

레시피 연결망에서 요리 난이도 및 유사성 분석

Analyzing the Difficulty and Similarity of Cooking in the Recipe Network

김수도*, 이윤정*, 윤성민**, 조환규***

부산대학교 사회급변현상연구소*, 부산대학교 경제학부**, 부산대학교 정보컴퓨터공학부***

Su-Do Kim(kimsd@pusan.ac.kr)*, Yun-Jung Lee(leeyj01@pusan.ac.kr)*,
Seong-Min Yoon(smyoon@pusan.ac.kr)**, Hwan-Gue Cho(hgcho@pusan.ac.kr)***

요약

인터넷을 통해 공개되고 있는 요리 레시피에 대한 분류 및 평가는 작성자의 문화적 배경, 요리능력, 요리 경험, 선호도 등 주관적 기준에 따라 제시되고 있다. 이 연구에서는 요리 난이도를 측정하기 위한 척도로서 정보 엔트로피 개념을 통해 객관화한다. 또한 요리의 공통 엔트로피를 계산하여 레시피 사이의 유사성을 측정하고, 레시피를 개체로 하는 유사도 기반의 사회연결망을 생성한다. 요리난이도를 측정한 결과, 동태해물찜(한식), 베지테리안 라자나(이탈리아) 등은 요리난이도 측면에서 가장 어려운 요리로, 초고추장(한식)과 두부스테이크(이탈리아)는 가장 쉬운 요리로 나타났고, 레시피 연결망의 거리공간을 통해 한식과 아시아 요리는 유사성이 높은 것을 확인할 수 있었다. 또한 활용적 측면에서 특정 요리와 유사한 요리는 무엇인지, 요리를 대체할 수 있는 유사한 요리 그룹은 어떤 것이 있는지, 요리용이성 관점에서 식단을 준비할 때 가장 합리적인 계획은 무엇인지를 보여주었다.

■ 중심어 : 레시피 | 요리재료 | 요리동사 | 엔트로피 | 요리 난이도 | 레시피 연결망 |

Abstract

The classification and evaluation of cooking that is being published on the internet are presented without scientific criteria based on individual subjective factors. In this paper, we objectified the degree of cooking difficulty based on the information entropy. And we measured the similarity by calculating the common entropy between recipes and constructed a social network based on the recipe similarity. As a result of measuring the cooking difficulty, 'Dongtae Haemul-jjim' (Korean) and 'Vegetarian Lasagna' (Italy) are the most difficult recipes and 'Gochu-jang' (Korean) and 'Tofu steak' (Italy) are the easiest recipes. Through the recipe network, the similarity between Korean and Asian cooking is higher than Western cuisine. We showed a similar recipe to a particular cooking, the group of similar recipes, and reasonable schedule when preparing the menu from the viewpoint of ease of cooking.

■ keyword : Recipe | Ingredient | Cooking Verb | Entropy | Recipe Difficulty | Recipe Network |

* 이 논문은 2013년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2013S1A3A2042747).

접수일자 : 2016년 03월 29일

심사완료일 : 2016년 05월 18일

수정일자 : 2016년 05월 12일

교신저자 : 조환규, e-mail : hgcho@pusan.ac.kr

I. 서론

최근 방송과 미디어에서는 음식과 요리가 ‘쿡방’이라는 새로운 문화 트렌드로 인기를 끌고 있다. 음식은 인간의 가장 기본적인 공통 관심사이며 창조 활동이라고 할 수 있다. 사람들은 인터넷을 통해 자신만의 요리와 레시피를 공개하고 이를 매개로 공동체 활동을 하고 있다. 레시피에 대한 분류 및 평가는 그것을 기술한 사람의 요리능력, 문화배경 등 개인적 특성과 밀접하게 연관되어 있다. 예를 들어 요리의 쉽고 어려움을 나타내는 요리난이도는 객관적이고 과학적 기준없이 레시피를 작성한 개인의 주관적 판단에 의해 작성되고 있다.

예를 들어, ‘수제비’를 만들기 위해 19 개(마늘, 밀가루, 식용유, 등) 재료가 필요하고 ‘약선 오골계탕’이 11 개(오골계, 업나무, 대추 등) 재료가 필요하다면, ‘수제비’가 ‘약선 오골계탕’ 보다 어려운 요리인가? 일반적으로 요리재료가 적을수록 쉬운 요리로, 많을수록 어려운 요리로 판단한다. 그러나 요리난이도를 요리재료의 개수만으로 판단하는 데는 어려움이 있다.

이 연구에서는 주관적 기준으로 기술된 레시피에 대한 분류 및 평가를 정보공학에 기반한 엔트로피 개념을 이용하여 객관화한다. 또한 레시피가 어떤 사회적 관계로 연결되어 있으며 어떻게 나타나고 공유되는가를 탐구하기 위해 각 요리들의 상호 관련성을 두 요리의 공통 엔트로피라는 개념으로 해석하여 레시피를 개체로 하는 사회연결망을 구성한다. 레시피 네트워크를 통해 레시피간의 상호관련성을 거리공간으로 매핑하고, 레시피에 대한 포괄적 네트워크 지도를 생성한다. 본 연구를 통해 특정 요리와 연관된 요리들은 어떤 것이 있으며, 유사한 요리 그룹은 어떻게 형성되고 있는지, 그리고 어떤 요리들이 융합되어 새로운 요리를 생성할 수 있는지를 보여준다. 또한 요리식단을 준비하거나 계획할 때 요리준비나 요리절차의 관점에서 어떻게 계획하는 것이 가장 합리적인지를 객관적 기준으로 제시한다.

