보를 추천에 반영하여야 한다. 또한 조리법 추천이나 사료추천은 음악이나 영화 분야의 선호도 기반 추천과는 달리 아이টে(조리법 또는 사료)의 수는 많은데 비해 평가(Rating)정보의 수가 상대적으로 작기 때문에 사용자-아이টে 선호도 매트릭스에 심각한 희박성(Sparse rating)이 존재한다. 특히 반려견 사료추천의 경우 단지 선호도 평가가 아니라 사료 급여 경험에 대한 평가이기 때문에 한 반려견에게 여러 사료에 대한 급여경험 정보를 수집하는 데는 한계가 존재한다. 따라서 협업필터링보다 콘텐츠 기반 추천이나 하이브리드 추천이 보다 적합한 방법이 될 것이다. 조리법 추천에 대한 기존 문헌 연구에서도 이러한 특성으로 인해 콘텐츠 기반 추천이 대세를 이루고 있으며 실험을 통해서도 콘텐츠 기반추천의 우수성을 확인하고 있다.

### 3. 사료추천 방법

본 연구에서는 반려견 사료추천을 위해 다음과 같은 몇 가지 원칙을 기반으로 추천 알고리즘을 설계하고자 한다. 첫째, 사료추천은 선호도 기반이 아니라 적합도 기반의 콘텐츠 기반 추천방식이 되어야 한다. 물론 반려견별로 좋아하는 향미와 재료가 다를 수 있으므로 선호도 기반의 추천도 고려해 볼 수 있으나 본 연구에서는 반려견의 신체 상태나

건강상태에 적합한 영양성분을 갖춘 사료를 추천하는 것이 더 필요할 것이란 판단에서 적합도 기반의 사료추천 방법을 제안한다. 둘째, 사료 급여 시 특정 사료는 해당 반려견에게 탈을 일으키기도 하므로 추천 시 적합도 외에도 부적합도를 고려하여 추천이 이루어져야 한다. 셋째, 사용자 후기정보와 전문가의 선형적 지식을 결합하여 사료추천이 이루어지도록 한다. 본 연구에서는 이러한 설계 원칙하에 영양성분 프로파일링 기반의 사료추천 알고리즘(NRA: Nutrient profiling-based Recommendation Algorithm)을 제안하기로 한다. 제안한 사료추천 알고리즘의 성능을 비교할 목적으로 베이스라인 모형과 협업필터링 기반 모형을 동시에 구성하여 제시하기로 한다.

### 3.1 베이스라인 모형

베이스라인 모형은 본 연구에서 제안한 알고리즘의 성능을 비교하기 위한 간단한 모형으로 반려견의 견종, 크기, 질병상태, 사료별 영양성분정보 등의 콘텐츠 정보를 사용하지 않고, 반려견에게 특정사료를 급여한 결과 만족했는지 또는 불만족했는지의 정보만을 활용하여 사료의 적합 및 부적합 여부를 예측하는 모형이다. 베이스라인 모형은 반려견 아이디  $u$ , 사료 아이디  $i$ 와 같은 두 개의 카테고리 값을 입력하여 평점  $\hat{r}_{ui}$ (만족 = 1, 불만 = 0)를 예측하는 가장 단순한 모형으로 반려견과 사료 특성에 의한 평균 평점의 합으로 구한다.

$$\hat{r}_{ui} = \mu + b_u + b_i,$$

$\mu$ : 전체평점평균

$b_u$ : 반려견  $u$ 평점조정값

$b_i$ : 사료  $i$ 평점조정값

베이스라인 모형의 오차 함수는 아래와 같이 정규화 항을 추가하여 정의되며, 경사하강법(SGD) 또는 대체최소사승법(ALS)과 같은 최적화 알고

리즘을 적용하여 구현한다[Koren and Sill, 2011; Bengio et al., 1994].

$$\sum_{r_{ui} \in R_{train}} (r_{ui} - (\mu + b_u + b_i))^2 + \lambda(b_u^2 + b_i^2)$$

