

머신러닝을 이용한 의사결정트리 기반의 식품교환표 구성 모델¹⁾

김지윤*, 이상민**, 전형준***, 김가은***, 김지현*, 박나은†, 진창균***,
권진영††, 김종완†

삼육대학교 {*식품영양학과, **컴퓨터·메카트로닉스 공학부, † 스미스학부대학}

†† 건국대학교 정보통신경영학과

**삼육대학교 인공지능·빅데이터 소프트웨어코드 연구소

† 성신여자대학교 융합보안공학과

wldbs3592@naver.com

Food Exchange Table Organization Model Based on Decision Tree Using Machine Learning

JiYun Kim*, Sangmin Lee**, Hyeongjun Jeon***, Gaeun Kim*** Ji-Hyun Kim**,
Naeun Park† ChangGyun Jin***, Jin young Kwon†, Jongwan Kim†

요 약

최근 국내에서는 식품에 대한 관심이 높아짐에 따라 먹거리에 건강·환경·미래지향적 가치가 부여되고 있으며 식품 산업에서도 신규 식품 개발이 증가하는 추세이다. 식단을 구성할 때 기준이 되는 식품교환표는 개정과정에서 많은 인력과 시간이 소요되기 때문에 식품 섭취 변화를 신속하게 반영하기 어렵다. 본 논문에서는 식품교환표의 활용도를 높이기 위한 식품교환표 갱신 기법을 제안한다. 제안 기법은 의사결정트리 모델을 학습하여 새롭게 추가된 식품의 정보를 바탕으로 식품군을 분류하여 식품교환표를 갱신한다. 이는 영양 관리가 필요한 당뇨병 환자 등에게 실용적이며 기호성·다양성이 높은 식단을 구성하는 데 도움을 준다.

1. 서론

경제성장과 소득의 증가로 국내에서는 먹거리에 대한 관심이 높아지고 있다. 식품은 단순히 허기를 채우는 것뿐만 아니라 개인의 기호를 보여주고 건강을 관리하는 수단으로 자리매김하고 있다.

식품 소비에서도 식품의 질적인 성장이 크게 나타났으며 식품의 간편화, 안전 지향, 고급화/다양화 트렌드가 나타났다 [1]. 따라서 식품에 다양한 가치를 부여한 새로운 식품에 대한 수요가 증가하는 추세이다. 대체 식품·메디푸드·고령친화식품 등 맞춤형·특수식품이 꾸준히 개발되고 있으며 건강 개선과 더불어 환경을 생각하는 미래지향적 가치를 지닌 식품들이 주목받고 있다. [2].

신규 식품 개발이 지속해서 이루어지는 데 반해 이를 식단에 적용할 수 있는 방법에 대한 연구는 미비한 실정이다. 영양 관리가 필요한 식단을 작성하

는 자료로는 식품교환표(Food exchange system)가 있다. 식품교환표를 사용하면 필요 영양소를 충족시킬 수 있는 균형 있는 식사 섭취가 가능하여 영양사 등 전문가들의 식단 작성과 영양 관리를 필요로 하는 당뇨병 환자의 교육수단에 활용된다.

식품교환표를 통해 새로운 식품을 활용하여 식단을 작성하는 것에는 한계가 있다. 왜냐하면 식품교환표를 개정하는데 이용되는 국민건강영양조사는 조사 및 공표에 약 1년이 소요되며, 많은 인력과 물적 비용을 수반하기 때문이다 [3]. 특히, 식품 목록을 추가하는데 몇 달간을 설문조사에 소요하기 때문에 한국인의 식품 섭취 변화를 신속하게 반영하기 어렵다는 단점을 가지고 있다.

기계학습은 인공지능의 한 분야로, 모델이 빅데이터의 패턴과 특징을 분석하여 학습하고 학습한 모델이 새로운 데이터의 특징값을 바탕으로 데이터의 클래스를 분류하는 방법이다 [4]. 식품 산업에서도 빅데이터 분석과 같은 ICT 기술을 접목하는 시도가 이루어지고 있으며, 기계학습이 시간·인력 소요에서 수작업을 보완할 수 있다고 판단하여 본 논문에서는

1) 이 논문은 2020년도 대한민국 교육부와 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2018S1A5A8027993).

† 교신저자 (kimj@syu.ac.kr)

기계학습에서 의사결정트리 모델을 선정하여 식품 분류를 수행하도록 하였다.