II. 관련 연구

레시피에 대한 가장 일반적인 연구는 개인별 맞춤 요

리를 제공해주는 추천시스템이다[1-3]. J. Freyne은 개인별 이전 선호도를 조사하여 체중조절이나 영양성분을 재조정할 수 있는 요리 추천 시스템과 주어진 재료를 이용하여 식단 메뉴에 대한 설계 방법을 보여주고 있다[1]. 또한 사용자가 원하는 요리와 준비된 재료를 참고하여 인접한 유사 요리로부터 식단을 설계하여 제시하는 연구도 진행되었다[2][3].

복잡계 과학 중에서 하나의 개체가 아닌 전체 연결망에서 나타나는 창발적 특성이나 개체에서는 관찰할 수 없는 특성을 연구하는 네트워크 연구는 오랫동안 연구되어 왔다[4-6]. 요리재료를 사회연결망 기법으로 연결하고 그 속에서 어떤 특징들을 찾는 연구는 최근 연구의 중심주제이다[7-9]. L.Wang은 요리법 분석에서 중국 요리에서 자주 나타나는 반복적 구조를 찾아내는 방법을 제시하고 있다. 이 연구에 따르면 중식 요리법은 더 이상 변화되지 않는 특정 기본 작업의 조합으로 구성되어 있고, 두 중식요리가 얼마나 유사한 것인지를 쉽게 계산할 수 있음을 보여주고 있다[7].

서양 요리와 재료에 대한 가장 광범위한 연구는 C.Teng에 의해서 공개되었다[8]. 웹에서 레시피 4만여 건을 분석하여 재료에 대한 연결망을 생성하였다. 우연히 어떤 재료에 같이 들어갈 확률과 실제 두 재료가 같은 요리에 포함된 정도를 비교하여 재료 보완 연결망을 생성하였고, 요리에서 대체재료를 추천해줄 수 있는 대체 재료 연결망을 생성하였다. 연구를 통하여 미국의 지역별 요리법의 특성이 어떤 것인지를 밝히기도 하였다.

또 다른 연구로는 재료의 화학성분과 영양성분을 활용하여 각 요리의 영양학적 의미를 분석하고 화학적 성분 수준까지 고려한 대체 재료 연결망에 대한 연구가 진행되어 연구의 새로운 면을 보여주고 있다[10-12].

우리는 이 연구에서 식단 설계나 요리 추천 시스템이 아니라 요리법, 또는 재료들을 독립적인 개체로 간주하여 이들의 연관관계로부터 새로운 사회연결망을 생성하고 전체적 특성을 규명하고자 한다.

III. 데이터 수집 및 통계분석

3.1 데이터 수집 및 정제

레시피 분석을 위해 'Allrecipes.kr' 한국 웹사이트에서 총 1135 개 레시피를 다운로드하였다. 수집한 레시피는 한국 775 개, 중국 52 개, 일본 73 개, 태국 18 개, 서양 217 개이다. 각 레시피는 요리재료, 요리시간, 요리절차 등의 정보를 포함하고 있으며, 각 요리재료는 재료명, 재료량, 재료형태 등을 포함하고 있다. [그림 1]은 'Allrecipes.kr' 에서 다운로드한 '된장찌개' 레시피를 보여주고 있다.

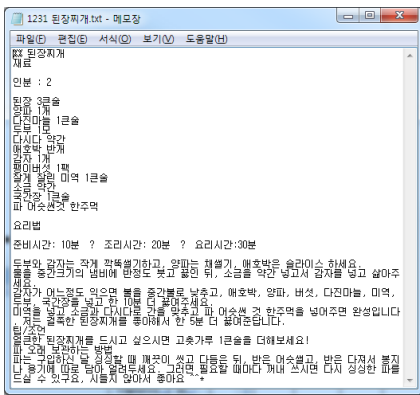


그림 1. '된장찌개' 레시피

표 1. 재료명과 동사 추출을 위한 데이터 처리방법

분류	작업종류	내용
요리 재료명 추출	재료명이 아닌 단어 제거	수량을 의미하는 단어들 예) '3큰술', '약간', '한주먹', '1개' 등 재료상태를 의미하는 단어들 예) 잘게 잘린', '어슷썬것', '불린', '얇힌' 등 (기호들 예) '+', '-', '1/2', '1/3' 등
	재료명이 일치하지 않는 단어 정제	동일한 재료명이지만 문법이 다른 단어들 예) 계파류, 계파류, 검은콩, 검은콩 등 동일한 재료명이지만 유사한 단어들 예) 말차가루, 말차가루, 말차육수 등
	형용사가 붙은 단어 정제	재료명에 형용사가 붙은 단어들 예) 다진마늘, 슬라이스마늘, '통마늘', '꽃마늘' 등을 '마늘' 로 분류
	재료원형으로 분류	일치하지 않지만 유사한 재료들을 하나의 재료원형으로 분류 예) 맛술, 마린, 마향, 요리술, 조식술 등을 '맛술' 로 분류
요리 동사 추출	형태소 분석	형태소의 동사에 해당하는 단어를 추출 예) '같은', '건져', '데쳐', '끓여' 등
	일반 동사 제거	요리방법과 거리가 먼 단어들 예) 찌다, 올리다, 내리다, 자르다 등
	동사원형으로 분류	하나의 동사 원형으로 분류 예) '끓는', '끓여', '끓여주면' 등을 '끓이다' 로 분류

필요한 정보만을 추출하기 위해 데이터 파싱, 형태소 분석기, 수작업 등을 통해 요리재료명과 요리동사를 추출하였다. 최종 정제된 요리재료명은 444 개이고, 요리동사는 129 개다. [표 1]은 요리재료명과 요리동사명을 추출하기 위한 데이터 정제 단계를 보여주고 있다.

