

과실 생산량과 기상요소간의 연관분석 마이닝

우중선^O, Erdenbileg Batbattar^{*}, 류근호^{*}

^{O*}충북대학교 컴퓨터학과

e-mail:jswoo@dblab.chungbuk.ac.kr^O, eegii@dblab.chungbuk.ac.kr^{*}, khryu@dblab.chungbuk.ac.kr^{*}

Association rule Mining between Climate factors and Fruits yields

Jong-Seon Woo^O, Erdenbileg Batbaatar^{*}, Keun-Ho Ryu^{*}

^{O*}Dept. of Computer Science, Chungbuk National University

● 요약 ●

이 논문에서는 기후조건과 농업 생산량을 포함하는 농업/기상 데이터에 데이터 마이닝의 연관규칙 기법을 적용하여 농업 생산의 기반이 되는 기후요인들과 생산량 간의 연관성을 분석하고자 한다. 기후 속성들의 값을 포함하고 있는 기상청 기후 데이터와 농업 생산량을 포함하는 통계청의 데이터를 통합 한 후 기후 속성들의 값을 이산화 하여 연관규칙 기법을 적용한다. 실험 결과 각 기후요소들과 생산량 간의 연관 규칙을 표현 할 수 있었다. 이를 통해 기후조건 변화에 따른 농업생산기반 취약성을 예방하는 지표를 마련하고 농업 생산성 향상에 기여 할 수 있을 것으로 기대한다.

키워드: 농업 생산량(Agricultural yield), 연관규칙(Association rule), 기후조건(Climature factors), 기상 데이터(Meteorological data)

I. 서론

최근 기후 변화가 산업에 미치는 영향이 관련 연구를 통해 밝혀지고 있다. 미국 국립 기후 자료 센터(NOAA) 연구조사에서 농업, 건설, 에너지, 레저 산업이 기후와 밀접한 연관성을 가지고 있는 것으로 조사되었다[1]. 특히, 농업 경영 측면에서 기후는 농업 작물 생산량의 주요 요인이다. 농업에서 정의하고 있는 기후 요인은 자연적인 요인과 인공적인 요인이 있으며, 자연적인 요인인 기온, 습도, 강수량, 일조량, 일사량이다. 인공적인 요인은 비료와 개량된 토양품질과 같은 요인이다.

온실에서는 인간의 직접적 조정을 통해 작물의 생육환경이 조정 될 수 있다 [2]. 하지만 많은 작물들은 노지에서 재배되고 있으며 농업생산량은 기후 변화에 많은 영향을 받는다[3].

따라서, 이 논문에서는 기상청으로부터 획득한 경북 지역의 기온, 습도, 일교차, 강수량, 적설량, 습도, 풍속, 일조량, 일사량, 운량 데이터와 통계청으로부터 획득한 농업 생산량 데이터를 통합하고 데이터 마이닝의 연관 규칙 기법을 적용하여 기후조건과 작물 생산량 간의 연관성을 분석하고자 한다.

이 논문의 구성은 다음과 같다. 2장에서는 기존에 연구되어진 관련연구의 국내외동향에 대해서 기술하였다. 3장에서는 수집 된 데이터 집단 및 연관 규칙 알고리즘에 대해서 기술하였으며, 4장에서는 실험결과 및 결과해석을 기술하였다. 마지막으로 5장에서는 이 논문의 결론을 맺는다.

II. 관련 연구

1. 농업 생산성의 기후와의 상관관계

1.1 국내 동향

최근 작물 생산량에 관한 통계적 연구가 활발히 수행 되고 있다[4]. 특히, 기상요소가 식량 작물의 생산량에 미치는 영향에 대해 회귀 분석을 실시하고 작물 생산량과 기후 인자들 간의 피어슨 상관 계수를 통하여 상관성을 표현하였다. 이 연구의 경우 통계적 모델은 적용함으로써 각 인자간의 독립적인 영향을 고려하는 한계를 가지고 있다. 따라서 이 논문에서는 데이터마이닝의 연관 규칙을 적용하여 단일 속성 간에 영향이 아닌 복합적 속성의 요인을 고려하였다.

