

협업필터링을 이용한 개인화 애완동물 돌보미 추천 시스템

김한이*, 황동현*, 박두순*
 *순천향대학교 컴퓨터소프트웨어공학과
 e-mail : gksdl1995@nate.com

A Personalized Pet Sitter Recommendation System based on Collaborative Filtering

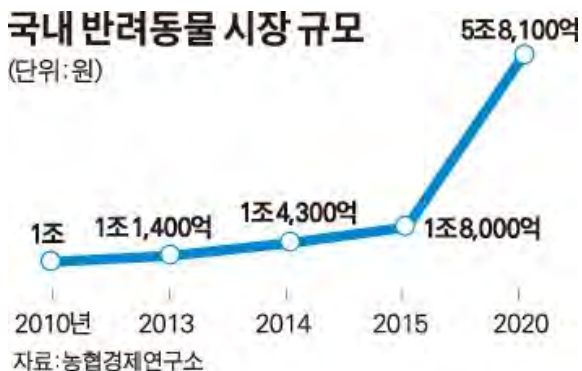
Han-Yi Kim, Dong-Hyun Hwang, Doo-Soon Park
 *Dept. of Computer Software Engineering, SoonchungHyang University

현대사회의 발전에 따라 1인가구가 증가하면서 애완동물을 키우는 애완 인이 증가하게 되었다. 애완동물을 가족의 구성원으로 여기는 사람들이 많아지면서 반려동물에게 사용하는 지출 규모가 폭발적으로 증가하였다. 자연스럽게 애완동물 사업 규모가 커지면서 서비스 산업이 확장되고 있다. 이에 따라 반려인들은 자신의 반려동물을 잘 돌봐줄 수 있는 애완동물 돌보미 서비스를 제공받기를 원한다. 본 논문에서는 협업 필터링방법에 사용자의 개인화 요소를 이용하여 애완동물 돌보미 중 사용자에게 적합한 애완동물 돌보미를 추천하는 시스템을 제안한다.

1. 서론

현대사회의 발전에 따라 핵가족, 노령인구, 1인가구가 증가하고 있다. 애완동물을 키우는 것으로 만족을 얻음에 따라서 애완 인이 100만에 다다르게 되었다. 경제 전반의 불황에도 불구하고 애완동물 시장은 활황세를 유지하고 있다. 현재 애완동물 사업 규모는 약 2조억 원으로 5년 내에 5조억 원까지 시장이 확대될 전망이다. 이에 따라 정부에서는 적극적으로 애완동물 산업을 육성하고 있다. 애완동물을 가족의 구성원으로 여기는 사람이 많아지면서 반려동물에게 지출하는 규모가 폭발적으로 늘어나고 있는 까닭이다. (그림 1)은 국내 반려동물 시장 규모 현황 이다.

자신의 애완동물을 돌볼 수 없는 상황이 발생하였을 때 이를 해결해 주는 서비스를 원한다. 이러한 서비스를 제공하기 위한 것으로 애완동물 호텔, 애완동물 돌보미가 존재한다. 애완동물 호텔은 반려동물이 케이지에 가두어져 있고 제대로 케어를 받지 못하는 단점이 있다. 그러나 애완동물 돌보미를 이용하게 된다면 이러한 단점을 보완할 수 있고 이전과 비슷한 환경에서 머무를 수 있다. 이를 위해 협업 필터링 방법에 개인화 요소를 사용하여 사용자에게 가장 적합한 애완동물 돌보미를 추천하는 시스템을 제안한다.



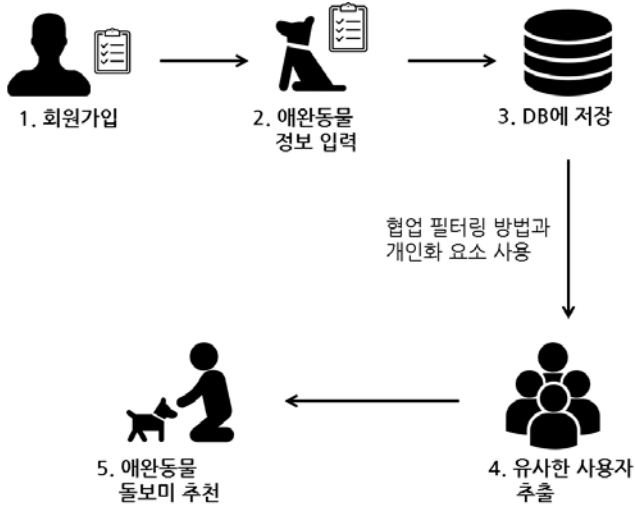
(그림 1) 국내 반려동물 시장 규모 현황[1].