3.2 데이터 통계분석

요리재료의 관점에서 '마늘'과 같이 어떤 재료는 많은 요리에 사용되지만 어떤 재료는 거의 사용되지 않는 희귀한 재료이기도 하다. 만약 어떤 요리를 준비하고자 할 때 흔히 사용되는 요리재료가 필요할 경우 쉽게 준비할 수 있지만 '방풍나물'처럼 희귀한 요리재료가 많이 필요할 경우 요리준비는 어려워진다. [표 2]는 전체 1135 개 레시피에 포함된 요리재료의 빈도수에 따른 상위 10 개 및 하위 10 개를 정리한 것이다. 레시피에서 가장 많이 사용되는 요리재료는 '마늘'로서 총 레시피의 590 개(약 52%)에 포함되고 있으며, 다음으로 '소금', '파', '간장' 순으로 나타나고 있다. [표 3]은 요리동사의 빈도수를 보여주며, 가장 빈번하게 사용되는 동사는 '끓이다'로 약 51%에서 사용되고 있다.

표 2. 요리재료 빈도수에 따른 상위 및 하위 순위

상위	재료명	빈도수	빈도율(%)	하위	재료명	빈도수	빈도율(%)
1	마늘	590	51.98%	1	감	1	0.09%
2	소금	571	50.30%	2	감초	1	0.09%
3	파	441	38.85%	3	곤드레	1	0.09%
4	간장	435	38.33%	4	곰취	1	0.09%
5	후추	424	37.36%	5	곰피	1	0.09%
6	참기름	373	32.86%	6	강낭콩	1	0.09%
7	설탕	373	32.86%	7	가지미	1	0.09%
8	양파	348	30.66%	8	꿀두기	1	0.09%
9	깨	292	25.73%	9	고구마줄기	1	0.09%
10	고추	228	20.09%	10	다슬기	1	0.09%

표 3. 요리동사 빈도수에 따른 상위 및 하위 순위

상위	요리동사	빈도수	빈도율(%)	하위	요리동사	빈도수	빈도율(%)
1	끓이다	575	50.67	1	각다	1	0.09
2	섞다	516	45.46	2	내리치다	1	0.09
3	썰다	480	42.29	3	늘리다	1	0.09
4	다지다	453	39.91	4	두들기다	1	0.09
5	볶다	409	36.04	5	삭히다	1	0.09

IV. 요리난이도 및 유사도 측정

재료가 흔히 사용되는 요리재료인지 그렇지 않은지에 따라 재료가 요리에 미치는 영향력은 달라진다. 이 논문에서는 확률에 기반한 엔트로피 측정법을 이용하여 재료 엔트로피를 계산하고, 레시피에 포함된 재료 엔트로피의 합을 통해 레시피 엔트로피를 계산한다.

4.1 엔트로피 측정

요리재료 i_x 의 엔트로피 $E(i_x)$ 는 요리재료 i_x 의 확률 $P(i_x)$ 를 이용하며, 식 (1)과 같이 정의한다.

$$E(i_x) = -\log P(i_x) \tag{1}$$

요리재료 i_x 의 확률 $P(i_x)$ 은 식 (2)와 같이 정의한다. $|R|$ 은 전체 레시피 개수를 의미하며, r_i 는 요리재료 i_x 가 포함된 레시피를 나타낸다.

$$P(i_x) = \frac{|\{r_i | i_x \in r_i\}|}{|R|} \tag{2}$$

레시피 r_x 의 엔트로피 $E(r_x)$ 는 레시피 r_x 에 포함된 재료 엔트로피의 합으로 계산되며, 식 (3)와 같이 정의한다.

$$E(r_x) = \sum_{i_k \in r_x} E(i_k) \tag{3}$$

4.2 요리난이도 측정

레시피 공유 사이트에서는 요리 난이도를 평가하기 위해 대부분 고급, 중급, 초급 등으로 분류하고 있다. 1135개 레시피 중에서 요리난이도가 고급인 경우는 전체 레시피 중에서 9개(0.79%), 중급은 140개(12.33%), 초급은 283개(24.93%)이고, 나머지 703개(61.94%)는 난이도 표시가 없었다. 또 레시피 중에서 ‘돼지갈비찜’은 고급으로 분류되고 있지만 ‘갈비찜’은 초급으로 분류되고 있고, ‘치즈케이크’는 고급, 중급, 초급 모든 분류에 포함되고 있어 난이도를 구분하는 기준없이 작성자

의 주관적 판단에 따라 작성되고 있다.

이 연구에서는 요리 난이도를 측정하는 하나의 척도로서 레시피 엔트로피를 이용한다. 레시피 엔트로피가 낮다는 것은 흔한 요리재료이거나 요리재료의 개수가 작다는 것을 의미하므로 요리 준비의 복잡성은 낮다는 것으로 나타낼 수 있으며, 반대로 레시피의 엔트로피가 높다는 것은 요리재료가 희귀하거나 재료의 개수가 많다는 것을 의미하므로 요리 준비의 복잡성은 높아진다.

