SLJoAS Sri Lankan Journal of Applied Sciences
Research Article
www.sljoas.uwu.ac.lk/index.php/sljoas

Association Rule Mining for a Climatic Condition based Recommender System: A Cinnamon Cultivation Case Study from Galle, Sri Lanka

K.M.G.T.R. Waidyarathna[a], S.M. Vidanagamachchi[b*]
[a]Advanced Technological Institute, Galle.
[b] Department of Computer Science, University of Ruhuna, Matara.

*** Corresponding author email address**: smv@dcs.ruh.ac.lk

(Received 14th May 2023; Accepted 14th August 2023)

Abstract:

Weather prediction has become complex due to ever-changing weather conditions. Effective analysis of the weather requires the identification of influential factors in changing weather conditions. Several machine learning techniques have been applied for whether predictions during the past few years. Linear regression, functional regression, neural networks, classification and regression trees, k-nearest neighbours, and Naïve Bayes are frequent among them. In this work, machine learning techniques are thoroughly analyzed, with particular emphasis on the utilization of recommender systems for weather prediction. Further, analyze meteorological data of Galle district in Sri Lanka using suitable techniques to find hidden patterns by transforming the historical information into usable knowledge. The main objective of this paper is to propose a recommender system for Cinnamon cultivation in the Galle district to predict future weather conditions based on association rule mining. This can further be used to identify the best sowing dates in the future for not only cinnamon, but also tea and other cultivations in the district.

Keywords: Machine-learning, Weather, Influential-factors, Recommender-system

1. Introduction

The weather forecasting is mainly based on historical weather conditions and has become complex due to ever-changing weather conditions. It might be associated with some uncertainty with the actual weather conditions. Changing weather conditions affect the growth of cinnamon as well as the harvest. Due to new trends in available technologies, storage of massively growing past meteorological data is available. Therefore, several attempts have been carried out utilizing above mentioned techniques to extract meaningful knowledge from these past meteorological data. Time series analysis plays a vital role in data mining for making effective predictions. Analyzing weather data requires effective identification of weather attributes and their correlation, for which we utilize time series analysis in this work.

In the field of weather forecasting, the integration of machine learning and data mining techniques has gained significant attention. Past efforts have encompassed a variety of methods, including neural networks [1-3] and regression analysis [4]. Similarly, data mining techniques such as decision trees [5] and the apriori algorithm [6] have been employed to analyze meteorological data.

A notable application of machine learning techniques is in rainfall forecasting. Deepti et al [7] conducted a comparative analysis using three different approaches: classification and regression trees, k-nearest neighbor, and Naïve Bayes. Their study revealed that the C4.5 decision tree algorithm exhibited superior performance in instance classification compared to Naïve Bayes. The accuracy of the C4.5 algorithm was recorded at 88.2%, outperforming Naïve Bayes with an accuracy of 54.8% [8].

The digital landscape, characterized by an abundance of online information, has posed the challenge of extracting relevant content for users. Notable search engines like Google, DevilFinder, and Altavista have made strides in information retrieval, though with limited focus on prioritization and personalization. As a result, the need for advanced recommender systems has surged. Recommender systems have found diverse applications in modern technology. These systems leverage user profiles to predict preferences, assist in item selection within e-commerce platforms, boost revenues through personalized product suggestions, and facilitate efficient catalog browsing within scientific libraries [9].

A recommender system plays a vital role in supporting users to select information or items based on their preferences. Serendipity is a good concept that can be used as a support for making good discoveries that are not explicitly searched by users. Recommender systems can be classified into three main categories: content-based, collaborative filtering and hybrid techniques, based on the filtering techniques utilize [10-11] Collaborative filtering is considered the most common and successful method for recommending items to the target user. These types of recommender systems can be useful in the agricultural field as well [12]. Association rule mining for recommender systems [13] and association rule-based collaborative filtering plays a vital role here in agriculture and other emerging fields [14] though that is not commonly used for weather prediction as other machine learning techniques such as regression and classification methods. Some hybrid recommendation systems were introduced based on association rules as well [15].

A recent research proposed a weather forecasting model based on data mining and an improved support vector machine algorithm [17]. In this article, the authors used association rule mining to mine frequent patterns in weather data and then used an improved SVM algorithm to predict temperature and precipitation. The observed results showed that the proposed model outperformed other traditional machine learning algorithms such as decision trees and back-propagation neural networks. In summary, while using association rule mining in weather prediction is not as common as other machine learning techniques, most studies show promising results and a good starting point for further research in this area.

An association rule mining approach was introduced recently for now-casting (short-term weather forecasting), which helps to distinguish between severe and normal weather conditions [18]. Another study was carried out to predict weather conditions in Malaysia using time series data from Association Rule Mining [19].

Context-aware recommender systems are mostly concerned with contextual information (eg: time, a place for watching a movie) when recommending the most relevant items to users. It is not sufficient to consider only users and items, but there is a high demand for recommending a system attaching contextual information to the process [16].