본 논문에서는 의사결정트리 모델을 학습하여 새로 추가된 식품의 식품군을 분류하여 식품교환표를 갱신함으로써 사용자가 식단 구성에 활용할 수 있도록 한다. 본 논문의 구성은 다음과 같다. 2장에서는 관련 연구에 관해 설명하고, 3장에서는 본 연구의 식품군 분류 모델이 학습되는 과정에 관해 서술한다. 4장에서는 의사결정트리가 실제 분류한 식품군을 확인하고 결과를 해석한다. 5장에서는 결론 및 향후 연구 방향을 기술한다.

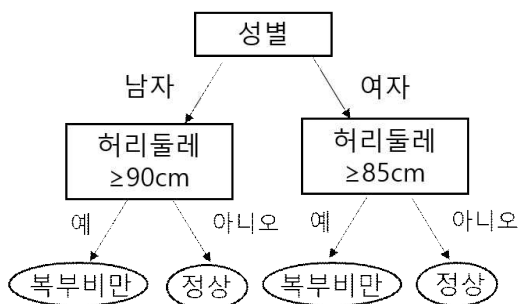
2. 관련연구

2. 1 식품교환표

식품교환표는 식단을 구성할 때 영양 관리를 위해 사용되는 도구이며, 같은 식품군 안에서 같은 교환단위끼리 바꾸어 먹을 수 있다는 개념이다. 식품교환표의 식품군에는 곡류군, 어육류군, 채소군, 우유군, 과일군, 지방군이 있으며, 어육류군은 지방함량에 따라 저지방, 중지방, 고지방으로 세분된다. 균형 있는 식사를 계획하기 위해 식품 교환단위를 계산하는 것이 필요하므로, 식품교환표는 영양 상담 프로그램 개발 [5]과 웹 기반 자기 영양 관리 프로그램 [6]등의 연구에서 활용되고 있다.

2. 2 의사결정트리(Decision Tree)

의사결정트리는 그림 1과 같이 데이터의 패턴을 인식하여 각 노드가 “IF~THEN”형식으로 이루어진 트리구조를 통해 주어진 문제를 해결하는 방법이다. 직관적인 분할정보 알고리즘을 이용해서 복잡한 문제를 순차적으로 해결한다. 의사결정트리 모델을 학습시켜 자동차 구매 패턴 분류 [7], 반도체 공정의 제품 이상 분류 [8], 식품의 품질 분류 [9] 등 다양한 분야에 활용된다.

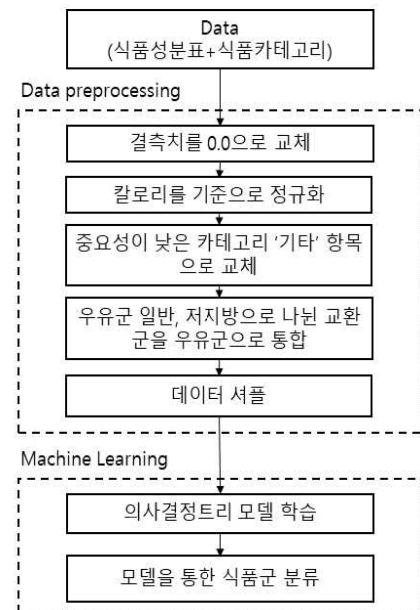


(그림 1) 의사결정트리의 예시

3. 식품교환표 구성 모델

수작업을 통한 식품교환표의 식품군 분류는 시간과 인력이 소요되므로 사용자가 식단을 구성하는데 제약이 생길 수 있었다. 본 연구는 음식 데이터를 바탕으로 식품군을 예측하는 의사결정트리 기반의 인공지능 모델을 학습한다. 식품군 분류를 위해 의사결정트리를 적용하였다. 의사결정트리는 의사결정 과정을 직관적으로 제공하는 추론 규칙으로 구성되어 있어 식품군 분류 과정을 확인하기 적합하다.

의사결정트리를 사용하여 식품군을 결정하는 순서도는 그림 2와 같다.



(그림 2) 식품군 분류 순서도

3.1 데이터 전처리

식품의 탄수화물, 단백질, 지방, 총 당류, 총 식이섬유 함량과 카테고리 정보를 데이터로 사용한다. 식품군 분류에서 데이터는 ‘당뇨병 영양 관리와 식품교환표’ 식품교환군별 e-식품교환량ver.2 [10]과 농촌진흥청 국가표준 식품성분표 제9 개정판을 참고하여 훈련 데이터로 하여 학습하였고, 2010년도 식품교환표 식품 목록을 검증 데이터로 하여 모델의 정확도를 확인하였다. 단, 식품교환표에는 세부적인 영양소가 제시되어있지 않아 식품 성분은 농촌진흥청 국가표준 식품성분표 제9 개정판을 참고하였다.