[표 4]와 [표 5]는 요리재료를 기준으로 계산된 레시피 엔트로피 순위에 따른 레시피명과 요리재료 리스트를 보여주고 있다. 가장 엔트로피가 높은 레시피는 한식의 ‘동태해물찜’으로 총 요리재료 개수는 27개로 나타나고 있고, 가장 엔트로피가 낮은 레시피는 ‘마늘 장아찌’로 ‘간장/마늘/소금/식초’만으로 요리가 가능하다. 국가별 평균 레시피 엔트로피는 한식은 약 24, 이탈리아식은 약 40, 일식은 약 38, 중식은 약 44, 태국식은 약 49로 계산되어 한식이 가장 낮게 나타나고 있다.

표 4. 요리재료에 의한 하위 레시피 엔트로피 순위

하위	국가	레시피명	요리재료		
			엔트로피	재료수	재료 리스트
1	한국	마늘장아찌	4.22	4	간장/마늘/소금/식초
2	한국	삶은감자	4.61	3	감자/소금/설탕
3	한국	달걀간장밥	5.36	3	밥/달걀/간장
4	한국	모시조개탕	5.40	3	소금/파/조개
5	한국	전자렌지 계란찜	6.16	5	달걀/소금/깨/파/물

표 5. 요리재료에 의한 상위 레시피 엔트로피 순위

상위	국가	레시피명	엔트로피	재료수	요리재료
					재료 리스트
1	한국	동태해물찜	78.43	27	콩나물/버섯/단호박/파/떡/해물/새우/양념장/고추장/생강/간장/액젓/물엿/소금/전분/숙갓/양파/양배추/깻잎/명태/조개/고추기름/마늘/굴소스/후추/매실청/깨
2	이탈리아	베지테리안 라자냐	73.23	23	올리브오일/생강/양파/피망/버섯/토마토/퓨레/로즈마리/소금/버터/우유/파마산치즈/치즈/마늘/당근/가지/오레가노/옥수/시금치/밀가루/체다치즈/라자냐면
3	이탈리아	잠발라야	64.06	21	마늘/소금/파/파리고추/양파/파슬리/피망/닭육수/와인/조개/쌀/후추/올리브오일/소세지/사프란/샐러리/월계수잎/파프리카/핫소스/타임/토마토
4	이탈리아	라자냐	63.90	21	소세지/양파/토마토/소스/설탕/회향/소금/파슬리/치즈/파마산치즈/소고기/마늘/페이스트/물/바질/시즈닝/후추/라자냐면/달걀/모짜렐라
5	태국	타이샐러드	62.72	19	당면/목이버섯/조개/양파/샐러리/레몬/깨나리/액젓/칠리소스/고추/물/오징어/새우/토마토/오이/숙갓/마늘/핫소스/설탕

[표 6]과 [표 7]은 요리동사로 계산한 레시피 엔트로피 순위와 요리동사 리스트를 보여주고 있다. 가장 엔트로피가 높은 레시피는 한식의 '전라도식 모시송편'으로 사용동사 종류는 총 18개이지만 동일한 동사를 여러 번 사용하여 총 요리동사 개수는 29이다. 가장 엔트로피가 낮은 레시피는 '초고추장'과 '파네토네'로서 '섞다'만 1번 사용되고 있다.

표 6. 요리동사에 의한 하위 레시피 엔트로피 순위

순위	국가	레시피명	요리동사		
			엔트로피	동사수	동사 리스트 및 빈도수
1	한국	초고추장	0.79	1	섞다(1)
2	이탈리아	파네토네	0.79	1	섞다(1)
3	이탈리아	그리스풍 파스타 샐러드	1.58	2	섞다(2)
4	일본	연두부 샐러드	1.65	2	섞다(1)/샐다(1)
5	한국	야채죽발냉채	2.23	2	뿌리다(1)/섞다(1)

표 7. 요리동사에 의한 상위 레시피 엔트로피 순위

순위	국가	레시피명	요리동사		
			엔트로피	동사수	동사 리스트 및 빈도수
1	한국	전라도식 모시송편	84.52	29	깎다(2)/끓이다(1)/누르다(1)/담그다(1)/데치다(2)/떼다(1)/떨어뜨리다(1)/볶다(1)/불리다(1)/빻다(3)/뺧다(2)/빼다(1)/살다(1)/섞다(3)/짜다(2)/찌다(2)/찜다(2)/헹구다(1)
2	한국	고추장 소스 닭날개 구이	79.89	31	가르다(1)/건지다(2)/굽다(1)/끓이다(1)/누르다(2)/담그다(1)/돌리다(1)/두르다(1)/떨어뜨리다(1)/버무리다(2)/빼다(2)/스미다(1)/익히다(2)/입히다(3)/줄이다(5)/주무르다(1)/털다(1)/튀기다(3)
3	한국	흰죽	61.17	26	건지다(1)/끓이다(8)/누르다(1)/불리다(5)/섞다(1)/썰다(3)/찢다(1)/줄이다(2)/퍼지다(3)/풀다(1)
5	한국	닭고기수삼샐러드	55.41	24	건지다(3)/굽다(1)/깎다(1)/담그다(3)/돌리다(1)/뜯다(1)/문지르다(1)/바라다(1)/버무리다(1)/벗기다(2)/뿌리다(1)/섞다(1)/식히다(1)/샐다(3)/찌다(1)/찜다(1)
4	일본	아시아인 오렌지 치킨	55.26	28	굽다(1)/끓이다(4)/볶이다(1)/두르다(1)/볶다(2)/빼다(1)/살다(1)/샐다(1)/찢다(1)

