HCTL Open Int. J. of Technology Innovations and Research
HCTL Open IJTIR, Volume 10, July 2014
e-ISSN: 2321-1814
ISBN (Print): 978-1-62951-619-6

Horticulture Plant Disease Prediction by using Graph Database

Shinkita Negi*and Suchi Juyal†
shinkita.negi@gmail.com

Abstract

Horticulture plant pathology is an important part among the various disciplines of horticulture. Plant pathology is diagnosis, management, prediction for plant disease which can help enhance yield and quality of horticultural crops. Plant disease Prediction System provides interface with user for detecting the disease in the plants.There are mainly three types of plant diseases such as Fungal, Bacteria and Virus Plant Disease. We focused here on the fungal diseases classification and prediction Model. We propose an approach for plant disease prediction model for the horticulture plants. To implement this we used the Graph database for the purpose of Integration of the Horticultural Plants database and then classify the data into classes of Fungal Disease. All the data has been represented and traversal by using the graph database, Neo4j.

*M.Tech Student, Department of Computer Science and Engineering, Graphic Era University, Dehradun, India.
†Asst. Professor, Department of Computer Science and Engineering, Graphic Era University, Dehradun, India. Email: bhadula.shuchi@gmail.com

HCTL Open Int. J. of Technology Innovations and Research
HCTL Open IJTIR, Volume 10, July 2014
e-ISSN: 2321-1814
ISBN (Print): 978-1-62951-619-6

Data has been classified into sub graph for the plant disease prediction model; where the user enters the input values for a plant and the system checks its compatibility mode for grow. We followed the approach of the FSM (Frequent Sub graph Mining). The objective of FSM is to extract all the frequent sub graphs, in a given data set, whose occurrence counts are above a specified threshold. To follow the FSM approach we calculated the threshold value of three main factors which are the casual agents of the disease such as soil PH value, Temperature and Rainfall Humidity. Then compare them with the user input for predicting the compatible mode for the growth of the plant. This application could be helpful to the decision makers to control over the infected plants.

Keywords

Plant Disease Prediction, Graph Database, Horticulture.

Introduction

Horticulture is the wing of agriculture that deals with the art, science, technology, and business involved in plant cultivation. The real potential of horticulture sector was highlighted only after the National Horticulture mission which has been started in 2005-06. To increase the growth rate of the horticulture field we must move to its research field, where we could find out the new opportunities and new trends and policies to increase horticulture growth rate.

Horticulture Data network has emerged as a novel field of research often applied to Organize data of Horticulture plants for analysing and classifying the data. In this working System an user interface has been provided for the analysing the result for checking the risk preferences of the target plant by providing its circumstances conditions.

Literature Review

Graph Databases model the whole dataset as one big dense network structure. Graph databases (GDB) has considered as the alternate of the Relational Database Systems (RDBMS). The examples of this applications are implemented on the field of Chemistry, social networking ,biology and semantics web. As the view of comparison between Graph database and Relational Database

e-ISSN: 2321-1814
ISBN (Print): 978-1-62951-619-6

could be find in these aspects such as data model features, Query facility and data size, and time execution. Graph database technology is an effective tool for modelling data when a focus on the relationship between entities is a driving force in the design of a data model. Modelling objects and the relationships between them means almost anything can be represented in a corresponding graph. Graph databases are specifically designed to store information about the relationships between nodes, and enable applications to use this information to traverse a graph extremely efficiently.

Graph Mining is used to extract patterns from graph structured data. The graph structure is a nice way of representing and explaining complex data forms and phenomena. Graph mining [2] finds its applications in various problem domains, including: bio-informatics, chemical reactions, Program flow structures, computer networks [2], social networks etc.

In literature various graph mining approaches have been proposed. Each of these approaches is based on either classification; clustering or decision trees data mining techniques. Here we used **Sub-Graph Mining**. Sub graph is a graph whose vertices and edges are subsets of another graph. The frequent sub graph mining problem is to produce the set of sub graphs occurring in at least some given threshold of the given input example graphs.

