International Journal of Computer Engineering & Technology (IJCET)
Volume 8, Issue 6, Nov-Dec 2017, pp. 95–105, Article ID: IJCET_08_06_012
Available online at
http://www.iaeme.com/ijcet/issues.asp?JType=IJCET&VType=8&IType=6
Journal Impact Factor (2016): 9.3590(Calculated by GISI) www.jifactor.com
ISSN Print: 0976-6367 and ISSN Online: 0976–6375

DEVELOPMENT OF KNOWLEDGE MINING TECHNIQUES FOR SPATIAL DECISION SUPPORT SYSTEM

S. D. Samantaray
Department of Computer Engineering, College of Technology, Pantnagar, India

Dr. Durgesh Pant
School of Information Technology, Uttarakhand Open University, Dehradun, India

ABSTRACT

The advent of remote sensing and survey technologies over the last decade has dramatically enhanced our capabilities to collect terabytes of agricultural geographic data on a daily basis. However, the wealth of remotely sensed data cannot be fully realized when information implicit in data is difficult to discern. Though, lot of research has been conducted in the area of spatial data mining, but still a very few works deal with knowledge discovery in agriculture spatial databases. Development of efficient Knowlege Mining approaches is an important and challenging area for developing efficient and effective Spatial Decision Support Systems (SDSS). With the availability of advanced computer and ICT technologies there is a great potential for the development of effective Knowledge Mining techniques for the development of better Decision Support systems to support Agriculture. This may include development of specific Knowledge Management techniques that will radically improve access, decision making, resource allocation, management strategies, and promulgate process know-how for overall performance improvement. Spatial data mining techniques are regarded simply as functional extensions of conventional data mining techniques, constructed on the same first principles but using algorithms designed specifically to handle the characteristics and requirements of spatial data and spatial data mining. The paper presents Spatial Association Rules mining techniques to extract interesting correlations, frequent patterns, and associations among sets of items in the spatial databases by integrating data mining and GIS techniques to extract patterns and rules from the data. It helps in developing a Spatial Decision Support System by discovering the possible influence of selected geographical objects on yellow rust of wheat crop. Further the Knowledge miner algorithm has been improved to take less number of scans and thus, improve the execution time of the developed system

Key words: Spatial Data Mining, Spatial Association rules, Apriori Algorithm Spatial Decision Support System (SDSS), Yellow rust.

Cite this Article: S. D. Samantaray and Dr. Durgesh Pant, Development of Knowledge Mining Techniques For Spatial Decision Support System. *International Journal of Computer Engineering & Technology*, 8(6), 2017, pp. 95–105.
http://www.iaeme.com/ijcet/issues.asp?JType=IJCET&VType=8&IType=6

1. INTRODUCTION

The advent of remote sensing and survey technologies over the last decade has dramatically enhanced our capabilities to collect terabytes of agricultural geographic data on a daily basis. However, the wealth of remotely sensed data cannot be fully realized when information implicit in data is difficult to discern. Though, lot of research has been conducted in the area of spatial data mining, but still a very few works deal with knowledge discovery in agriculture spatial databases. Development of efficient Knowledge Mining approaches is an important and challenging area for developing efficient and effective Spatial Decision Support Systems (SDSS). With the availability of advanced computer and ICT technologies there is a great potential for the development of effective Knowledge Mining techniques for the development of better Decision Support systems to support Agriculture. This may include development of specific Knowledge Management techniques that will radically improve access, decision making, resource allocation, management strategies, and promulgate process know-how for overall performance improvement. Spatial data mining techniques are regarded simply as functional extensions of conventional data mining techniques, constructed on the same first principles but using algorithms designed specifically to handle the characteristics and requirements of spatial data and spatial data mining

The paper presents Spatial Association Rules mining techniques to extract interesting correlations, frequent patterns, and associations among sets of items in the spatial databases by integrating data mining and GIS techniques to extract patterns and rules from the data. It helps in developing a Spatial Decision Support System by discovering the possible influence of selected geographical objects on yellow rust of wheat crop. Further the Knowledge miner algorithm has been improved to take less number of scans and thus, improve the execution time of the developed system