요리동사는 요리방법을 표현하는 단어로서 요리동사에 의한 레시피 엔트로피가 높을수록 요리방법의 복잡성은 증가하게 된다. 국가별 평균 레시피 엔트로피는 한식은 약 16, 이탈리아식은 약 14, 일식은 약 16, 중식은 약 15, 태국식은 약 18로서 요리방법은 태국이 가장 복잡하였고, 한식은 중간으로 나타나고 있다. 요리동사

에 의한 레시피 엔트로피는 나라별로 차이가 크지 않았다.

[그림 2]는 레시피 엔트로피의 분포도를 보여주고 있다. x값이 증가할수록 총 요리재료 개수 또는 희귀한 요리재료가 많아지므로 요리준비의 복잡도는 증가한다. y값이 증가할수록 총 요리동사 개수 또는 잘 사용되지 않는 요리동사를 사용하므로 요리방법의 복잡도는 증가한다.

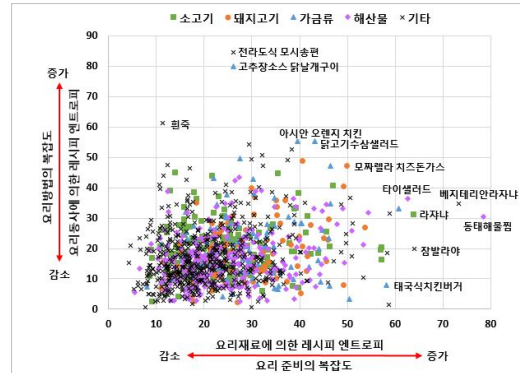


그림 2. 레시피 엔트로피 분포도

예를 들어 '흰죽'은 6가지 재료만 필요하여 요리준비는 간편하지만, 26번의 요리동사가 필요하여 요리방법은 복잡하다. 요리준비와 요리방법의 전체 관점에서 요리의 난이도를 평가하기 위해 커버 레벨을 이용한다. 커버 레벨 1은 어떤 좌표(x,y) 범위에도 포함되지 않는 좌표를 말하며, 가장 테두리에 위치하는 좌표들이다. 커버 레벨 1에 속하는 좌표들을 제거한 후, 다음 커버 레벨에 해당하는 좌표들을 찾는 것을 반복한다. 요리복잡도의 관점에서 레벨 1은 가장 난이도가 높은 요리들이며, 레벨 숫자가 커질수록 요리난이도는 쉬워진다.

표 8. 커버 레벨에 의한 요리난이도 상위 레시피

순위	커버 레벨	국가	레시피명	요리재료		요리동사	
				재료수	엔트로피	동사수	엔트로피
1	1	한국	동태해물찜	27	78.43	19	30.47
		이탈리아	베지테리안라자냐	23	73.23	24	34.86
		태국	타이샐러드	19	62.72	20	36.30
		한국	청국장 시리얼 쿠키	14	50.63	19	36.52
		일식	모짜렐라 치즈 돈가스	17	49.97	17	47.29
		이탈리아	곱창전골	18	45.76	24	49.08
		한국	닭고기수삼샐러드	14	43.21	24	55.41
		한국	전라도식 모시송편	8	26.12	29	84.52

표 9. 커버 레벨에 의한 요리난이도 하위 레시피

하위	커버 레벨	국가	레시피명	요리재료		요리동사	
				재료수	엔트로피	동사수	엔트로피
1	66	한식	초고추장	3	7.78	1	0.79
2	65	이탈리아	두부스테이크	5	9.16	3	2.50
3	64	한식	김치쌈밥	2	9.70	1	2.52
4	63	한식	단팥죽	3	10.33	4	2.72
5	62	한식	달걀간장밥	3	5.36	3	6.52
		이탈리아	모시조개탕	3	5.41	4	5.40
		한식	고구마스틱	4	8.40	3	5.03
		한식	새송이버섯전	3	9.42	3	4.84
		한식	콩밥	3	10.60	3	4.15
		한식	호두김주먹밥	5	11.59	2	2.84

4.3 유사도 측정

이 논문에서는 두 레시피 r_x 와 r_y 의 유사도 $Sim(r_x, r_y)$ 을 다음과 같이 정의한다. 두 레시피의 공통 요리재료가 많을수록 그리고 엔트로피가 높은 재료를 공유할수록 두 레시피의 유사도는 높아진다.

$$Sim(r_x, r_y) = \frac{\sum_{i_k \in \{r_x \cap r_y\}} E(i_k)}{\sum_{i_k \in \{r_x \cup r_y\}} E(i_k)} \quad (4)$$