2. 애완동물 돌보미 추천 시스템의 구성

추천 시스템은 사용자의 과거 선호도에 기반 하여 예측한 후 새로운 아이템에 대한 평가치를 예측하여 제공하는 시스템이다[2]. 가장 성공적인 추천기법은 협업 필터링이다. 협업 필터링 중 사용자 기반 협업 필터링은 선호도가 비슷한 유사 고객들이 동일하게 평가한 상품에 대하여 상대적으로 높은 예측력을 보이고 있다. 하지만, 특징 벡터의 차수가 적을수록 객체 사이의 유사성을 발견하기 어렵게 되기 때문에 선호도 예측 정확도가 하락 하는 현상을 보인다. 이것을 희박성(sparsity)이라 한다. 이 현상을 해결하기 위하여 협업 필터링방법에 사용자 유사도를 사용하지 못하고, 사용자의 개인화 요소를 입력받고 피어슨 상관관계를 이용하여 유사도를 측정한다. 개인화(personalization)는 어떤 특정 고객만의 고유하고도 특정한 요구에 대하여 민첩하게 대응할 수 있는 것을 의미한다[3]. (그림 2)는 애완동물 돌보미 추천시스템의 시나리오이고, (그림

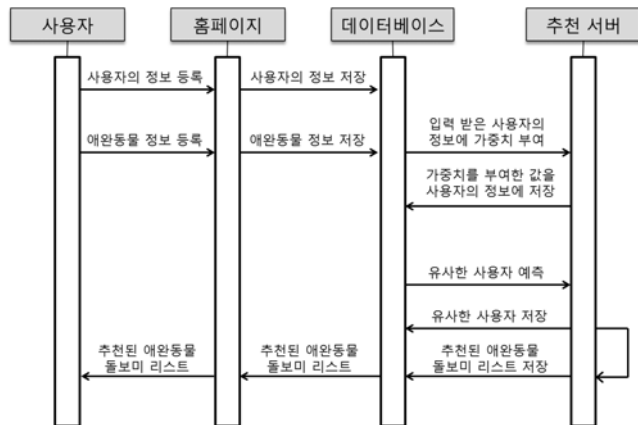
※ 본 연구는 미래창조과학부 및 정보통신기술진흥센터의 대학ICT연구센터육성 지원사업의 연구결과로 수행되었음 (IITP-2016-H8601-16-1009)

3)은 시퀀스 다이어그램이다.



(그림 2) 추천 시스템 시나리오

(그림 2)의 추천 시스템에서 (1)새로운 사용자가 회원가입을 할 때, 개인화 요인으로 사용하게 될 나이, 성별, 주소, 직업을 입력받는다. (2)회원이 가입이 완료된 후 개인화 요인으로 사용하게 될 애완동물의 정보인 종, 성별, 중성화 여부, 크기, 나이, 무게, 친화력, 짚음 정도, 질병 여부, 마킹 여부, 분리불안증에 대한 정보를 입력받는다. (3)입력 받은 정보들을 데이터베이스에 저장한다. (4)협업 필터링 방법과 입력받은 개인화 요소를 이용하여 가중치를 부여하고 근접이웃을 구성한다. 그리고 근접이웃과 피어슨 상관관계(Person product -moment correlation coefficient)를 이용하여 유사도를 측정하고, 유사도가 높은 Top -n리스트를 작성한다. (5)Top -n리스트로부터 사용자와 유사한 다른 사용자가 추천한 애완동물 돌보미를 사용자에게 추천한다. (그림 3)은 추천시스템 시퀀스 다이어그램이다.



(그림 3) 추천 시스템 시퀀스 다이어그램

3. 애완동물 돌보미 추천 시스템의 구현

추천 시스템을 사용하기 위해 개인화 요소들을 입력받

는 회원가입 양식은 (그림 4)과 같다. 회원가입 한 후에 애완동물의 정보를 입력받는 양식은 (그림 5)와 같다. 입력한 정보들은 사용자 테이블에 저장되며 사용자 데이터베이스의 일부는 (표 1)과 같다.