전처리 과정 중 총 식품성분표에서 조사되지 않아 발생한 결측치는 학습에 영향을 주지 않기 위해 0.0으로 교체하였다. 칼로리 (100kcal)를 기준으로 탄수화물, 단백질, 지방, 총 식이섬유, 총 당류 정보를 정규화하여 각 영양소 비율을 구하고 해당 비율에 맞

<표 1> 데이터 전처리 결과

식품 명	에너지(kcal)	단백질(g)	지방(g)	탄수화물(g)	총 당류(g)	총 식이섬유(g)	카테고리	식품군
귀리, 겉귀리, 도정, 생것	100.0	3.000000	0.059701	22.507463	0.0	4.029851	곡류 및 그 제품	곡류군
귀리, 오토밀	100.0	2.688312	0.103896	22.584416	0.0	2.077922	곡류 및 그 제품	곡류군
...								
토마토 소스, 케첩	100.0	11.780822	0.753425	36.438356	0.0	0.000000	조미료류	과일군

는 식품군으로 분류될 수 있게 하였다. 전체 카테고리 중 중요성이 낮은 카테고리는 기타 항목으로 교체하였다. 총 8가지의 식품군으로 식품을 분류하였고, 2010 식품교환표 개정에서 세분화했던 저지방 우유군과 일반 우유군은 개별 데이터 수가 적으므로, 유의미한 학습을 위해 우유군으로 통합시켜 분류할 수 있게 하였다. 또한, 모델의 일반화 성능을 높이기 위해 데이터 내부 항목들의 순서를 섞었다.

전처리된 데이터는 표 1과 같다. 의사결정트리의 학습에는 탄수화물, 단백질, 지방, 총 당류, 총 식이 섬유, 카테고리 정보를 데이터의 속성으로 하고, 식품군을 학습의 목표변수로 하였다. 검증은 데이터의 속성을 가지고 목표변수를 예측하는 방식으로 식품군 분류를 수행한다.

3.2 의사결정트리 학습

본 연구의 의사결정트리는 앞서 전처리한 훈련 데이터를 통해 학습된다. 의사결정트리가 학습되며 생길 수 있는 가지의 깊이는 최대 40으로 설정했으며, 세부적인 파라미터는 표 2와 같다. 학습 데이터를 통해 생성된 식품군 분류 모델은 검증 데이터를 통해 완성도가 평가된다.

<표 2> 의사결정트리의 파라미터

Parameter	Value
criterion	gini
splitter	best
max_depth	40
min_samples_split	2
min_samples_leaf	1
min_weight_fraction_leaf	0.0
max_features	None
random_state	55
max_leaf_nodes	None
min_impurity_decrease	0.0
min_impurity_split	0
class_weight	None
presort	deprecated
ccp_alpha	0.0

4. 실험

제안 방법의 실험 환경은 표 3과 같으며, 3장에서 생성한 학습 데이터 1192개를 통해 의사결정트리를 생성하였다.

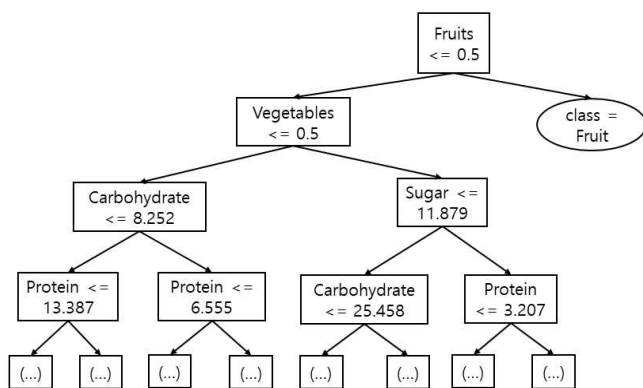
<표 3> 실험환경

CPU	Intel Xeon Silver 4114 2.20GHz * 2EA
RAM	128GB
VGA	RTX 2080 ti * 2EA

학습데이터로 의사결정트리를 학습한 결과는 본 논문에서는 지면 관계상 그림 3과 같이 나타내었다.