표 10. 공통 요리재료에 의한 레시피 유사도 및 순위

순위	레시피 및 재료수	레시피 및 재료수	공통 요리재료		유사도
			요리재료 리스트	재료수	
1	고구마맛탕(4)	고구마스틱(4)	살랑/고구마/소금/물/	4	1
2	콩나물무침(6)	콩나물무치기(6)	마늘/깨/소금/파/콩나물/참기름/	6	1
3	도라지무침(10)	도라지오이생채(10)	마늘/설탕/깨/고추가루/소금/파/도라지/식초/오이/참기름/간장/설탕/계피/소금/대추/참쌀/물/밤/자/참기름/	10	1
4	약식(10)	전기밥솥약식(10)	콩/쌀/물/	3	0.91
6	도라지무침(10)	도라지생채(9)	마늘/설탕/깨/고추가루/소금/파/도라지/식초/참기름/	9	0.84
7	도라지생채(9)	도라지오이생채(10)	마늘/설탕/깨/고추가루/소금/파/도라지/식초/참기름/	9	0.84
8	고사리나물(7)	불린 고사리나물(9)	간장/마늘/깨/고사리/파/물/참기름/간장/마늘/소고기/설탕/깨/소금/파/후추/송이버섯/참기름/간장/마늘/깨/액젓/꽃게/고추가루/파/생강/양파/물엿/후추/참기름	7	0.81
9	송이산적(10)	완자꼬치(12)		10	0.78
10	양념계장(15)	양념계장떡(12)		12	0.74

[표 10]은 레시피 유사도의 순위에 따른 레시피와 공통 요리재료 리스트를 보여주고 있다. ‘고구마맛탕’과 ‘고구마스틱’, ‘콩나물 무침’과 ‘콩나물무치기’, ‘도라지무침’과 ‘도라지오이생채’, ‘약식’과 ‘전기밥솥 약식’으로 모두 유사도가 1이며, 실제 유사한 요리인 것을 알 수 있다. 그러나 레시피명은 등록시 사용자에게 의해 직접 작성되기 때문에 동일한 요리라도 레시피명이 다를 경우 이름만으로는 구별하기가 쉽지 않다. 레시피 유사도는 실제 같은 요리인지 아니면 유사한 요리인지 구별할 수 있는 객관적 정보가 되며, 유사도를 통해 어떤 요리와 동일하거나 유사한 요리들을 쉽게 검색할 수 있게 한다.

V. 레시피 연결망 생성

이 연구에서는 레시피 유사도에 기반한 연결망 구성을 위해 임의의 레시피 r_x 를 정점으로 하고, 두 레시피 r_x 와 r_y 의 유사도 $Sim(r_x, r_y)$ 을 간선의 가중치로 하는 무방향 네트워크 $G_R(V, E)$ 를 구축한다.

5.1 최소비용신장트리에 의한 레시피 연결망

[그림 3]은 레시피 연결망을 최소비용신장트리(MST, Minimum Spanning Tree)를 이용하여 생성하였고, 정점의 색상과 모양으로 나라별 레시피를 구분하고 있다.

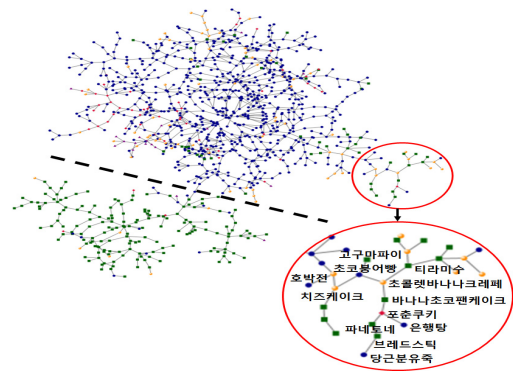


그림 3. 최소비용신장트리 기반의 레시피 연결망 (파란색 원은 한국, 핑크색 다이아몬드는 중국, 오렌지 구는 일본, 보라색 삼각형은 태국, 초록색 사각형은 서양 요리를 표시)

연결망에서 점선을 기준으로 아시아(한국, 일본, 중국, 태국) 요리는 대부분 위쪽에 분포되고 서양(이탈리아, 멕시코 등)은 아래쪽에 분포되고 있어 유사도 관점에서 차이가 뚜렷하게 구별되고 있었다. 그러나 아시아 요리 안에서 한국 레시피와 비교하여 중국, 일본, 태국 레시피는 나라별로 일부 모여있지만 한식 레시피와 근접거리에 위치하였고, 특히 일본 레시피는 대부분 한식과 연결되고 있어 유사성이 매우 높은 것으로 나타났다. 일본 레시피 중에서 ‘고로케’와 ‘쇠고기참깨볶음’은 한식 레시피에도 이와 유사한 ‘감자뽕’과 ‘쇠고기볶음’이 있고, 한식 레시피의 ‘고추잡채’와 ‘생선완자탕’은 실제 중국요리에 포함된다. 대부분 웹에서 레시피를 등록할 때 어느 나라의 요리인지는 작성자의 주관적 기준에 따라 선택되기 때문에 레시피 연결망을 이용하여 어느 나라의 요리들과 유사한지 객관적 기준을 제시할 수 있다.

5.2 유사도 순위에 의한 레시피 연결망

유사도 순위에 따라 간선을 연결할 때, 유사도 임계값에 따라 떨어져 있는 레시피들이 하나의 그룹으로 연결되기도 한다. 예를 들어 유사도 임계값이 1일 경우, ‘고구마맛탕’과 ‘고구마스트’, ‘도라지무침’과 ‘도라지오이생채’ 등 레시피명은 다르지만 실제 동일한 요리들이 하나의 그룹으로 연결된다.