1.2 해외 동향

해외에서도 기후의 변화에 따른 작물 생산성의 변화에 대한 연구가 활발히 진행되고 있다[5]. 이 연구에서는 타이완의 특정 지역의 작물 생산량을 제시하는 공식을 적용하여 예측한다. 그러나 이 연구에서의 기후데이터는 기후속성이 두 가지 뿐이며, 제안하는 공식은 다른 지역이나 국가에서 적용되기 어렵다.

중국에서는 기후요인이 겨울철 밀과 여름철 옥수수 생산량에 미치는 영향에 관한 연구가 수행되었다[6]. 이 연구에서는 기후요인과의 상관분석을 통해 해당 지역의 밀과 옥수수의 수확에 중요한 기간을

산정했다.[6] 그러나 이 연구 또한 세 가지의 속성만을 사용하였기에 속성의 개수가 적었다.



그림 2. 실험 과정

Fig. 2. The experimental process

III. 본 론

1. 데이터 집단

1999년부터 2009년까지의 경북 지방의 기온, 일교차, 습도, 적설량, 습도, 풍속, 운량, 일사량, 일조량에 대한 데이터와 사과 생산량에 관한 데이터를 각각 기상청과 통계청에서 습득했다.

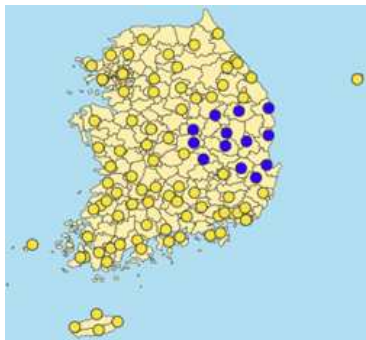


그림 1. 데이터 수집 지역

Fig. 1. Map of selected fruits plantation

표 1. 데이터 코딩

Table 1. Data coding

Attributes	Abbr	Code1 X > Average	Code 2 Average > X
일교차	TG	A	B
일강수량	DRF	A	B
풍속	WS	A	B
습도	H	A	B
기온	T	A	B
적설량	SC	A	B
일사량	SR	A	B
기압	AP	A	B
일조량	SS	A	B
운량	C	A	B
사과 생산량	AY	A	B

표2. 실험 결과 apple

Table 2. Experimental result

RULES	CONFIDENCE
RULE 1: T=B TG=A ==> AY=A	1
RULE 2: DRF=B WS=B ==>AY=A	1
RULE 3: T=B TG=A DRF=B ==> AY=A	1
RULE 4: T=B TG=A H=B ==> AY=A	1
RULE 5: TG=A DRF=B WS=B ==> AY=A	1
RULE 6: DRF=A SC=A ==> AY=B	1
RULE 7: TG=B AP=A C=A ==> AY=B	1
RULE 8: TG=B AP=A SR=B ==> AY=B	1
RULE 9: TG=B SC=A C=A ==> AY=B	1
RULE 10: TG=B AP=A C=A SR=B ==> AY=B	1

2. 연관규칙

연관규칙이란 동시에 발생하는 사건들을 규칙의 형태로 표현한 것으로 특정 사건이 발생하면 동시에 혹은 일정한 시간 간격 사이에 다른 사건이 일어나는 관련성을 의미한다[7]. $I = \{i_1, i_2, \dots, i_n\}$ 를 항목이라 부르는 리터럴들의 집합이라 하고, D 를 트랜잭션들의 집합이라 했을 때 각 트랜잭션 T 는 $T \subseteq I$ 인 항목들의 집합이다. 트랜잭션들은 TID 라 부르는 식별자를 가지고 있으며 X, Y 를 항목들의 집합이라 하면 $X \subseteq T$ 이고, $Y \subseteq T$ 이다. 이 때 트랜잭션 T 는 X, Y 를 포함한다고 말한다. 전체 D 에 해당하는 조건 항목이 결론 D 에 해당하는 항목들을 야기한다고 정의하며 연관규칙은 $R: X \Rightarrow Y$ 로 나타내고 이 때 $X, Y \subseteq I$ 이고 $X \cap Y = \emptyset$ 이다. 이 때 규칙의 타당성을 검증하기 위한 척도로서 지지도와 신뢰도가 적용된다. 지지도는 전체 항목 중에 연관규칙 $R: X \Rightarrow Y$ 를 지지하는 비율을 의미하는 척도를 말하며, 신뢰도는 X 의 모든 항목을 포함하고 있는 트랜잭션의 개수에 대하여 Y 또한 포함하는 트랜잭션의 비율을 의미한다.