Frequent Subgraph Mining (FSM) [4] is the essence of graph mining. The objective of FSM is to extract all the frequent subgraphs, in a given data set, whose occurrence counts are above a specified threshold [6]. FSM can be classified into mainly two approaches:

1. Apriori-based Approach [17] and
2. Pattern Growth Based approach.

Here in this algorithm, it share similar characteristics frequent item set. This algorithm uses Adjustment matrix for graph representation. BFS [17] technique is uses in this algorithm. Pattern growth based approach is especially used for pattern mining. Pattern growth [17] includes algorithm gSpan [17], MoFa, SPIN [17]. The pattern-Growth mining algorithm extends a frequent graph by adding new edge, in every possible position.

e-ISSN: 2321-1814
ISBN (Print): 978-1-62951-619-6

Proposed Work

In this project we propose to use Graph Database Algorithms for the purpose of Integration of the Horticultural Plants database and then classify the data on the basis of plant disease, then purpose the system for the plant disease prediction model for the forecasting risk of the targeted plant.

Here we used the graph database, through which the data has been integrated and traversal for the searching over the graph through node to node. Graph database is the collection of the nodes which are connected through the relationship link between each other.

After creating the graph structure for the Horticulture plant database, we classified the data on the basis of the Disease for plant disease prediction. Here the approach has been used for the sub graph mining for the getting the result. We followed the approach of the FSM (Frequent Sub graph Mining).

The objective of FSM is to extract all the frequent sub graphs, in a given data set, whose occurrence counts are above a specified threshold. We focused on the three main factor of the plant for analysis of plant disease prediction. These parameters are Soil PH value, RH, Temperature of the plant. In this working System an user interface has been provided for the analysing the result for checking the risk preferences of the target plant by providing its circumstances conditions.

Unstructured data is represented in graphical structure and database is presented in Neo4j. Each node contains information about a specific entity, and each connection specifies a relationship between entities. Each relationship in Neo4j has a type. Traversing is done in the graph database. The overall work is explained through a flow diagram as shown in figure 1. This flow diagram shows the actual progress of the system and gives the complete information about the system at each & every step.
In the project flow diagram the work has been broadly classify in four stages.

Stage 1: Integration of Data & Display of the Graph

After collecting the plant data structured into a main graph form and display it by using Neoclipses.

HCTL Open Int. J. of Technology Innovations and Research
HCTL Open IJTIR, Volume 10, July 2014
e-ISSN: 2321-1814
ISBN (Print): 978-1-62951-619-6