1.1. Spatial Association

In a typical spatial data base, a large number of associations may exist among objects but only a umber of small n them are of real significance or interest to users. Thus, Korperski and Han suggested that it is necessary to use two thresholds, namely minimum confidence and minimum support, to filter out associations describing a small percentage of objects and rules with low confidence. Korperski and Han [12, 13] defined the rule of spatial association as $P1 \wedge P2 \ldots\ldots \wedge Pm \rightarrow Q1 \wedge Q2 \ldots \wedge Qn$ (c%, s %), where at least one of the predicates P1,…Pm, Q1,..Qn is a spatial predicate, and c% and s% are confidence and support. There are various kinds of spatial predicates that can be used to constitute a spatial association rule, including topological relations (for example, intersect, overlap, disjoint), spatial orientation (for example, east_of, left_of) and distance expressions (for example, close_to, far_away_from). They proposed a top-down tree search approach that made use of a spatial concept hierarchy to control the spatial association rule mining process by progressively eliminating uninteresting associations. Co-location is a special type of spatial association. It is defined as the occurrence of two or more spatial objects at the same location or at significantly close proximity to one another. Co-location differs from ordinary spatial associations in that there is no natural notion of a transaction between the antecedent and consequence spatial objects, and user-defined neighborhood information is an important factor in constructing co-location rules [15].

Spatial autocorrelation is the method used in exploratory spatial data analysis to measure the correlation of a variable with itself (that is, the relationship among multiple occurrences of values on the same variable over space). Spatial autocorrelation is present when occurrences of similar values cluster together spatially. There are two common measures of spatial autocorrelation. One of these is Moran's I, which has a value between –1 and +1, where a positive value implies similarity among nearby or neighboring data objects. A negative value implies dissimilarity, and a zero value indicates independent and random distribution of data values. The other commonly used measure is Geary's C, which is suitable for use in the analysis of aggregated spatial data, such as population statistics reported by census tracts and land use figures reported by planning zones. Spatial autocorrelation is used to measure the strength of the relationships among spatial objects of the same type. It helps to uncover the extent to which the occurrence of an event or feature at a certain point in space will constraint, or make more probable, the occurrence of another event or feature in its neighborhood. Thus, spatial autocorrelation analysis is particularly useful for knowledge discovery about spatial association [16].

2. SPATIAL ASSOCIATION RULE MINING FOR DISEASE MANAGEMENT IN WHEAT CROP

Wheat being the world's most important food crop feeds about 2.5 billion poor peoples in around 90 countries and is a crucial source of calories and protein. But the demand for wheat currently outstrips the world's ability to produce it and the reason is reduction in the production of wheat crop because of increasing new kinds of fungus infections, such as wheat rusts, that are killing these crops. There is a need to deploy controlling and monitoring systems that reduces the adverse effect of yellow rust disease on wheat crop. Yellow rust spread due to weather parameters like humidity, wind speed, temperature, sun shine hours and rainfall etc. An, attempt has been made by incorporating computer intelligence methods including spatial association rule mining to find associations between different types of rust disease and there associations with location, weather parameters etc. This helps in devising strategies for disease and crop management [6,7,8].Computer intelligence and Knowledge Mining techniques can be used to develop more effective decision support system for predictions of disease and to suggest precautionary measures.

2.1. Pre-Process

The remotely sensed data of wheat crop (quac.ndvi file) is opened in GIS software and classification is done for the wheat yellow rust disease data using decision tree on the basis of percentage of vegetation index values. Based on the vegetation index values, the crop is classified as no wheat, healthier and diseased. The diseased crop was further classified into various severity levels viz. mild, moderate and severe. After executing the classification, on the basis of majority/minority analysis post classification of the above results is done. Then the file is converted from raster to vector and saved as .evf. The .evf so obtained is exported as shape file using ENVI. The shape file so obtained is opened in Arc GIS as a vector layer and is used for applying spatial predicate queries for further analysis.