A mobile context-aware recommender system called STS (South Tyrol Suggests) was suggested including forecasting of current and previous weather conditions that uses contextual factors to recommend POI (Places of Interest). Here they have relied on a matrix factorization algorithm integrating up-to-date weather forecast data and generated personalized context-aware recommendations. They developed a variant of STS called STS-S including the same model and contextual factors in STS except for the weather conditions [18].

Time series analysis can be applied when the measurements are made at regular time intervals. This helps in identifying hidden patterns of the data and provides a better understanding by fitting a forecasting model. Univariate and multivariate models are the two main models in time series analysis. Linear regression and least square-based models can also be utilized in the time series analysis. The ARIMA (Autoregressive Integrated Moving Average) model and exponential smoothing provide complementary forecasting types for time series. ARIMA (p, q, r) models provide facilities for identifying different seasonal changes. For example, a model with one AR term, a first difference, and one MA term would show as ARIMA (1, 1, 1).

ARIMA and regression analysis has been used by Malgorzata et.al. [20] for forecasting weather in four European areas: Jokioinen, Dikopshof, Lleida and Lublin, which are having different climatic zones. They showed the dynamics of time-series data and forecasted sensible results. Further, ARIMA has been used to analyse temperature variability and trends in the area of Karachchi, Pakistan and found that there is an increasing trend in the temperature [21]. Our main objective was to develop a useful recommendation model with association rule mining technique based on past weather patterns, which can be used for as a support for the cinnamon cultivation. However, in the proposed method, association rules have been used and aim to extend the work with the ARIMA model.

2. Materials and methods

In this analysis, a recommender model was established based on association rules and tested it on a dataset from an online dataset mentioned in the data availability section. Some possible suggestions for improving the recommendation effectiveness are suggested and the discussion section. The procedure for recommendation model analysis includes the steps: 1. Missing value imputation/ removal in pre-processing and 2. Utilization of association rule mining to develop the model.

According to the above procedure, the developed rules can be then used in developing the recommender system. Our methodology utilized Python programming language in Google Colaboratory environment. If someone needs to predict the system with more recent weather data, high computational power is required.

3. Results and Discussion

Based on the association rules that are described in the later paragraph, the framework of the recommendation model in Fig. 01 has been developed.

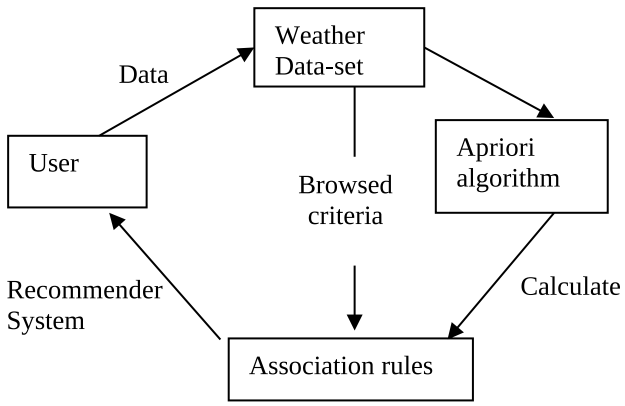

Fig. 1. Framework of the association rules based recommender model

In order to start with association rules, the Apriori algorithm is adopted. Association rules are the main inputs to the recommender system. Apriori algorithm is known as the most classic and actively utilized algorithm under association rules. The "support" and "confidence" can be used as the measurements of the usefulness of association rules. The support of a rule is the probability of events occurs in weather conditions, and the confidence of a rule is the conditional probability.

Consider X, Y as difference weather conditions, the support of the rule "$X \Rightarrow Y$"is decided as

$$Support(X \Rightarrow Y) = P(X \cup Y). \quad (1)$$

The confidence of the rule "$X \Rightarrow Y$"is decided as:

$$(X \Rightarrow Y) = (Y|X) = P(X \cup Y) P(X) \quad (2)$$

If the support or confidence is insignificant, it is required to set minimum support and confidence. An example of the process of Apriori algorithm can be seen in the following tables;

Table 1: Example of Apriori Algorithms with different weather conditions:Database

DayID	Weather Condition
01	A,B.C
02	B,C,D,E
03	A,C,E,F
04	B,C,E

Table 2: Example of Apriori Algorithms with different weather conditions: C1

Weather Set	Support
[A]	2
[B]	3
[C]	4
[D]	1
[E]	3
[F]	1

Table 3: Example of Apriori Algorithms with different weather conditions: L1

Weather Set	Support
[A]	2
[B]	3
[C]	4
[E]	3

Table 4: Example of Apriori Algorithms with different weather conditions: C2

Weather Set	Support
{A,B}	1
{A,C}	2
{A,E}	1
{B,C}	2
{B,E}	2
{C,E}	3

Table 5: Example of Apriori Algorithms with different weather conditions: L2

Weather Set	Support
{A,C}	2
{B,C}	2
{B,E}	2
{C,E}	3

Following two steps get repeated.