회원 가입

(그림 4) 회원가입 양식

애완동물 정보

(그림 5) 애완동물 정보 양식

[표 1] 사용자의 데이터베이스 일부분

User													Dog					Pet Site
ID	Age	Gender	Job	Address	Species	Sex	Neutralization	Size	Age	Weight	Sociability	Bark	Disease	Marking	Separation anxiety	Pet Site		
User_H_1	30대	여	판매	서울	사슴	암컷	중성화	중형	성인	15kg 이상	있음	없음	없음	없음	없음	Pet_26		
User_H_2	20대	여	학생	서울	포메라니안	암컷	행동	소형	성인	0-5kg 미만	없음	없음	없음	없음	없음	Pet_25		
User_H_3	30대	여	사무직	경기도	몰티쉬	수컷	행동	중형	성인	10-15kg 미만	없음	없음	없음	없음	없음	Pet_25		
User_H_4	30대	남	건축사	서울	세이브	수컷	행동	대형	성인	15kg 이상	없음	없음	없음	없음	없음	Pet_17		
User_H_5	40대	여	단순노동	전라도	요크셔테리어	수컷	행동	소형	성인	0-5kg 미만	없음	없음	없음	없음	없음	Pet_36		
User_H_6	20대	여	학생	부산	시츄	수컷	행동	소형	성인	5-10kg 미만	없음	없음	없음	없음	없음	Pet_49		
User_H_7	40대	여	서비스	서울	스피츠	암컷	행동	소형	성인	5-10kg 미만	없음	없음	없음	없음	없음	Pet_42		
User_H_8	20대	여	전문가	경기도	말티즈	암컷	행동	소형	성인	0-5kg 미만	없음	없음	없음	없음	없음	Pet_8		
User_H_9	50대	여	통역사	전라도	말티즈	수컷	행동	소형	노년	0-5kg 미만	없음	없음	없음	없음	없음	Pet_16		
User_H_10	20대	남	학생	강원도	푸들	암컷	행동	소형	성인	0-5kg 미만	없음	없음	없음	없음	없음	Pet_36		
User_H_11	60대	여	황제	충청도	포메라니안	수컷	행동	중형	성인	5-10kg 미만	없음	없음	없음	없음	없음	Pet_20		
User_H_12	20대	여	전문가	서울	믹스견	암컷	행동	중형	노년	10-15kg 미만	없음	없음	없음	없음	없음	Pet_7		

개인화 요소인 연령대, 성별, 직업 세 가지를 사용했을 때 user_1에 대한 근접이웃 구성은 (표 2)와 같다. 개인화 요소인 주소를 사용하여 user_1에 대한 근접이웃 구성은 (표 3)과 같다. 사용자의 개인화 요소와 애완동물의 정보인 종, 성별, 중성화 여부, 크기, 나이, 무게, 친화력, 짚음 정도, 질병 여부, 마킹 여부, 분리불안증 등 11가지를 추가하였을 때의 uae_1에 대한 근접이웃 구성은 (표 4)와 같다. (표 2)와 (표 3)을 비교하면 개인화 요

소 중 주소의 정보에 민감한 것을 알 수 있다. (표 3)과 (표 4)를 비교하면 개인화 요소는 애완동물의 정보에 매우 민감하다는 것을 알 수 있다.

[표 2] 개인화 요소 3개 사용한 근접이웃 구성

ID	Union	Intersection	Similarity	Pet Sitter
User_1	X	X	X	X
User_74	3	3	1.00000	Pet_10
User_12	3	3	1.00000	Pet_50
User_87	3	3	1.00000	Pet_22
User_3	4	2	0.50000	Pet_25
User_33	4	2	0.50000	Pet_26
User_67	4	2	0.50000	Pet_42
User_97	5	1	0.20000	Pet_31
User_64	5	1	0.20000	Pet_21
User_15	6	0	0.00000	Pet_38
User_65	6	0	0.00000	Pet_2

[표 3] 개인화 요소 주소를 사용한 근접이웃 구성

ID	Union	Intersection	Similarity	Pet Sitter
User_1	X	X	X	X
User_87	4	4	1.00000	Pet_22
User_74	5	3	0.60000	Pet_10
User_12	5	3	0.60000	Pet_50
User_67	4	2	0.50000	Pet_42
User_33	6	2	0.33333	Pet_26
User_3	6	2	0.33333	Pet_25
User_65	7	1	0.14286	Pet_2
User_15	7	1	0.14286	Pet_38
User_97	7	1	0.14286	Pet_31
User_64	8	1	0.12500	Pet_21

[표 4] 애완동물의 개인화 요소를 사용한 근접이웃 구성

ID	Union	Intersection	Similarity	Pet Sitter
User_1	X	X	X	X
User_15	15	15	1.00000	Pet_38
User_87	16	14	0.87500	Pet_22
User_97	17	13	0.76471	Pet_31
User_65	18	12	0.66667	Pet_2
User_64	18	12	0.66667	Pet_21
User_12	18	12	0.66667	Pet_50
User_33	19	11	0.57895	Pet_26
User_67	20	10	0.50000	Pet_42
User_3	21	9	0.42857	Pet_25
User_74	22	8	0.36364	Pet_10

최종적으로 도출된 결과를 이용하여서 본 논문에서는 사용자의 주소, 애완동물의 성별, 중성화, 크기, 나이, 무게, 사회성, 짖음, 질병, 영역표시, 분리불안 총 11개의 개인화 요소를 사용하였다. 주소(15%), 성별(5%), 중성화(5%), 크기(15%), 나이(5%), 무게(5%), 사회성(5%), 짖음(10%), 질병(15%), 영역표시(10%), 분리불안(10%)의 가중치를 부여

하였을 때 user_1에 대한 근접이웃 구성은 (표 5)와 같다.