### 3.2 협업필터링(Collaborative Filtering) 기반 모형

협업필터링기반 모형 또한 베이스라인 모형과 마찬가지로 반려견의 특성정보와 사료의 영양성분과 같은 콘텐츠 정보를 무시하고 오직 반려견에게 사료급여 결과 만족했는지 또는 불만족했는지 여부만으로 예측을 수행하는 방법이다. 협업필터링 기반 모형에서 사용되는 반려견 사료급여 평점 행렬은 빈 원소가 많아 희박도(Sparsity)가 높은 행렬이기 때문에 이를 P와 Q행렬로 분해한 후 재조합하는 방식을 사용하여 희박도 문제를 완화할 수 있으며 이 때 주로 사용되는 방법이 잠재요인(Latent Factor) 모형이다. 사용자의 특성 벡터나 상품의 특성 벡터의 길이는 수천에서 수십억에 달하는 긴 크기가 될 수도 있다. 잠재요인모형은 이렇게 긴 사용자 특성이나 상품 특성을 몇 개의 요인 벡터로 간략화 할 수 있다는 가정에서 출발한다. 여기서는 잠재요인 모형 중 SVD(Singular Value Decomposition)를 사용하여 반려견 급여 사료에 대한 적합성 여부를 예측하기로 한다. SVD적용을 위한 오차함수는 다음과 같다.

$$\sum_{r_{ui} \in R_{train}} (r_{ui} - (\mu + b_u + b_i + q_i^T p_u))^2 + \lambda(b_u^2 + b_i^2 + \|q_i\|^2 + \|p_u\|^2)$$

### 3.3 영양성분 프로파일링 기반 사료추천

알고리즘(NRA: Nutrient Profiling-based Recommendation Algorithm)

본 연구에서 제안하는 사료추천 방법은 사용자정형에 기반한 집단지성과 전문가의 사전지식

을 동시에 활용하여 적합사료 또는 부적합 사료를 추천하는 방법이다. 즉 반려견 커뮤니티 회원 을 대상으로 수집된 사료급여 후기 정보를 토대 로 반려견 특성별로 적합 또는 부적합 영양성분 이 무엇인지를 도출하고, 여기에 전문가의 사료 추천 지식을 결합하여 추천에 활용하는 콘텐츠

기반 추천 방법이다. 영양성분 프로파일링 기반 사료추천 알고리즘은 사료급여 후기정보를 영양 성분으로 분해하는 단계와 추천단위별 프로파일 링 단계 그리고 적합사료 또는 부적합사료 추천 목록 작성단계로 구분되며 알고리즘은 다음과 같다.

**Step 1:** 사용자급여 후기의 적합(부적합)사료제품을 제품에 표시된 영양성분으로 분해하여 각 각의 후기를 반려견 특성별 적합(부적합) 영양성분 정보로 변환한다.

**Step 2:** 영양성분으로 분해된 사료급여 후기 중 사료추천을 요청한 반려견과 동일한 특성의 반려견 에게 급여했던 후기들만을 추출하여 다음식과 같이 해당 반려견에게 적합(부적합)한 영양 성분에 대한 프로파일을 생성한다.

$$x_{ki} = \sum_{j \in POS_i} w_j p_{kj}, y_{ki} = \sum_{j \in NEG_i} w_j p_{kj}$$

$x_{ki}$ :  $i^{th}$  반려견에게 영양성분  $k$  급여시 적합도

$y_{ki}$ :  $i^{th}$  반려견에게 영양성분  $k$  급여시 부적합도

$p_{kj}$ :  $j^{th}$  급여후기에 포함된 영양성분  $k$ 의 함량비율( $0 \leq p_{kj} \leq 1$ )

$POS_i$ :  $\begin{cases} i^{th} \text{ 견종/색상/연령의 급여후기(전문가지식) 중 긍정적 후기} \\ \text{(해당하는 급여후기가 없을 경우 동일한 색상/몸무게/연령의 후기)} \end{cases}, \text{if } i^{th} \text{ 반려견} \in \text{정상견}$   
 $\begin{cases} i^{th} \text{ 견과 동일 질병/연령의 급여후기(전문가지식) 중 긍정적 후기} \\ \text{(해당하는 급여후기가 없을 경우 각각의 질병에 대한 후기)} \end{cases}, \text{if } i^{th} \text{ 반려견} \in \text{질병견}$