1. If the support of the item set is lower than the threshold remove the item sets of CK to generate frequent set LK (called prune);

2. Generate the candidate set CK+1 (called Join) by combining the item sets of LK

The generated association rules from the weather sets are summarized in Table 8. These association rules can then be used in the recommender system. For eg: if a weather condition is A, C condition is recommended to the farmer with the probability of 75% for cultivating. Experiments were performed on the selected 2543 data. The support threshold of 0.001 and the confidence threshold of 0.1 were used. The association rules were found based on weather condition records. Using the Apriori algorithm, 5 frequent weather sets of size 2 and 1 frequent weather sets of size 3 were obtained, with total 10 association rules. The confidence distribution of association rules were ranged from 0.35 to 1, and 80% was higher than 0.6.

Table 6: Association Rules of the Weather Conditions

Association rules	Support	Confidence
A→C	0.5	1
C→E	0.75	0.75
E→C	0.75	1
B,E→ C	0.5	1
B,C → E	0.5	1

4. Conclusion

We have developed a framework of a recommender system based on past weather conditions, which can be used in cultivation purposes (i.e. predicting the sewing dates) in Galle district, Sri Lanka. This can further be extended with more computational power and dataset and improved for future weather changes too. Further, utilizeing this technique can be recommended in any kind of cultivation in the area such as cinnamon, tea and coconut.

Author Contributions

Both authors have contributed equally.

Conflicts of Interest

No conflict of interest to declare.

References

[1] Wojciech Z, Jacek M and Jaroslow S 2015 Theory and Engineering of Complex Systems and Dependability, Proc. of the 10thInternational Conference on Dependability and Complex Systems p 78-85

[2] Saktaya S and Warangkhana K 2016 Neural Network training model for weather forecasting using Fireworks Algorithm, International Computer Science and Engineering Conference (ICSEC) p 1-7

[3] Kumar A, Singh MP Saswata G and Abhishek A 2012 Weather Forecasting Model using Artificial Neural Network, Procedia Technology p 311-318

[4] Ting D and Zongjian K A 2012 Comparison of Statistical Approaches for Seasonal Precipitation Prediction in Pakistan, Journal of American Meteorological Society

[5] Geetha A and Nasira G M 2014, Data mining for meteorological applications: Decision trees for modeling rainfall prediction, IEEE International Conference on Computational Intelligence and Computing Research (ICCIC) p 1-4

[6] Athiyaman B, Sahu R and Srivastava A 1978 Hybrid Data Mining Algorithm: An Application to Weather Data, J. of Indian Research p 71-83

[7] Deepti G and Udayan G 2015 'A comparative study of classification algorithms for forecasting rainfall, 4th International Conference on Reliability Infocom Technologies and Optimization

[8] Fahad Sheikh S, Karthick D, Malathi J, S Sudarsan and C Arun 2016 Analysis of Data Mining Techniques for Weather Prediction, Indian Journal of Science and Technology 9

[9] Isinkaye F O, Folajimi Y O and Ojokoh B A 2015 Recommendation systems: Principles, methods and evaluation, Egyptian Informatics Journal p 261-273

[10] Vanda V 2014 Masters thesis (University of Zagreb)

[11] George A S, Aimilia T and George A P 2015 Recommender Systems Review: Types, Techniques and Applications, University of Cyprus, Accessed January 2021 from University of Cyprus Repository.

[12] Anji K R and Kiran K R 2018 Recommendation system: A Collaborative Model for Agriculture, International Journal of Computer Sciences and Engineering p 120-123.

[13] Rick S 2013 Masters Thesis (Tilburg University, Netherlands)

[14] Behzad S N, Nasim S, Reza M and Mohammad H N 2019 Improving Performance of Association Rule-Based Collaborative Filtering Recommendation Systems using Genetic Algorithm, International Journal of Information Technology and Computer Science p 48-55.

[15] Ahmed A 2013, Masters Thesis (Western Kentucky University)

[16] Adomavicius G, Mobasher B, Ricci F and Tuzhilin A 2011 Context-Aware Recommender Systems AI Magazine p 67-80.

[17] Kumar R and Khatri R 2016 A Weather Forecasting Model using the Data Mining Technique, International Journal of Computer Applications p 4-12

[18] Gabriela C, Andrei M and István G C 2020 RadRAR: A relational association rule mining approach for nowcasting based on predicting radar products' values, Procedia Computer Science p 300-309

[19] Ab. Rashid Rabiatul, Nohuddin Puteri and Zainol Zuraini. 2017 Association Rule Mining Using Time Series Data for Malaysia Climate Variability Prediction, International Visual Informatics Conference p 120-130.

[20] Murat M, Malinowska I, Gos M, Krzyszczak J 2018 Forecasting daily meteorological time series using ARIMA and regression models, Int. Agrophys. P 253-264.

[21] Amjad M, Khan A, Fatima K, Ajaz O, Ali S, Main K 2023 Analysis of Temperature Variability, Trends and Prediction in the Karachi Region of Pakistan Using ARIMA Models, Atmosphere.