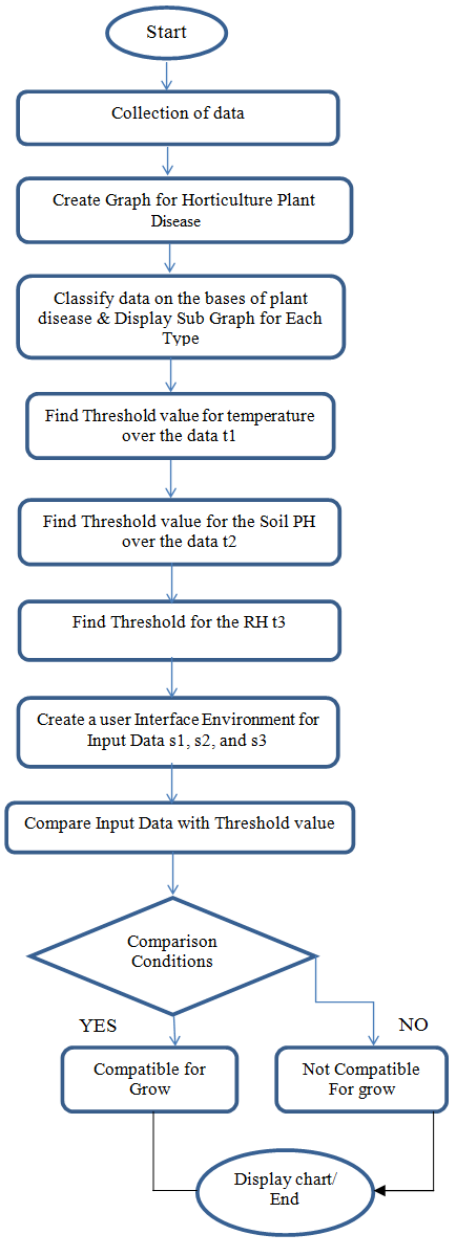

Figure 1: *Flow diagram of the system*

Stage 2: Classification of Data on the basis of Plant Disease

Classify the data on the basis of the disease type and create a sub graph by using the implemented algorithm.

HCTL Open Int. J. of Technology Innovations and Research
HCTL Open IJTIR, Volume 10, July 2014
e-ISSN: 2321-1814
ISBN (Print): 978-1-62951-619-6

Stage 3: Finding the Threshold Value for Three Parameters

In this step, we set the threshold value on the basis of three parameters such as Soil PH , Temperature and Rainfall Humidity. To get the threshold value we applied here the mean for each field. Mean and expected value are used synonymously to refer to one measure of the central tendency either of a probability distribution or of the random variable characterized by that distribution.

$$x' = \frac{x_1 + x_2 + ... + x_n}{n} \tag{1}$$

Stage 4: Comparing Threshold Value with the Input & Result

After set the threshold value, we need to compare this with given input value by the user to find out the compatible mode for the grow plant. And display the chart.

Algorithm Implemented

```
Algorithm Implemented Step-wise
Create Graph Database
Create Node
Set Properties of nodes
Create Relationship
Select  property                                /* Property to be searched */

Algo Traversing
for Node position traverse
    if  p <-  node.property                     // if property matches
        Split Graph
        sort lexically
        traverse <- depth
        r <- node-relationship                  // store relationship of first match
        if node.left.relationship   equals r
                    display properties
                    continue traverse down
        else
                if node.right.relationship  equals r
                        display properties
                        continue traverse down
                else
                        traverse <- down next node
                end if
        end if
        if p <- node.property
                    repeat Step 8 through18
        end if
    end if
end for
```

HCTL Open Int. J. of Technology Innovations and Research
HCTL Open IJTIR, Volume 10, July 2014
e-ISSN: 2321-1814
ISBN (Print): 978-1-62951-619-6

Experimental Results

The Research work has been carried out in the following manner. The work starts with the Creating graph and sub-graph on Neo4j [2]. Several programs have also been developed for the proposed work. Main Graph is created with four nodes. Main nodes is `fungal disease prediction` having Id 1. Next three nodes are the child nodes; `ascomycotina, plasmodiophoromycetes, uredinales` having id 2, 3 and 4. The graph is represented by the help of Neoclipes [1]. Here all the nodes are connected through a relationship , such as node 1 is connected to node 2 with the relationship `plasmodiophoromycetes` as well as node 1 is connected to node 3 with relationship `uredinales` and from node 4 with the relationship `ascomycotina`. Each child node is further classify into sub-nodes having its properties. The main graph given as below in figure 2.