$w_j = \frac{r_j}{\sum_j r_j}$ ,  $r_j$ :  $j^{th}$  급여후기(전문가지식)에 부여된 평점

최종  $i^{th}$  반려견에 대한 적합성분 프로파일  $\vec{c}_i$ 와 부적합성분 프로파일  $\vec{d}_i$ 는 다음과 같이 구성된다.

$$\vec{c}_i = (x_{1i}, \dots, x_{Ti}), \vec{d}_i = (y_{1i}, \dots, y_{Ti})$$

$T$ : 영양성분의 수

**Step 3:**  $i^{th}$  반려견에 대한 적합성분 프로파일과 사료별 성분함량 벡터의 코사인 유사도에 의해 유 사도를 구하고 유사도가 경계 값보다 큰 Top-N 사료를 추천한다. 이때 성분함량 벡터의 원소가 음수가 될 수 없으므로 유사도는 0에서 1사이의 값을 가진다.

$$\text{적합사료 점수} = \frac{\vec{c}_i \cdot \vec{p}_j}{\| \vec{c}_i \| \times \| \vec{p}_j \|}$$

$$\text{부적합사료 점수} = \frac{\vec{d}_i \cdot \vec{p}_j}{\| \vec{d}_i \| \times \| \vec{p}_j \|}$$

$\vec{p}_j$ : 사료  $j$ 의 영양성분함량 벡터

먼저 Step 1에서는 사용자급여후기의 적합(부적합)사료 제품을 제품에 표시된 영양성분으로 분해하여 각 각의 후기에 해당하는 적합(부적합) 영양성분 행렬을 도출한다. 예를 들어 “포메라니안 견종이고 슬개골탈구 증상이 있었던 반려견에게 K사료를 급여한 결과 급여후기 평점이 5점(만족)”이고, “K사료의 성분이 콘드로이친 25%, 조단백질 16% 로 구성”되었다면, 이 후기는 “포메라니안 견종이고 슬개골탈구 증상이 있었던 반려견에게 적합했던 영양성분은 콘드로이친 25%, 조단백질 16%이다”로 변환된다. Step 2는 사료급여 후기를 추천단위별로 프로파일링하는 단계이다. 이는 각각의 사료급여 후기 중 추천대상 반려견에게 적용될 수 있는 사료급여 후기만을 추출하여 해당 반려견에게 급여시 적합한 영양성분 또는 부적합한 영양성분이 무엇인지를 요약하는 과정이다. 이 과정에서 반려견 사료전문가의 인터뷰 결과를 토대로 다음과 같은 선택적인 지식을 추천에 반영하기로 한다. 먼저 질병이 있는 반려견의 경우, 견종이나 색상, 몸무게와 무관하게 동일한 연령대이면서 같은 질병을 가진 반려견에게 급여한 후기를 토대로 적합(부적합) 영양성분을 프로파일링한다. 반면 건강상태가 정상적인 경우, 견종, 색상, 연령 정보가 동일한 반려견에게 급여한 후기를 토대로 적합(부적합) 영양성분을 프로파일링한다. 이때 동일한 견종, 색상, 연령의 반려견에게 급여한 사료후기가 없을 경우 동일한 색상, 몸무게, 연령으로 단위를 넓혀 프로파일링을 한다.

한편, 프로파일링은 사료급여 후기 평점정보를 가중치로 사용하여 각 영양성분을 가중 평균하는 방식으로 프로파일을 계산한다. 또한 사료급여 후기 외에도 전문가의 선택적 지식을 추천에 동시에 반영하기 위해 전문가가 입력한 반려견 상태별 적합 영양성분함량 정보에도 가중치를 부여함으로써 사용자급여후기와 전문가정보를 종합하여 추

천이 이루어지도록 한다. 전문가 정보에 부여하는 가중치는 입력파라미터로 설계하여 추천 실행 시 중요도를 조정할 수 있도록 한다. 추천단위별 각 영양성분에 대한 프로파일링은 적합성 피드백을 위한 Rocchio's Algorithm[Pazzani and Billsus, 2007]을 변형하여 적용한다. 마지막으로 Step 3에서는 적합 또는 부적합 추천사료 목록을 작성한다. 최종 추천은 적합 추천사료 목록과 부적합 추천사료 목록을 별도로 작성하는 것을 원칙으로 한다. 사료추천은 적합성분(부적합성분) 프로파일벡터와 사료별 성분함량 벡터의 내적을 통해 코사인 유사도 점수를 계산하여 추천순위를 결정한다.