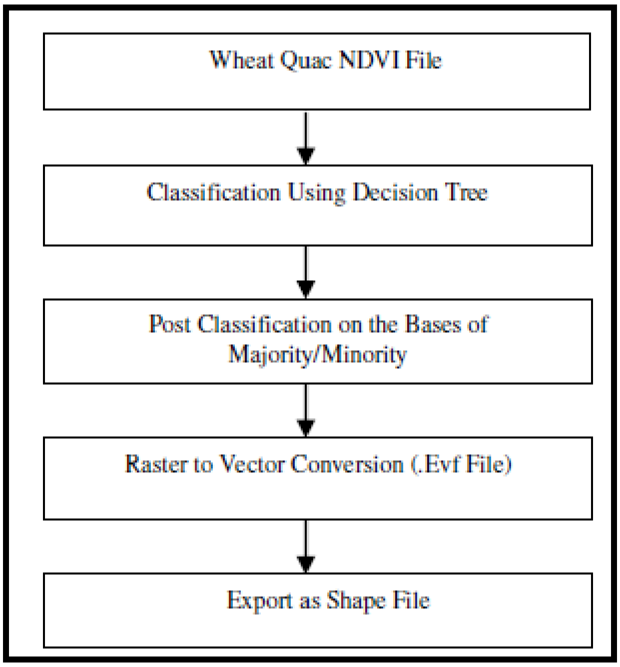


Figure 1 Flow chart of pre-processing steps

The data is analysed for extracting spatial relationship among various features. Spatial queries is used for finding the details of the yellow rust disease that had occurred within a given distance from water bodies. (e.g. 250 meters or 500 meters from water bodies). Figure 2 and Figure 3 shows the process of query application:

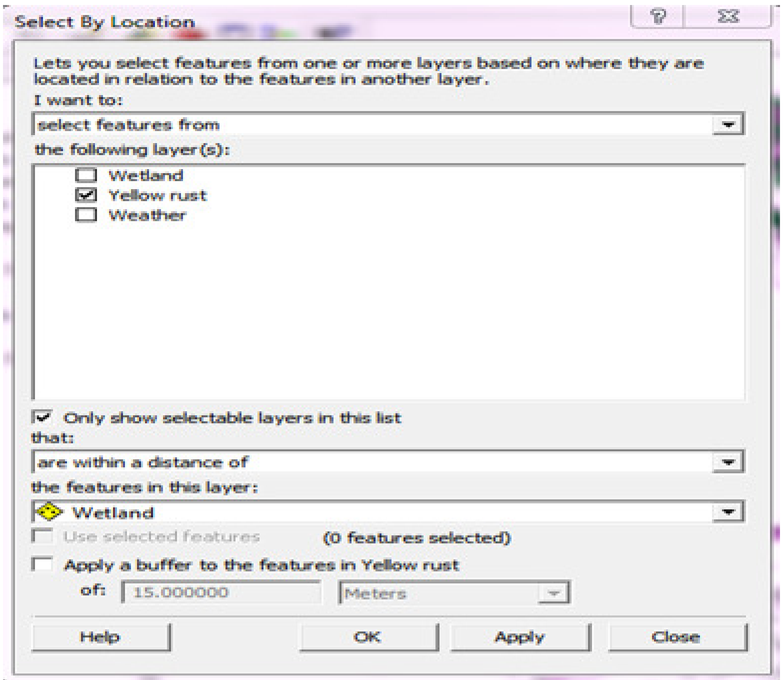


Figure 2 Spatial queries in select by location option

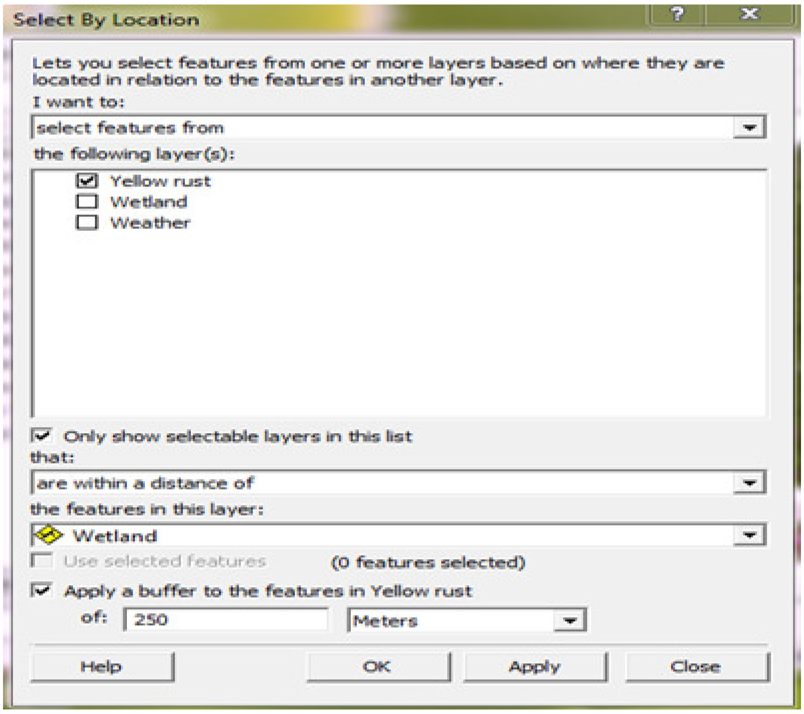


Figure 3 Selecting appropriate buffer value of 250 meter

The points in the Figure 4 show the existence of yellow rust that occurs within a distance of 250 meters from the wetlands areas.

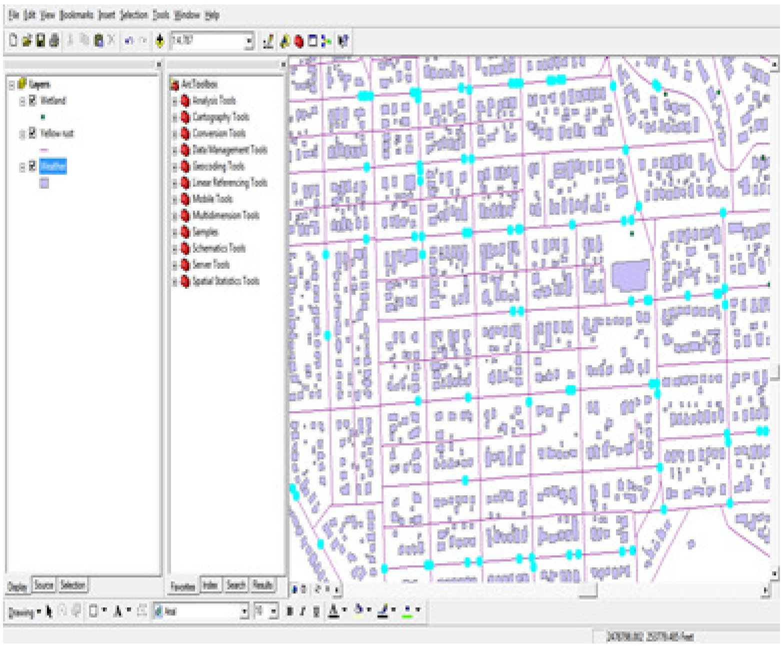


Figure 4 Yellow rust 250 meters from wetland

The attribute table is modified with two more attributes dis_wet and dis_temp added into it. The dis_wet attribute is assigned a value _dw_low for all those records where rust occurred within a distance of 250 meters from the wetlands. The rest records were assigned a value –db_high. The_dis_temp attribute was assigned a value _dt_low for all those records where rust occurred at temperature range 18-25 degree Celsius. The rest records were assigned a value _dt_high. The databases now were exported to dbf format. Since the dbf file so generated was a spatial database file, it was converted into a form suitable for carrying out association rule mining. For further processing of the data primary attribute id of integer type was created to uniquely identify each level of rust in case of wheat rust and type of land in case of land use. Next, in case of wheat crop, separate attributes for each level of rust was created e.g. severe, moderate, mild, healthier, no wheat. Default value of the attributes was set as 0. A bit value of 1 was set to each particular tuple according to the level of rust. For close_wetland and close_temprature attribute the value was set to 1 if dis_wet and dis_temp was dw_low and dt_low respectively. After the execution of the query the database was inserted into the new table. This transaction database is used by Spatial Data Mining algorithm. Weather data containing all weather parameters (humidity, temperature, cloud, wind speed) for five years was acquired along with latitude and longitude and stored in different excel file for every year. Each weather parameters individually for every year was interpolated and preprocessed in GIS software. The weather data was interpolated in order to obtain the values of unsampled point using sampled points of the area of study.