3. 실험 및 결과

실험의 과정은 다음과 같다. 농업/기상데이터를 평균기반 이산화 [8] 기법을 적용하여 데이터 값을 A와 B로 가공한 후 연관규칙 기법을 적용하였다.

이 논문에서 우리는 기후 인자들과 생산량 간의 상관관계를 확인할 수 있다. RULE 1은 기온이 낮고 일교차가 크면 사과 생산량이 많음을 보인다. 일강수량이 적고 풍속이 느리면 사과 생산량이 많음이 RULE2에서 보여 진다. RULE3에서 기온이 낮고 일교차가 크며 일강수량이 낮으면 사과 생산량이 높음을 볼 수 있다. RULE4에서 기온이 낮고 일교차가 크며 습도가 낮으면 작물 생산량이 높음을 볼 수 있다. 일교차가 크고 일강수량이 적으며 풍속이 느리면 생산량이 많음이 RULE5에 나타난다. RULE6는 일강수량이 많고 풍속이 빠르면 생산량이 적음을 보인다. 일교차가 작고 기압이 높으며 운량이 많으면 생산량이 적음이 RULE7에서 보여 진다. 일교차가 작고 기압이 높으며 일사량이 작으면 생산량이 적음이 RULE8에서 보여 진다. RULE9에서 일교차가 작고 적설량이 많으며 운량이 많으면 생산량이 적음을 볼 수 있다. 일교차가 작고 기압이 높으며 운량이 많고 일사량이 적으면 생산량이 낮음이 RULE10에서 보여 진다. 실험을 통해 사과의

생산량은 습도와 운량이 많으며 일사량이 적은 환경에서는 낮음을 알 수 있고, 일교차가 있으며 습도와 강수량이 낮은 환경에서는 높음을 알 수 있다.

IV. 결 론

이 논문에서 우리는 농업/기상 데이터를 이산화 한 후 데이터 마이닝의 연관규칙 기법을 적용하여 농업경영에 영향을 미치는 기후 인자들과 생산량 간의 연관성을 분석했다. 그 결과, 이 연구에서는 기존의 알려진 단일 속성 간에 연관성이 아닌 복합적 속성 연관성에 대한 흥미로운 지식을 산출하였다.

이 연구의 분석 결과를 기반으로 향후 품종을 확대하여 농업 경영 생산성 향상을 위한 연구가 더 활발히 수행 될 수 있으리라 기대한다.

Acknowledgment

이 논문은 2013년도 미래창조과학부의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No.2013R1A2A2A01068923). / 본 연구는 농림 축산 식품 부 수출전략기술개발사업에 의해 이루어진 것임(No.114083-3).

Reference

[1] United States Department of Commerce National Oceanic and Atmosphere Administration (NOAA), 2008. Economic statistics for NOAA(2008).
[2] Jeong, W. J., Myoung, D. J., & Lee, J. H. (2009). Comparison of climatic conditions of sweet pepper's greenhouse between Korea and the Netherlands. Journal of Bio-Environment Control

[3] Jones, P. G., and Thornton, P. K., "The potential impacts of climate change on maize production in Africa and Latin America in 2055", Global environmental change, 13(1), 2003, pp. 51-59.
[4] J.W. Lee, Y.J. Jang, 2013.Effects of Meteorological Elements in the Production of Food Crops: Focused on Regression Analysis using Panel data. Journal of Environmental Science International 22(9), 1171~1180.
[5] CHANG, Ching-Cheng. The potential impact of climate change on Taiwan's agriculture. Agricultural Economics, 2002, 27.1: 51-64.
[6] SHAO-E, Yang; BING-FANG, Wu. Research on the Relationship between Meteorological Factors and Yields of Winter Wheat and Summer Maize in North China Plain. In: Bioinformatics and Biomedical Engineering(iCBBE),20104th International Conference on. IEEE, 2010. p. 1-4.
[7] J.H. Kim, H.M.Kim, Introduction to Concept in Association Rule Mining. Korea information science society,29(1), 2002, pp 100-102.
[8] KOTSIANTIS, Sotiris; KANELLOPOULOS, Dimitris. Discretization techniques: A recent survey. GESTS International Transactions on Computer Science and Engineering, 2006, 32.1: 47-58.