## 4. 데이터 및 실험

### 4.1 데이터셋

제안한 영양성분 프로파일링 기반 사료추천 알고리즘(NRA)의 성능을 평가하기 위하여 시중에 유통되고 있는 사료와 주요 영양성분을 대상으로 코드화하고 각 사료별 영양성분 함량 정보를 수집하였다. 먼저 사료의 주요 영양성분으로 조단백질, 조지방, 조회분, 조섬유, 수분, 칼슘, 인, 오메가6, 오메가3, 글루코사민, 타우린, 탄수화물 등 12가지의 주요 영양성분을 추출하여 코드화 하였다. 또한 시중에 유통되고 있는 758개의 사료들을 수집하여 코드화 하고 사료별 영양성분 함량정보를 수집하였다. 사료추천에 필요한 또 다른 중요한 정보는 사료급여 후기평점 정보이다. 사료급여 후기평점은 특정 반려견에게 특정 사료를 급여한 결과 얼마나 만족스러웠는지를 나타내는 정보이다. 사료급여 후기평점은 반려견 커뮤니티 회원을 대상으로 조사하는 방식으로 수집하거나 크롤링을 통해 수집할 수 있으나, 본 연구에서는 제안한 알고리즘의 상용화 가능성을 확인하기 위한 성능비교 및 평가가 목적이기 때문에 난수를 발생하여 사료급여 후

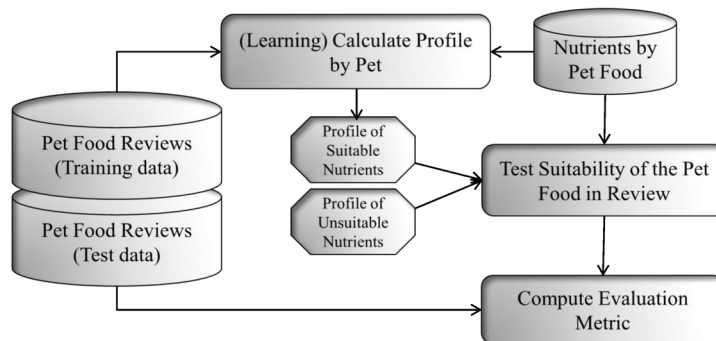
기평점 정보를 생성하였다. 커뮤니티 회원을 대상으로 사료급여 후기평점을 수집한다면 반려견 특성과 급여한 사료 간에 내재된 패턴이 존재할 가능성이 높는데 비해, 랜덤으로 생성된 사료급여 후기평점 정보에는 이러한 패턴이 존재하기 어렵다. 따라서 랜덤으로 생성된 사료급여 후기평점 정보를 이용한 알고리즘의 성능비교 및 평가는 보다 보수적인 접근방법이라 할 수 있다. 사료급여 후기평점 정보의 생성은 엑셀의 INDEX 함수를 이용하였으며, 총 1,000개의 후기를 유니폼 분포로 생성하였다. 생성된 사료급여 후기평점 데이터셋에 따르면, 질병 없는 정상견에 대한 후기는 220건, 1개 이상 질병이 있는 질병견의 후기는 780건이다. 전체 후기 중 평점이 3(보통)인 사례를 제거하고, 적합도 필드를 추가하여 평점이 1또는 2인 경우 불만족(0), 평점이 4 또는 5인 경우 만족(1)으로 코딩하였으며 최종 774개의 사례가 실험에 사용되었다.