2.2. Spatial Queries

Spatial queries through –"select by location option" is used for finding the yellow rust that had occurred within a distance of 15 meters from the rivers and that occur at the temperature range of 18-25 degree Celsius.

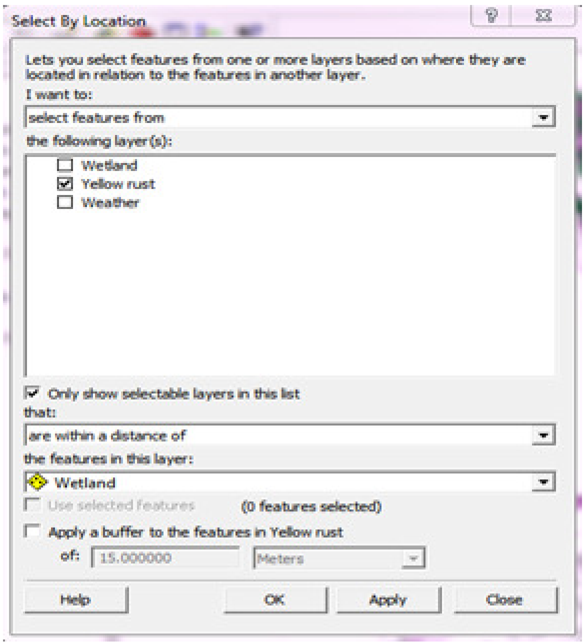


Figure 5 Spatial queries in select by location option

This option requires a database (.shp format) from which attributes or features are to be selected. So at first the yellow rust wheat database was selected and then the _are within a distance of filter was selected and was made to set the buffer of 15 meters from the wet lands database.

2.3. Database Conversion

The attribute table was modified and two more attributes dis_wet and dis_temp was added into it. The _dis_wet attribute was assigned a value _dw_low for all those records where rust occurred within a distance of 15 meters from the wetlands. The rest records were assigned a value –db_high. The_dis_temp attribute was assigned a value _dt_low for all those records where rust occurred at temperature range 18-25 degree Celsius. The rest records were assigned a value _dt_high. The database now was exported to dbf format. Since the dbf file so generated was a spatial database file, it was converted into a form suitable for carrying out association rule mining. For further processing of the data primary attribute id of integer type was created to uniquely identify each level of rust. Next separate attributes for each level of rust was created e.g. severe, moderate, mild, healthier, no wheat. To do this following spatial query as was applied:

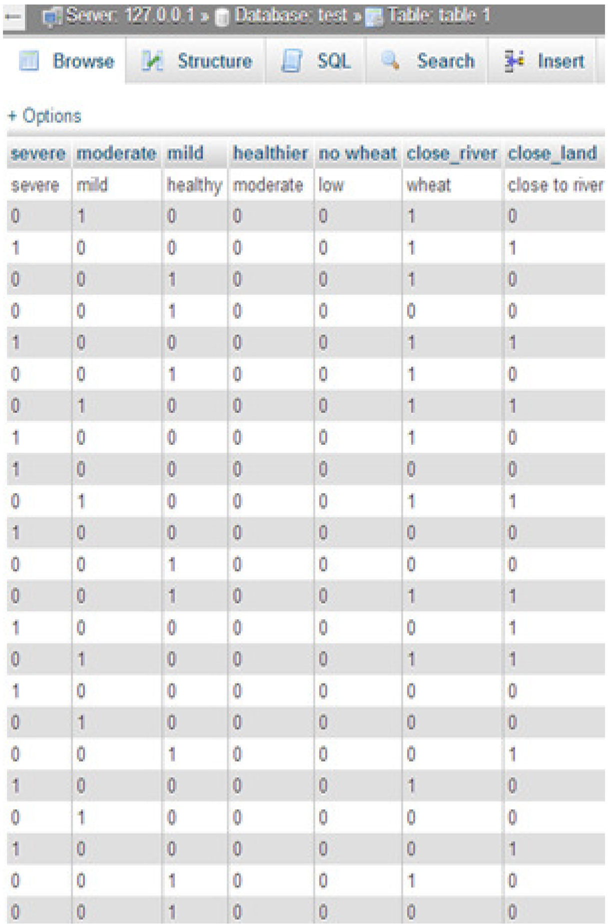
ALTER TABLE test ADD COLUMN _SEVERE BIT NULL DEAULT 0;