[표 5] User_1에 대한 가중치를 이용한 근접이웃 구성

ID	Union	Intersection	Similarity	Pet Sitter
User_1	X	X	X	X
User_87	1.16	0.84	0.72414	Pet_22
User_67	1.19	0.81	0.68067	Pet_42
User_15	1.24	0.76	0.61290	Pet_38
User_3	1.29	0.71	0.55039	Pet_25
User_64	1.23	0.67	0.54472	Pet_21
User_33	1.23	0.67	0.54472	Pet_26
User_97	1.38	0.62	0.44928	Pet_31
User_74	1.43	0.56	0.39161	Pet_10
User_12	1.47	0.53	0.36054	Pet_50
User_65	1.53	0.47	0.30719	Pet_2

근접이웃과의 유사도는 피어슨 상관계수를 통하여 구한다. 피어슨 상관계수는 -1부터 1의 값을 가지며, 관계가 크면 1에 가까워지고 서로 반대되는 관계를 가지면 -1에 가까워진다. 피어슨 상관계수를 이용한 수식은 아래 (그림 6)과 같다.

$$r = r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

(그림 6) 피어슨 상관계수 공식

[표 6]은 사용자들과 유사도의 측정값을 나타낸 것이다. 사용자는 자신과의 유사도는 항상 1.0이다. 특정 사용자에게 애완동물 돌보미를 추천할 때, 다른 사용자들과의 유사도 측정값을 이용하여 측정값이 가장 높게 나온 사용자들이 선호하는 애완동물 돌보미를 사용자에게 추천한다.

[표 6] User_1에 대한 유사도

ID	Compare User	Similarity
User_1	user_87	0.0678
	user_67	0.0597
	user_15	0.0512
	user_3	0.0347
	user_64	0.0133
	user_33	0.0012
	user_97	-0.0033
	user_74	-0.0459
	user_12	-0.0982
	user_65	-0.1123
	...	

마지막으로 유사도가 가장 높은 근접 이웃이 추천하였던 애완동물 돌보미를 사용자에게 보여준다. 서울, 암컷, 중성화 함, 중형견, 10살 이상, 5~10kg미만, 사회성 없음, 조용함, 질병 없음, 영역표시 안함, 분리불안 있음 이라는 개인화 요소를 입력한 후에, 피어슨 상관계수를 이용해 유사도

를 계산하였을 때 추천되는 애완동물 돌보미를 나타낸 결과는 (그림 7)과 같다.

애완동물 돌보미 추천



(그림 7)사용자에게 추천된 결과

4. 결론

본 논문은 사용자의 정보뿐만 아니라 애완동물의 정보를 입력받아 사용자에게 적합한 애완동물 돌보미를 추천하였다. 최종적으로 개인화 요소를 사용하였던 결과들을 비교분석 해보았을 때 사용자의 주소와 애완동물의 정보가 민감하다는 것을 보았다. 즉, 사용자의 정보 중 연령, 성별, 직업과 애완동물의 종에는 가중치를 사용하지 않는다. 그리고 애완동물의 성별, 중성화, 크기, 나이, 무게, 사회성, 짝음, 질병, 영역표시, 분리불안에 가중치를 부여하고 추천하는 것이 좋다.

향후 연구과제는 현재 사용한 애완동물의 요소 외에도 또는 요소보다 고유하고도 특정한 요구에 대하여 민첩하게 대응할 수 있는 것을 찾아내는 것이다. 요소를 정확히 제시하여 애완동물 돌보미를 추천하여 준다면 더욱 더 정확한 시스템을 구현해낼 수 있다.

참고문헌

[1] <http://www.hankookilbo.com/v/78506f98d20a4b058d91bf87993c20>
 [2] 김경록, 변재희, 문남미, “장르유사도와 선호장르를 이용한 협업필터링 설계”
 [3] 홍태호, 이희정, 서보밀, “클러스터링 기반 사례기반추론을 이용한 웹 개인화 추천시스템”