#### 4.2 성능평가 방법

성능평가를 위한 절차는 다음과 같다. 먼저 사료급여 후기평점 정보의 2/3는 훈련용으로 1/3은 시험용으로 구분한다. 그리고 훈련용 데이터를 입력받아 반려견 특성별로 적합사료성분 프로파일과 부적합사료성분 프로파일을 각각 생성한다. 이렇게 생성된 프로파일을 급여한 사료의 성

분벡터와 비교하여, 급여한 사료의 성분벡터와 적합성분 프로파일과의 유사도가 부적합성분 프로파일과의 유사도 보다 높으면 적합한 사료로 그렇지 않으면 부적합한 사료로 판정하는 방식으로 추천 성능을 평가하였다. 이 과정을 그림으로 표현하면 <Figure 1>과 같다.

성능평가 지표로는 정밀도(Precision)와 재현율(Recall)을 사용하기로 한다. 정밀도는 “반려견에게 특정사료를 급여했을 때 적합할 것으로 예측된 경우의 수 중 실제 후기조사 결과에서 적합했던 경우의 수의 비율”로 계산한다. 또한 재현율은 “후기조사 결과 반려견에게 특정사료를 급여했을 때 적합했던 경우의 수 중 추천결과도 동일하게 적합할 것으로 예측된 경우의 수의 비율”로 계산한다. 한편 정밀도와 재현율을 동시에 고려한 F1 score도 예측 품질을 가늠하는 중요한 지표가 될 것이다. 또한 모형의 ROC 그래프는 가로축을 특이도(Specificity)로 하고 세로축을 재현율(Recall) 또는 민감도(Sensitive)로 하여 시각화 한 그래프이다. 일반적으로 ROC 그래프가 위로 향할수록 성능이 우수한 모형이라 할 수 있다. 이때 ROC는 그래프이기 때문에 모형의 정확도를 하나의 숫자로 나타내기 쉽도록 AUC(Area Under Curve)라는 값을 사용하는데, AUC값은 ROC 그래프의 면적으로 계산하며 최대값은 1이 된다.

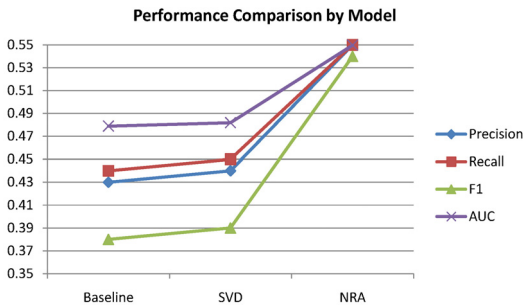


<Figure 1> The Process of Performance Evaluation



### 4.3 실험 결과

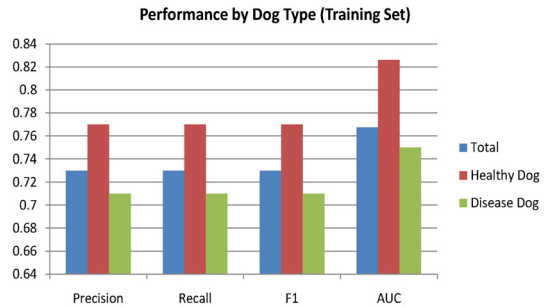
먼저 동일한 조건 하에서 시험용 데이터에 적용된 베이스라인 모형, SVD 모형, 그리고 NRA의 성능을 비교하면 <Figure 2>와 같다. <Figure 2>는 세 가지 모형의 정밀도, 재현율, F1, AUC값을 나타내고 있다. 네 가지 성능지표 모두에서 베이스라인 모형과 SVD 모형은 유사한 결과를 나타내고 있는데 비해 본 연구에서 제안한 NRA의 성능 평가치는 베이스라인 모형 대비 AUC의 경우 15%, F1 Score의 경우 42% 가량 향상된 성능 값을 보여주고 있다. 이는 영양성분 함량을 고려하지 않고 단순히 사료에 대한 적합여부만을 이용하여 추천에 활용하는 방식에서는 반려견-사료 행렬의 희박성이 높기 때문에 추천 성능이 저하되는 결과를 초래한 것으로 추측해 볼 수 있다.