In this query default value of the attributes was set as 0. A bit value of 1 was set to each particular tuple according to the level of rust. For close_wetland and close_temprature attribute the value was set to 1 if dis_wet and dis_temp was dw_low and dt_low respectively. The application of queries was as under:

UPDATE test SET SEVERE=1 WHERE RUST= severe;

UPDATE TEST SET CLOSE_WETLAND = 1 WHERE dis_wet = _dblow;

Now the final query was applied to insert only the required attributes into the table.



severe	moderate	mild	healthier	no wheat	close_river	close_land
severe	mild	healthy	moderate	low	wheat	close to river
0	1	0	0	0	1	0
1	0	0	0	0	1	1
0	0	1	0	0	1	0
0	0	1	0	0	0	0
1	0	0	0	0	1	1
0	0	1	0	0	1	0
0	1	0	0	0	1	1
1	0	0	0	0	1	0
1	0	0	0	0	0	0
0	1	0	0	0	1	1
1	0	0	0	0	0	0
0	0	1	0	0	0	0
0	0	1	0	0	1	1
1	0	0	0	0	0	1
0	1	0	0	0	1	1
1	0	0	0	0	0	0
0	1	0	0	0	0	0
0	0	1	0	0	0	1
1	0	0	0	0	1	0
0	1	0	0	0	0	0
1	0	0	0	0	0	1
0	0	1	0	0	1	0
0	0	1	0	0	0	0

Figure 6 Bit values for attributes

2.4. Spatial Knowledge Mining Algorithm

After exporting the database to csv file, the knowledge mining algorithm is applied. In order to generate Strong Association rules following algorithm based on A-priori algorithm is applied:

Input

D// Database in the form of a text file including spatial and non-spatial attributes.

S// Minimum support level entered by the user.

C// Minimum confidence entered by user which are to be included in the final rules.

Output

R// Spatial association rules.

Step1: Procedure find_ spatial association rules (d).

Step2: From the given database use only those attributes specified by Q to generate L1= {frequent items};

Step 3: for (k=2; L (k-1)! = ⦰; k++) do begin

C (k) = candidate generated from L (k-1);

for each transaction in database do

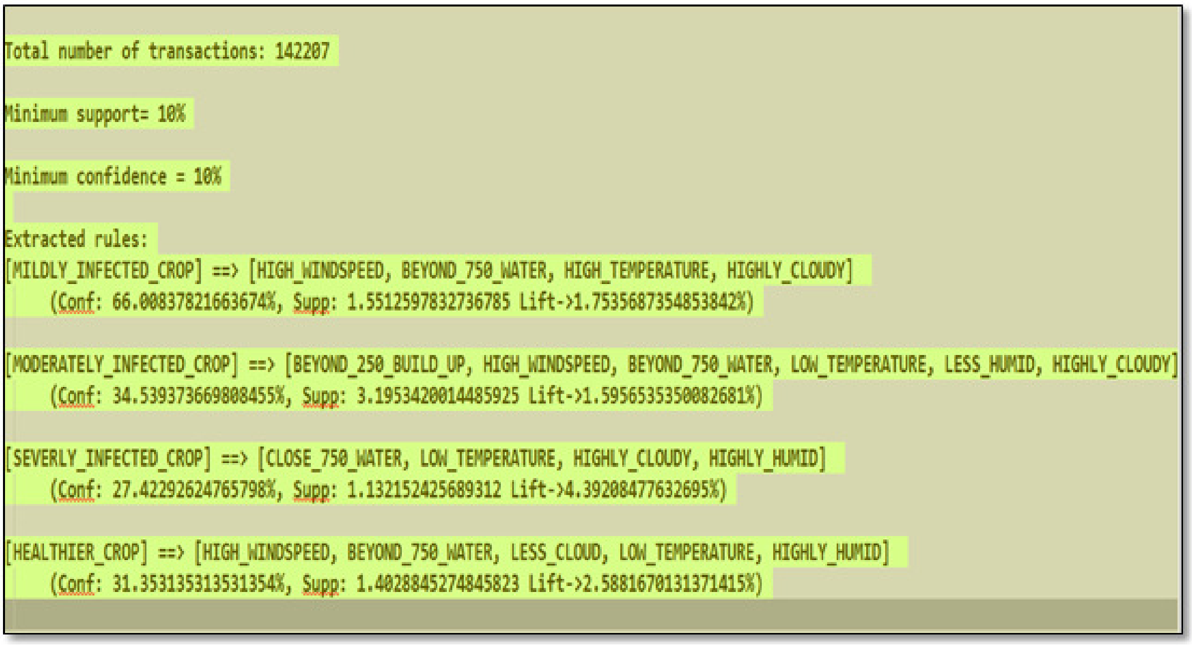
Increment the count of all candidates in C (k) that contained L (k) = candidate in C (k) with minimum support S.