<표 4> 식품군 분류 결과

식품 명	단백질(g)	지방(g)	탄수화물(g)	총 당류(g)	총 식이섬유(g)	카테고리	식품군	예측 값
멥쌀밥, 백미	1.973684	0.065789	21.84211	0	0	곡류 및 그 제품	곡류군	곡류군
멥쌀밥, 현미	1.976048	0.11976	22.33533	0	0	곡류 및 그 제품	곡류군	곡류군
멥쌀죽, 백미	1.971831	0	21.97183	0	0	곡류 및 그 제품	곡류군	곡류군
멥쌀, 백미, 생것	1.763085	0.110193	21.90083	0	0	곡류 및 그 제품	곡류군	곡류군
...								
프루트카테일, 통조림	0.253333	0.04	24.69333	25.2	1.6	과실류	과일군	과일군
복숭아, 백도, 통조림	0.430769	0.030769	24.67692	21.87692	3.230769	과실류	과일군	과일군
복숭아, 황도, 통조림	0.382716	0.024691	24.62963	23.22222	3.08642	과실류	과일군	과일군
파인애플, 통조림, 고형물,	0.397436	0.051282	24.58974	20.26923	2.179487	과실류	과일군	과일군



(그림 3) 의사결정트리의 결정 과정

생성된 모델에 검증데이터를 넣어서 예측한 결과 정확도는 95.37%로 나타났다. 잘못 분류된 항목들은 검증데이터를 구축할 때 2010년도 식품교환표에 제시되어있지 않은 식품 성분을 농촌진흥청 국가표준 식품성분표 제9 개정판에서 참고하였기 때문에 측정 시기에 따른 영양소 차이가 있었기 때문으로 보인다.

의사결정트리 모델이 검증데이터를 분류한 결과는 표 4와 같다. 분류한 식품군은 총 8개이며 각각 곡류군 55개, 과일군 51개, 어육류군 저지방 52개, 어육류군 중지방 20개, 어육류군 고지방 14개, 우유군 4개, 지방군 29개, 채소군 84개에서 예측값과 식품군이 일치하였다.

5. 결론

본 연구에서는 의사결정트리를 사용한 식품교환표 식품 추가 방법을 제안하였다. 제안된 방법으로 식품을 분류한 결과 95.37%의 정확도를 확인할 수 있었다. 인공지능이 분류한 식품 분류 결과가 높은 정확도를 보이므로 수작업으로 분류한 식품 목록과 비

교하여 유의미한 사용이 가능하다고 볼 수 있다. 의사결정트리를 통한 식품 분류는 수작업 업무를 줄여 기존 식품교환표 식품 분류에서 소요되었던 시간과 비용을 줄일 수 있다. 또한 이전에 식품교환표에 포함되지 않았던 식품을 신속하게 추가할 수 있기 때문에 개인의 선호가 반영된 식단을 구성하는 데 도움을 줄 수 있다.

본 연구에서는 의사결정트리를 사용하여 식품을 분류하였지만, 향후에는 다양한 머신러닝 모델을 사용하여 정확도를 높이는 연구로 확장할 것이다.

참고문헌

- [1] 이계임, 김상효, 허성운, “한국인의 식품소비 심층 분석”, 한국농촌경제연구원 기본연구보고서, 2016.
- [2] 홍석인, “미래지향적 식품산업 R&D 추진전략”, 식품과학과 산업, 2020.
- [3] Ju DL, Jang HC, Cho YY, Cho JW, Yoo HS, Choi KS, Woo MH, Sohn CM, Park YK, Choue RW, “Korean Food Exchange Lists for Diabetes: Revised 2010”, J Korean Diabetes, 12(4), pp. 228-244, 2011.
- [4] 문성은, 장수범, 이정혁, 이종석, “기계학습 및 딥러닝 기술동향”, 한국통신학회지 (정보와통신), 33(10), pp. 49-56, 2016.
- [5] 한지숙, 이숙희, “당뇨병 환자를 위한 전산화된 영양상담 시스템”, 한국식품영양과학회지, 22(6), pp. 734-742, 1994.
- [6] 안운, 배제현, 김희선, “당뇨환자를 위한 웹 기반의 유헬스 자가영양관리 프로그램 개발”, 대한지역사회영양학회지, 19(4), pp. 372-385, 2014.
- [7] 이병엽, 박용훈, 유재수, “의사결정트리를 통한 자

동차산업의 구매패턴분류”, 한국콘텐츠학회논문지, 10(2), 372-380, 2010.

[8] 손지훈, 고종명, 김창욱, “반도체 공정의 이상 탐지와 분류를 위한 특징 기반 의사결정 트리”, 산업공학, 22(2), 126-134, 2009.

[9] Velásquez, Lía, et al. "An application based on the decision tree to classify the marbling of beef by hyperspectral imaging", Meat Science 133, pp. 43-50, 2017.

[10] 홍순명, “당뇨병 영양관리와 식품교환표”, 울산대학교출판부, 2019.