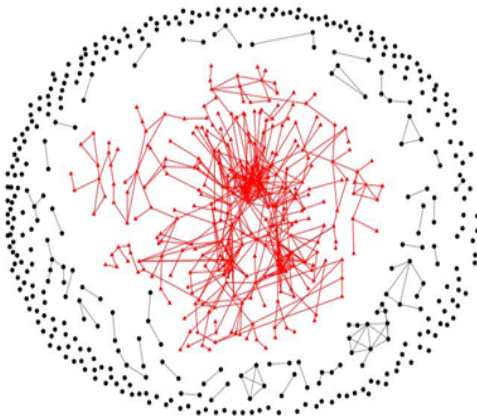


그림 4. 유사도 $Sim(r_x, r_y) \geq 0.36$ 일 때, 한식 레시피 연결망

[그림 4]는 유사도 임계값 0.36일 때 한식 레시피 연결망을 보여주고 있다. 302개의 레시피가 상호연결된 그룹 1개와 2개 또는 그 이상의 레시피들이 연결된 소규모 그룹들, 그리고 어떤 노트와도 연결되지 않은 하나의 레시피로만 구성된 328개의 레시피를 보여주고 있다. 주어진 요리재료로 할 수 있는 식단 메뉴나 요리 추천시스템에 관한 연구들은 오랫동안 진행되어왔다. 만약 추천된 요리를 사용자가 만들기 어렵다면 다른 대안은 무엇인가? 레시피 연결망을 통해 추천된 ‘송이산적’이 어렵다면 유사한 ‘오이쇠고기볶음’, ‘완자꼬지’, ‘쇠고기청고추찜’ 등 대체할 수 있는 다양한 레시피를 제시할 수 있을 것이다.

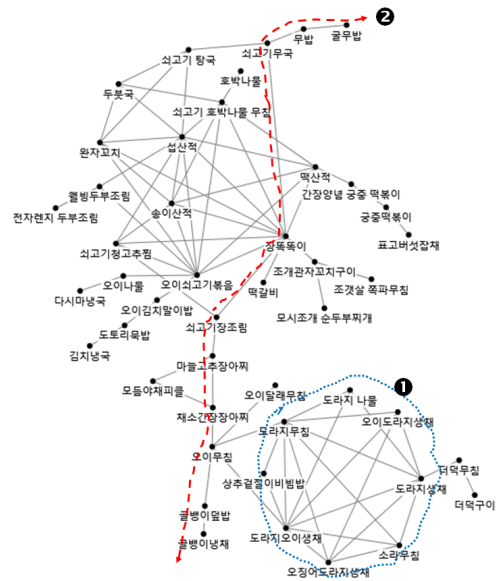


그림 5. 유사도 $Sim(r_x, r_y) \geq 0.43$ 일 때, 한식 레시피 연결망

[그림 5]는 유사도 임계치 0.43일 때 연결된 하나의 레시피 그룹을 보여주고 있다. 만약 오늘의 요리로 ‘도라지생채’를 하였다면 내일은 똑같은 요리가 아닌 다른 요리를 원할 것이다. 그러나 요리재료가 전혀 다른 새로운 요리를 계획할 경우 모든 재료를 새로 구입해야 하므로 요리준비는 어려워지지만, 기존 재료로 가능한 요리를 계획한다면 손쉽게 요리준비를 할 수 있으므로

준비용이성은 높아지게 된다. [그림 12]의 ①은 ‘도라지생채’에서 일부 재료의 변경으로 ‘도라지나물’, ‘소라무침’, ‘오이도라지생채’, ‘오징어도라지생채’ 등으로 쉽게 요리를 바꿀 수 있다는 것을 보여준다. [그림 12]의 ②는 ‘굴무밥’에서 ‘골뱅이냉채’까지의 최단경로에 포함된 레시피 순서를 보여준다. 만약 일주일동안의 요리 스케줄을 계획한다면 다양한 요리를 손쉽게 준비하기를 원할 것이다. 최단경로는 요리방법은 다르지만 일부 재료의 변경으로 다양한 요리로 변형할 수 있는 예를 보여주고 있다.

VI. 결론

기존 요리 추천 시스템은 특정 재료나 요리법을 포함하는 레시피를 찾는데 목적을 두고 있다. 그러나 사용자의 관점에서는 특정 재료로 할 수 있는 요리 추천뿐만 아니라, 요리난이도 측면에서 준비하기 쉬운 요리 또는 요리절차가 쉬운 요리를, 또 추천된 요리가 싫다면 유사한 대체 레시피를 원할 것이다. 만약 한식 스타일을 원한다면 단순히 요리이름이 한국식이 아닌 요리 내용에서 한식과 유사한 요리를 원할 것이다.

이 연구는 엔트로피 개념을 이용하여 주관적으로 분류되고 있는 요리 난이도를 정량적 값으로 객관화시켰고, 레시피의 유사성을 측정하여 레시피 연결망을 생성하였다. 연결망 분석을 통해 요리의 특징과 활용적 측면을 탐구하였다. 레시피 연결망에서 거리공간을 통해 국가별 유사도가 높은지 낮은지, 유사도가 높은 레시피들은 무엇인지 확인할 수 있으며, 상호관련성이 높은 유사그룹을 확인할 수 있었다. 또한 두 레시피 간의 최단거리를 이용하여 복수의 요리를 준비해야하는 식단을 구성할 때 가장 적절한 재료와 요리과정에 드는 노력을 최소화할 수 있는 방법을 제시하였다.