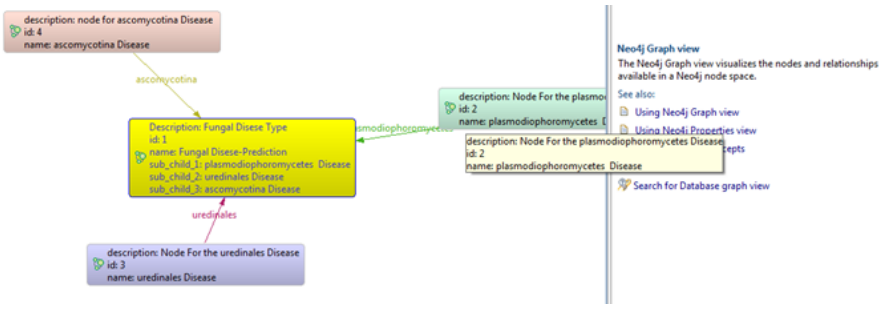

Figure 2: *Main Graph*

The Child nodes having the various properties such as Disease Name, Symptoms, Temperature_Max, Temperature_Min, Soil_PH_Max, Soil_Ph_Min, RH_Min, Rh_Max Plant Host. A node has many sub-nodes which all are connected through the relationship. Figure 3 display nodes with its properties.

To classify the data for the child node each class of fungal has been created as sub-graph by using the implemented algorithm. This classify the data by traversing over the nodes in the graph. Figure 4 shows the sub-graph for the `Plasmodiophoromycetes` node (class of fungal disease).

Finally the result will display in the form of chart which shows the comparison graph between user inputs and threshold values. The figure 5 shows the output.

HCTL Open Int. J. of Technology Innovations and Research
HCTL Open IJTIR, Volume 10, July 2014
e-ISSN: 2321-1814
ISBN (Print): 978-1-62951-619-6

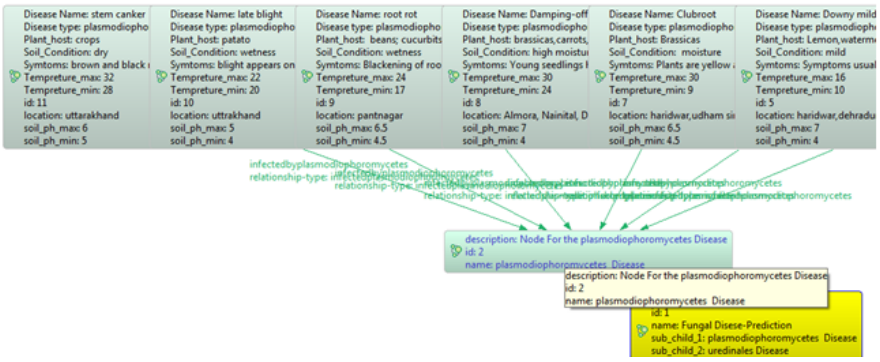

Figure 3: *Child Node*

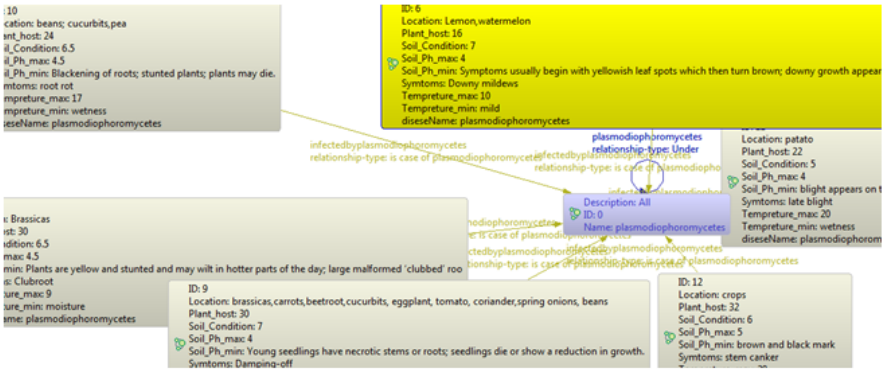

Figure 4: *Sub-graph*

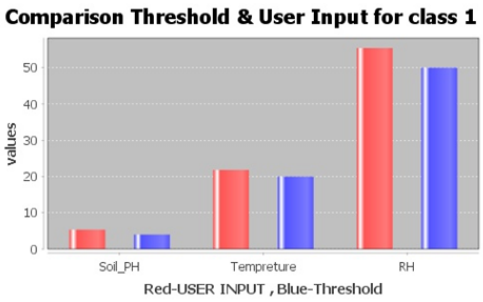

Figure 5: *Comparison between Threshold Value and User Input*

e-ISSN: 2321-1814
ISBN (Print): 978-1-62951-619-6

Conclusion and Future Scope

In this proposed System Horticulture Plant Disease data has been classified into three classes of Fungal Disease. All the data has been represented and traversal by using the graph database, Neo4j. Data has been classified into sub graph for the plant disease prediction model; where the user enters the input values for a plant and the system checks its compatibility mode for grow.

