Integrating Heterogeneous Agriculture Information Using Naïve Bayes and FCA

J. S. Kanchana and S. Sujatha

Department of Information Technology, K.L.N College Of Engineering, Pottapalayam
Asst. Prof, MCA Dept, Anna University Regional Office, BIT Campus, Tiruchirapalli

Received 27 May 2016; Accepted 28 June 2016; Available 12 July 2016

Address For Correspondence:
J. S. Kanchana, Department of Information Technology, K.L.N College Of Engineering, Pottapalayam

Copyright © 2016 by authors and American-Eurasian Network for Scientific Information (AENSI Publication). This work is licensed under the Creative Commons Attribution International License (CC BY).
http://creativecommons.org/licenses/by/4.0/

ABSTRACT
Data mining is an interdisciplinary subfield of computer science. It is the computational process of discovering patterns in large data sets. Naïve Bayes is used in this paper which makes use of probability to classify real and discrete data. The Formal Concept Analysis is used in mapping the results of the Naïve Bayes and the given data. In this paper, Naïve Bayes algorithm is used to classify agricultural datasets and the results are mapped to farmer data using Formal Concept Analysis to provide recommendation for farmers.

KEYWORDS:

INTRODUCTION

Agriculture is one of the fields that have not yet made full use of the potential of technology. It is one field where technology and techniques should be applied and help farmers who provide us our basic needs, the food. There are many factors that affect the growth of a crop. All the factors have to be analysed before investing on a particular crop. The project aims to use data mining technique to help farmers. The location of the farmers is got from them and based on the location all the factors are analysed and data sets of past years production are analysed. Thus the farmers can predict the crops to be planted in future. This will help farmers to maximize their profit.

There are research paper help to identify mechanisms to get good quality and improved crop yields using a new algorithm named as “Agro algorithm” implemented in Hadoop platform and uses Hadoop framework to handle large amount of data sets. A paper presents an ontology-based approach and a formal concept analysis (FCA) approach to integrating heterogeneous tourism information for online tour planning. An analytic hierarchy process is used to rank the tourism attractions suggested by the ontology and FCA-based approaches.

Naïve Bayes and Genetic Algorithm are applied for classifying texture of agriculture soil data in a research. The database contains measurements of various Soil profiles. The achieved performances were compared and analyzed on the collected Soil data. However, Naïve Bayes algorithm has much better performance than Genetic algorithms. A paper also compares the different classifiers and the outcome of the research could improve the management and systems of soil uses throughout a large number of fields.

Our goal is to find the crop that can be planted by the farmers so that they can get maximum profit. The location of the farmer is got and the location is analysed for its soil, rainfall and climate. The crop that best suits the location of the farmer is recommended by this recommendation system.

To Cite This Article: J. S. Kanchana and S. Sujatha., Integrating Heterogeneous Agriculture Information Using Naïve Bayes and FCA. Advances in Natural and Applied Sciences. 10(10); Pages: 308-311
Architecture:

The recommendation process consists of two stages. The first stage is to apply Naïve Bayes algorithm and classifies the preprocessed data sets. Weka a free open source data mining tool is used to classify the data based on the threshold value.

Side by side input data is collected from the user such as their location, investment and area of the user.

The next stage is to map the results of the Naïve Bayes algorithm and the user input the crop that matches the location are marked and to find the best of results the profit is calculated and the recommendation is given to the user as which crop would yield them the maximum profit.

![Data Flow Diagram](image)

**Fig. 1:** Data Flow Diagram

**Naïve Bayes Algorithm:**

Naïve Bayes classifiers are linear classifiers that are known for being simple yet very efficient. The probabilistic model of naive Bayes classifiers is based on Bayes’ theorem, and the adjective naive comes from the assumption that the features in a dataset are mutually independent. In practice, the independence assumption is often violated, but naive Bayes classifiers still tend to perform very well under this unrealistic assumption. Especially for small sample sizes, naive Bayes classifiers can outperform the more powerful alternatives.

Being relatively robust, easy to implement, fast, and accurate, naive Bayes classifiers are used in many different fields. Some examples include the diagnosis of diseases and making decisions about treatment processes, the classification of RNA sequences in taxonomic studies, and spam filtering in e-mail clients. However, strong violations of the independence assumptions and non-linear classification problems can lead to very poor performances of naive Bayes classifiers. We have to keep in mind that the type of data and the type problem to be solved dictate which classification model we want to choose. In practice, it is always recommended to compare different classification models on the particular dataset and consider the prediction performances as well as computational efficiency.

**Formal Concept Analysis:**

Formal concept analysis (FCA) is a method of data analysis with growing popularity across various domains. FCA analyzes data which describe relationship between a particular set of objects and a particular set of attributes. Such data commonly appear in many areas of human activities. FCA produces two kinds of output from the input data. The first is a concept lattice. A concept lattice is a collection of formal concepts in the data which are hierarchically ordered by a subconcept-superconcept relation. Formal concepts are particular clusters which represent natural human-like concepts such as “organism living in water”, “car with all-wheel drive system”, “number divisible by 3 and 4”, etc. The second output of FCA is a collection of so-called attribute implications. An attribute implication describes a particular dependency which is valid in the data such as “every number divisible by 3 and 4 is divisible by 6”, “every respondent with age over 60 is retired”, etc. A distinguishing feature of FCA is an inherent integration of three components of conceptual processing of data and knowledge, namely, the discovery and reasoning with concepts in data, discovery and reasoning with dependencies in data, and visualization of data, concepts, and dependencies with folding/unfolding capabilities.
Integration of these components makes FCA a powerful tool which has been applied to various problems. Examples include hierarchical organization of web search results into concepts based on common topics, gene expression data analysis, information retrieval, analysis and understanding of software code, debugging, data mining, and design in software engineering. Internet applications including analysis and organization of documents and e-mail collections, annotated taxonomies, and further various data analysis projects described in the literature.