향후 연구에서는 이번 연구에서 고려하지 못했던 요리의 주재료·부재료 구분과 방향성을 고려한 연결망을 진행할 예정이다. 요리 A의 재료로 요리 B를 만들 경우 재료를 추가하는 것과 반대인 재료를 빼는 것은 대칭적인 난이도가 아니다. 두 요리의 유사성은 무방향 예지가 아니라 방향이 있는 예지로 표시되어야 할 것이

다. 이 방향성을 고려하면 보다 현실적인 요리의 변화 방향에 따른 난이도를 구할 수 있을 것으로 기대되며, 기존의 요리에서 새로운 요리가 생성되는 과정을 모형화하기 위한 연구를 진행할 것이다.

참고 문헌

- [1] J. Freyne and S. Berkovsky, “Intelligent food planning: Personalized recipe recommendation,” Proc. of Intelligent User Interfaces, pp.221-324, 2010.
- [2] G. Geleijnse, P. Nachtigall, P. V. Kaam, and L. Wijgergangs, “A personalized recipe advice system to promote healthful choices,” Proc. of 16th Intelligent User Interfaces, pp.437-438, 2011.
- [3] F. Kuo, C. Li, M. Shan, and S. Lee, “Intelligent menu planning: Recommending set of recipes by ingredients,” Proc. of the ACM Multimedia Workshop on Multimedia for Cooking and Eating Activities, pp.1-6, 2012.
- [4] S. E. Ahnert, “Generalized power graph compression reveals dominant relationship patterns in complex networks,” Scientific Reports 4, No.4385, pp.1-6, 2014.
- [5] C. A. Bliss, Morgan R. Frank, Christopher M. Danforth, and Peter S. Dodds, “An evolutionary algorithm approach to link prediction in dynamic social networks,” Journal of Computational Science, Vol.5, pp.750-764, 2014.
- [6] A. Papadimitriou, P. Symeonidis, and Y. Manolopoulos, “Fast and accurate link prediction in social networking systems,” The Journal of Systems and Software, Vol.85, pp.2119-2132, 2012.
- [7] L. Wang, Q. Li, N. Li, G. Dong, and Y. Yang, “Substructure similarity measurement in Chinese recipes,” Proc. of World Wide Web, pp.979-988, 2008.
- [8] C. Teng, Y. Lin, and L. Adamic, “Recipe

recommendation using ingredient networks,” Proc. of 4th Annual ACM Web Science Conference, pp.298-307, 2012.

[9] S. D. Kim, Y. J. Lee, S. H. Kang, H. G. Cho, and S. M. Yoon, “Constructing cookery network based on ingredient entropy measure,” Indian Journal of Science and Technology, Vol.8, No.23, pp.1-9, 2015.

[10] Y. Shidochi, T. Takahashi, I. Ide, and H. Murase, “Finding replaceable materials in cooking recipe texts considering characteristic cooking actions,” Proc. of 9th ACM Multimedia Workshop on Multimedia for Cooking and Eating activities, pp.9-14, 2009.

[11] Y. Yeol Ahn, S. E. Ahnert, James P. Bagrow, and Albert-L. Barabasi, “Flavor network and the principles of food pairing,” Nature Scientific Reports, pp.1-21, 2011.

[12] M. Ueda, S. Asanuma, Y. Miyawaki, and S. Nakajima, “Recipe recommendation method by considering the user’s rreference and ingredient quantity of target recipe,” Proc. of the International Multi-Conference of Engineers and Computer Scientists, Vol.1, pp.12-14, 2014.

저 자 소 개

김 수 도(Su-Do Kim)

정회원



- 1995년 : 부경대학교 전자계산학과(이학사)
- 2001년 : 부경대학교 교육대학원 전산교육전공(교육학석사)
- 2008년 : 부경대학교 정보시스템 협동과정(이학박사)

▪ 2012년 ~ 현재 : 부산대학교 사회급변현상연구소 전임연구원

<관심분야> : 웹 콘텐츠, 사회연결망, 융합 콘텐츠

이 윤 정(Yun-Jung Lee)

정회원



- 1995년 2월 : 부경대학교 전자계산학과(이학사)
- 1999년 2월 : 부경대학교 전산정보학과(이학석사)
- 2008년 8월 : 부경대학교 전자계산학과(이학박사)

▪ 2008년 9월 ~ 2014년 2월 : 부산대학교 박사후연구원
 ▪ 2015년 3월 ~ 현재 : 부산대학교 사회급변현상연구소 전임연구원

<관심분야> : 얼굴 애니메이션, 웹 콘텐츠 시각화, 사회연결망 분석

윤 성 민(Seong-Min Yoon)

정회원



- 1983년 : 고려대학교 경제학과(경제학사)
- 1985년 : 고려대학교 일반대학원 경제학과(경제학석사)
- 1989년 : 고려대학교 일반대학원 경제학과(경제학박사)

▪ 1991년 ~ 2008년 : 부경대학교 경제학부 교수
 ▪ 2008년 ~ 현재 : 부산대학교 경제학부 교수

<관심분야> : 계량경제학, 경제물리학

조 환 규(Hwan-Gue Cho)

정회원



- 1984년 : 서울대학교 계산통계학과(이학사)
- 1986년 : KAIST 대학원 전산학과(공학석사)
- 1990년 : KAIST 대학원 전산학과(공학박사)

▪ 1990년 3월 ~ 현재 : 부산대학교 전기컴퓨터공학부 교수

<관심분야> : 계산이론, 생물정보학