Step 4: Return union L (k);

3. RESULTS AND DISCUSSION

Figure 8 represents the snapshot of some of the selected strong spatial association rules that was generated after applying the mining algorithm. The rule considered as strong contains high value of confidence and lift ratio. Where Lift is defined as follows:

Lift of X → Y = Confidence (X U Y) / Support(Y)



```
Total number of transactions: 142207

Minimum support= 10%

Minimum confidence = 10%

Extracted rules:
[MILDLY_INFECTED_CROP] ==> [HIGH_WINDSPEED, BEYOND_750_WATER, HIGH_TEMPERATURE, HIGHLY_CLOUDY]
    (Conf: 66.00837821663674%, Supp: 1.5512597832736785 Lift->1.7535687354853842%)

[MODERATELY_INFECTED_CROP] ==> [BEYOND_250_BUILD_UP, HIGH_WINDSPEED, BEYOND_750_WATER, LOW_TEMPERATURE, LESS_HUMID, HIGHLY_CLOUDY]
    (Conf: 34.539373669808455%, Supp: 3.1953420014485925 Lift->1.5956535350082681%)

[SEVERLY_INFECTED_CROP] ==> [CLOSE_750_WATER, LOW_TEMPERATURE, HIGHLY_CLOUDY, HIGHLY_HUMID]
    (Conf: 27.42292624765798%, Supp: 1.132152425689312 Lift->4.39208477632695%)

[HEALTHIER_CROP] ==> [HIGH_WINDSPEED, BEYOND_750_WATER, LESS_CLOUD, LOW_TEMPERATURE, HIGHLY_HUMID]
    (Conf: 31.353135313531354%, Supp: 1.4028845274845823 Lift->2.5881670131371415%)
```

Figure7 Selected strong spatial association rules

3.1. Comparison of Proposed and Existing Approach

In order to determine the impact of various parameters on the execution time and to compare the execution time of the proposed approach and the existing approach various experiments were performed.

Case1: Comparison between varying minimum support and execution time and keeping minimum confidence, number of items and number of transactions constant. A series of experiments were run with increasing number of support. The results are showed in the Figure 8. It can be observed that as we increase the support the execution time decreases. Other than this for the smaller value of support our algorithm perform better than the original algorithm while as we increase the support both takes approximately same amount of time.

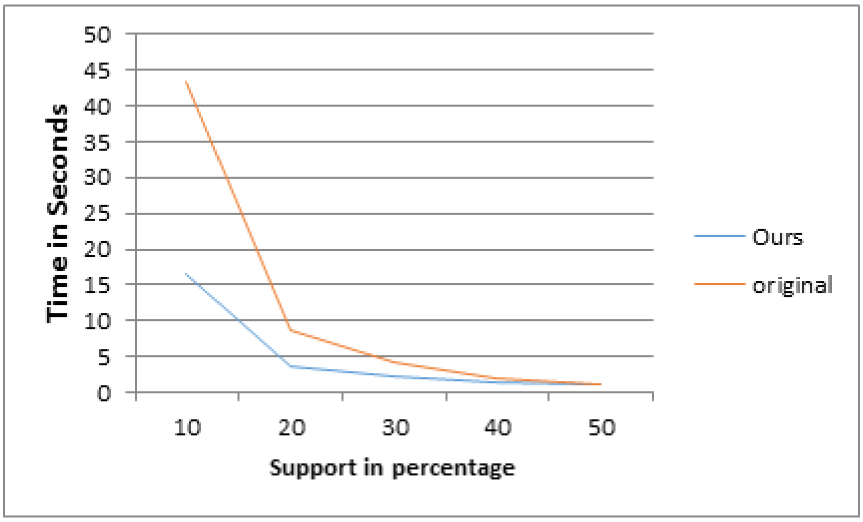


Figure 8 Graph between execution time and support

Case2: Comparison in order to determine the impact of varying minimum confidence on the execution time when minimum support, number of transactions and number of items are fixed: A series of experiments were run with increasing number of confidence. The results are showed in the Figure 9.