A table with logical attributes can be represented by a triplet \((X,Y,I)\) where \(I\) is a binary relation between \(X\) and \(Y\). Elements of \(X\) are called objects and correspond to table rows, elements of \(Y\) are called attributes and correspond to table columns, and for \(x \in X\) and \(y \in Y\), \((x,y) \in I\) indicates that object \(x\) has attribute \(y\) while \((x,y) \notin I\) indicates that \(x\) does not have \(y\). For instance, Fig. 1 depicts a table with logical attributes. The corresponding triplet \((X,Y,I)\) is given by \(X = \{x_1,x_2,x_3,x_4\}, Y = \{y_1,y_2,y_3\} \) \(x_1 \times x_2 \times x_3 \times x_4\) \(y_1 \times y_2 \times y_3\) \(I = \{(x_1,y_1), (x_2,y_2), (x_2,y_3), (x_3,y_1), (x_3,y_3), \ldots \}\). Since representing tables with logical attributes by triplets is common in FCA, we say just “table \((X,Y,I)\)” instead of “triplet \((X,Y,I)\)” representing a given table”. FCA aims at obtaining two outputs out of a given table. The first one, called a concept lattice, is a partially ordered collection of particular clusters of objects and attributes. The second one consists of formulas, called attribute implications, describing particular attribute dependencies which are true in the table.

| \(y_1\) | \(y_2\) | \(y_3\) | \(
\ldots\)
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(x_1)</td>
<td>(\times)</td>
<td>(\times)</td>
<td>(\times)</td>
</tr>
<tr>
<td>(x_2)</td>
<td>(\times)</td>
<td>(\times)</td>
<td>(\ldots)</td>
</tr>
<tr>
<td>(x_3)</td>
<td>(\times)</td>
<td>(\times)</td>
<td>(\ldots)</td>
</tr>
</tbody>
</table>

\[\text{Fig. 1: Applying fca on crisp set}\]

The clusters, called formal concepts, are pairs \((A,B)\) where \(A \subseteq X\) is a set of objects and \(B \subseteq Y\) is a set of attributes such that \(A\) is a set of all objects which have all attributes from \(B\), and \(B\) is the set of all attributes which are common to all objects from \(A\). For instance, \((\{x_1,x_2\},\{y_1,y_2\})\) and \((\{x_1,x_2,x_3\},\{y_2\})\) are examples of formal concepts of the (visible part of) table in Fig. 1. An attribute implication is an expression \(A \Rightarrow B\) with \(A\) and \(B\) being sets of attributes. \(A \Rightarrow B\) is true in table \((X,Y,I)\) if each object having all attributes from \(A\) has all attributes from \(B\) as well. For instance, \(\{y_3\} \Rightarrow \{y_2\}\) is true in the (visible part of) table in Fig. 1, while \(\{y_1,y_2\} \Rightarrow \{y_3\}\) is not \((x_2\) serves as a counterexample).

**Experimental Results:**

Datasets are collected from various sources available on the internet. The dataset are analyzed to understand what data is provided. All unwanted, missing and noisy data from the dataset are removed and is used for further processing. A threshold value is selected for classifying the preprocessed data. The preprocessed data is given as input to the Naïve Bayes algorithm. The algorithm classifies the data based on the probability values. The values above the threshold are considered for further processing. Input is got from the user using a user interface. The users fill input fields such as investment capability, soil type, rainfall, etc. The input which is received is given as input for the formal concept analysis. The formal concept analysis is a matrix mapping technique for mapping data. The two entities that are passed as input to FCA are the output of Naïve Bayes algorithm and the data received from the user. Both the entities are mapped to derive efficient results. The best results of the formal concept analysis are chosen. The crops that produce maximum profit are found. These results are further analyzed using F-measure value for recommendation to users.

**Related Work:**

The existing system classifies only the soil and Genetic algorithm is used in existing system. With the Genetic Algorithms, the binary decision tree explains the prediction of the category of Soil data that is built first. To generate the decision tree for the dataset it takes more time and then the rules are classified with that decision tree. Genetic Algorithms are not suitable for managing imprecise or uncertainty in soil data.

**Proposed Work:**

Naïve Bayes and Formal Concept analysis is used in the proposed system. Datasets of previously collected data over years is used to predict the future. Naïve Bayes is used in this project which makes use of probability to arrive at the results. It handles real and discrete data well. The Formal Concept Analysis is used in mapping the results of the Naïve Bayes and the farmer requirements.
Experimental Results:

Table:

<table>
<thead>
<tr>
<th></th>
<th>GENETIC ALGORITHM</th>
<th>NAIVE BAYES AND FCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recall</td>
<td>0.8701</td>
<td>0.8403</td>
</tr>
<tr>
<td>Precision</td>
<td>0.6912</td>
<td>0.9318</td>
</tr>
<tr>
<td>F-Measure</td>
<td>0.7696</td>
<td>0.8838</td>
</tr>
</tbody>
</table>

Graph:

Conclusion:

Data mining is the analysis step of the “knowledge discovery in databases” process, it means not extracting data but knowledge and discovering patterns from data. Agriculture is one field which needs to make use of the knowledge. A pattern is analyzed from the large data sets and the current production is predicted based on probability using Naïve Bayes algorithm. Formal Concept Analysis is then used to map the output of the Naïve Bayes algorithm with the user input. Thus both the future prediction and the user requirements are implemented. Further the profit from that prediction is also approximately calculated and the recommendation is provided. This recommendation creates an opportunity for farmers to make effective use of technology for future prediction. Thus two algorithms are employed to provide recommendation in the field of agriculture.

REFERENCES