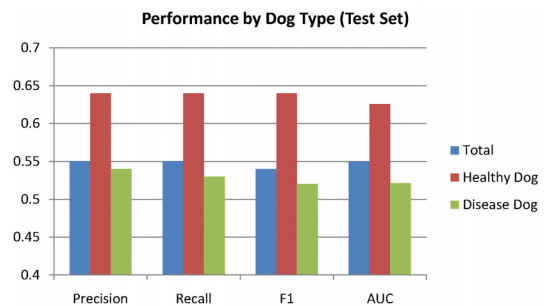
<Figure 2> Performance Comparison by Model

<Figure 3>과 <Figure 4>는 훈련용 데이터와 시험용 데이터를 구분하여 NRA의 성능을 보다 세부적으로 평가한 결과를 보여주고 있다. 두 그림 모두에서 건강견에 대한 사료추천 성능이 질병견에 대한 사료추천 성능보다 높게 나타나고 있다. 건강견과 질병견의 추천성능이 다른 것은 프로파일링의 단위가 달랐기 때문이다. 즉 건강견의 경우 건종, 색상, 연령별로 1차 프로파일링하고 해당 급여후기가 없을 경우 색상, 몸무게, 연령별로 프로파일링한데 비해, 질병견은 연령

과 동일한 질병 별로 1차 프로파일링하고 해당 후기가 없을 경우 연령과 최소 한 개 질병 일치 단위로 프로파일링 하였기 때문이다. 정상견에 비해 질병견의 추천성능이 낮은 것은 연령과 모든 질병이 일치하는 반려견에 대한 급여후기가 별로 없기 때문에 연령과 최소 한 개의 질병이 일치하는 급여후기를 대상으로 프로파일링이 주로 이루어졌으며, 이로 인해 추천성능이 저하되었을 것으로 추측된다. 따라서 질병견에 대한 프로파일링 단위를 조정한다면 추천 성능을 보다 높일 수도 있을 것이다.



<Figure 3> Performance Applied to Training set by Dog Type



<Figure 4> Performance Applied to Test Set by Dog Type

## 5. 결론

본 연구에서는 반려견 사료추천을 위한 콘텐츠 기반 추천 알고리즘을 제안하였다. 본 연구에서 제안하는 사료추천 방법은 사용자경험에 기

반한 집단지성과 전문가의 사전지식을 동시에 활용하여 적합사료 또는 부적합 사료를 추천하는 방법이다. 보다 구체적으로 사용자가 반려견 특성정보를 입력하면 반려견 특성정보별 사용자 후기평점정보와 반려견 특성정보별 적합한 영양성분에 대한 전문가의 지식을 결합하여 해당 반려견 특성정보에 요구되는 영양성분을 발견하며, 이들 영양성분을 가장 적절하게 포함하고 있는 사료를 추천하는 시스템을 제안하고 기존 협업필터링 기반의 모형과 성능을 비교하였다. 성능평가는 재현율, 정밀도, F1, AUC 관점에서 이루어졌으며 성능평가 결과 제안한 NRA가 베이스라인 모형 및 SVD 모형에 비해 높은 성능을 보여주는 것을 확인하였다. 영양성분 함량을 고려하지 않고 단순히 사료에 대한 적합여부만을 이용하여 추천에 활용하는 방식인 베이스라인 모형 및 SVD 모형에서는 반려견-사료 행렬의 희박성이 높기 때문에 추천 성능이 저하되는 결과를 초래한 것으로 추측해 볼 수 있다. 또한 NRA 추천알고리즘이 질병건 보다는 정상건에 대한 사료추천에 있어서 보다 높은 성능을 나타내었다. 이러한 결과는 반려견의 질병상태에 적합한 사료를 추천하기 위한 당초 추천시스템 개발 동기를 충족시키지 못하는 결과이다. 그러나 질병건에 대한 성능평가가 건강건에 비해 낮게 나타난 것은 건강건에 대한 급여후기에 비해 건종별 질병별 급여후기가 상대적으로 작아서 나타난 현상이므로 건종별 질병별 급여후기를 충분히 수집하여 추천에 활용하거나 질병건에 대한 프로파일링 대상을 달리하여 추천에 적용한다면 이러한 문제는 극복될 수 있을 것이라 판단된다.