Here the approach has been used for the sub graph mining for the getting the result. We followed the approach of the FSM (Frequent Sub graph Mining). The objective of FSM is to extract all the frequent sub graphs, in a given data set, whose occurrence counts are above a specified threshold.

After the set the threshold value compare it with the user input on the bases of three main factor which are the casual agents of the disease such as soil PH value, Temperature and RH Value. On this bases we analysis the data for predicting the compatible mode for the growth of the plant.

In future work we can add this approach for the large scale of the data and work with more than three factor for the prediction for efficient results.

References

[1] Shradha Chawla, Vivek Panwar and Preeti Kakkar, Aggregation and then classification of medical data using subgraph algorithm, International Journal of Computer Engineering and Applications, Volume VI, Issue II, May 2014.

[2] Saif Ur Rehman, Asmat Ullah Khan, Simon Fong, Graph Mining: A Survey of Graph Mining Techniques, IEEE, 2012.

[3] Renzo Angles, A Comparison of Current Graph Database Models, IEEE 28th International Conference on Data Engineering Workshops, 2012.

[4] Zhaonian Zou, Jianzhong Li, Member, IEEE, Hong Gao and Shuo Zhang, Mining Frequent Subgraph Patterns from Uncertain Graph Data, IEEE Transactions On Knowledge And Data Engineering, Volume 22, No. 9, September 2010.

[5] Chad Vicknair, Michael Macias, Zhendong Zhao, Xiaofei Nan, A Com-

e-ISSN: 2321-1814
ISBN (Print): 978-1-62951-619-6

parison of a Graph Database and a Relational Database, Oxford, MS, USA,2010.

[6] Andres Gago-Alonsoa, Alfredo Munoz-Brisenoa, Niusvel Acosta-Mendozaa, New proposal for Graph Classification using Frequent Geometric Subgraphs.

[7] J. E. Yuen and G. Hughes, Bayesian analysis of plant disease prediction, University of Edinburgh, Edinburgh EH9 3JG, UK,2012.

[8] Ming Syan Chan,Jiawai Han,Philip S. Yu, Data Mining: A Overview From database Prospective, IEEE Transactions On Knowledge And Data Engineering, Volume. 8, No. 6, December 1996.

[9] Mike Buerli, The Current State of Graph Databases, Cal Poly San Luis Obispo, 2012.

[10] Maryam Gholami, Afshin Salajegheh, A Survey on Algorithms of Mining Frequent Subgraphs, International Journal of Engineering Inventions, Volume 1, Issue 5, September 2012.

[11] K. Lakshmi and Dr. T. Meyyappan, A Comparative Study Of Frequent Subgraph Mining Algorithms, International Journal of Information Technology Convergence and Services (IJITCS) Vol.2, No.2, April 2012.

[12] Neo4j, http://neo4j.org/

[13] http://www.linkeddatatools.com/introducing-rdf/

[14] http://ausveg.com.au/intranet/technical-insights/cropprotection/fungal-diseases.htm/

[15] http://davesgarden.com/guides/articles/view/721/#b/

[16] http://www.britannica.com/EBchecked/topic/463327/plant-disease/63344/Symptoms-and-signs/

[17] Sujata J. Suryawanshi, Prof. S.M. Kamalapur, Algorithms for Frequent Subgraph Mining, International Journal Of Advance Research in Computer and Communication Enginerring,Vol 2,Issue 3, March 2013.