본 연구에서는 반려견의 신체 및 건강상태에 따른 적합 또는 부적합 사료를 추천하는 알고리즘을 제안하였다. 그러나 반려견이 건식사료를 좋아하는지 습식사료를 좋아하는지, 또는 반려견이 좋아하는 향미나 재료가 무엇인지 등과 같

이 반려견의 사료에 대한 선호도를 동시에 고려하여 추천할 수 있다면 한층 사료추천의 만족도를 높일 수 있을 것이다. 이 때 적합영양성분은 동일한 특성을 가지는 반려견에게 동시에 적용될 수 있기 때문에 동일한 특성의 반려견별로 프로파일링 하는 것이 적합한데 비해 선호사료 정보는 동일한 특성의 반려견이라도 개별적으로 선호도는 다를 수 있기 때문에 적합영양성분과 선호사료 정보를 동시에 고려하는 형태로 추천이 이루어지기 보다는 적합영양성분 기준의 추천과 선호기반의 추천이 별도로 이루어지도록 구현하는 것이 바람직 할 것이다. 본 연구는 반려견 시장의 급성장과 때를 같이하여 반려인들의 사료 선택의 복잡성 문제를 해결하기 위해 시도된 첫 번째 연구라는 점에서 의의를 찾을 수 있다.

## References

- [1] Almeida, D., *Personalized Food Recommendations*, Dissertation of MS, Tecnico Lisboa, 2015.
- [2] Bengio, Y., Simard, P., and Frasconi, P., "Learning long-term dependencies with gradient descent is difficult", *IEEE Transactions on Neural Networks*, Vol. 5, No. 2, 1994, pp. 157-166.
- [3] Forbes, P. and Zhu, M., "Content-boosted Matrix Factorization for Recommender Systems: Experiments with Recipe Recommendation", *RecSys 2011 Proceedings of the fifth ACM conference on Recommender systems*, 2011, pp. 261-264.
- [4] Freyne, J. and Berkovsky, S., "Recommending food: Reasoning on recipes and ingredients", In *Proceedings of the 18th International Conference on User Modeling, Adaptation, and Personalization*, Vol. 6075, 2010, pp. 381-

- 386.
- [5] Koren, Y. and Sill, J., "Collaborative Filtering on Ordinal User Feedback", *ACM Conference on Recommendation Systems (RecSys'11)*, 2011.
- [6] Lin, C. J., Kuo, T. T., and Lin, S. D., "A Content-Based Matrix Factorization Model for Recipe Recommendation", *Advances in Knowledge Discovery and Data Mining*, Vol. 8444, 2014.
- [7] Pazzani, M. J. and Billsus, D., "Content-Based Recommendation Systems", *The Adaptive Web*, Vol. 4321, 2007, pp. 325-341.
- [8] Rendle, S., Freudenthaler, C., and Schmidt-Thieme, L., "Factorizing personalized markov chains for next-basket recommendation", In *WWW Conference*, 2010, pp. 811-820.
- [9] Sato, M., Izumo, H., and Sonoda, T., "Discount Sensitive Recommender System for Retail Business", *Proceedings of the 3rd Workshop on Emotions and Personality in Personalized Systems*, 2015, pp. 33-40.
- [10] Ueda, M., Takahata, M., and Nakajima, S., "User's food preference extraction for personalized cooking recipe recommendation", In *CEUR Workshop Proceedings*, Vol. 781, 2011, pp. 98-105.
- [11] Wang, J., Sarwar, B., and Sundaresan, N., "Utilizing related products for post-purchase recommendation in e-commerce", *Proceedings of the fifth ACM Conference on Recommender Systems*, 2011, pp. 329-332.

## ■ 저자소개



Hee Seok Song

Song, Hee Seok is a professor of management information systems department at Hannam University in Korea. He received bachelor's degree from Korea University and PhD degree from Korea Advanced Institute of Science and Technology. His research interests include intelligent computing technology, business intelligence, and social network. He published his studies to Knowledge-based systems, Expert systems with applications, Artificial Intelligence Review, and many international and domestic journals.