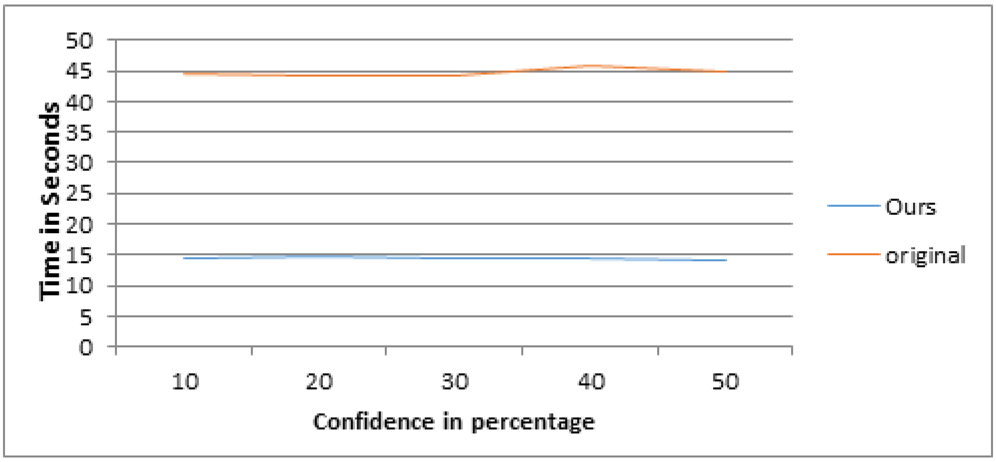


Figure 9 Graph between execution time and confidence

As evident from Figure 9, there is no significant impact of varying confidence in the execution time of the algorithm. As can be observed from the Figure the graph between execution time and different confidence is almost constant. However, it can be observed that our algorithm perform well than the existing algorithm.

Case3: Comparison between varying number of transactions and execution time when minimum confidence, minimum support and number of items are fixed: A series of experiments were run with decreasing number of transactions. The results are showed in the Figure 10.

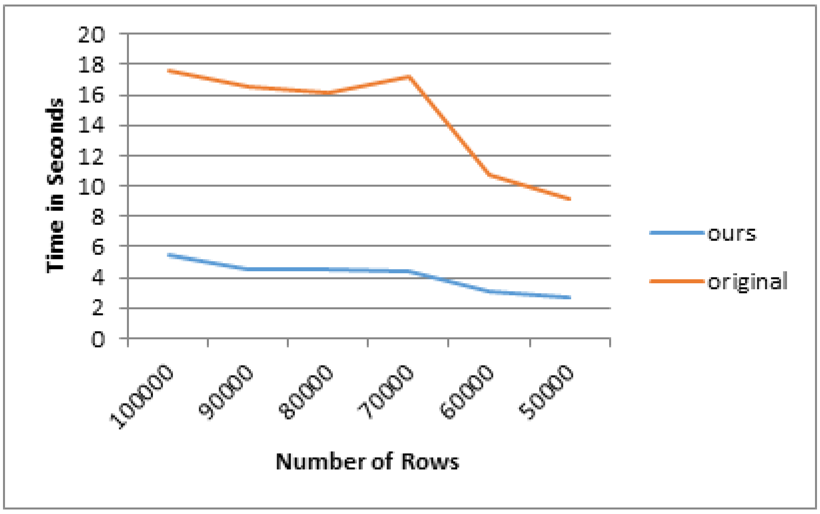


Figure10 Graph between execution time and number of rows (transactions)

As we decrease the number of rows o transactions the execution time decreases. Also it can be observed that our algorithm performs better than the original algorithm.

Case4: Comparison between varying number of items and execution time considering minimum confidence, minimum support and number of transactions constant:

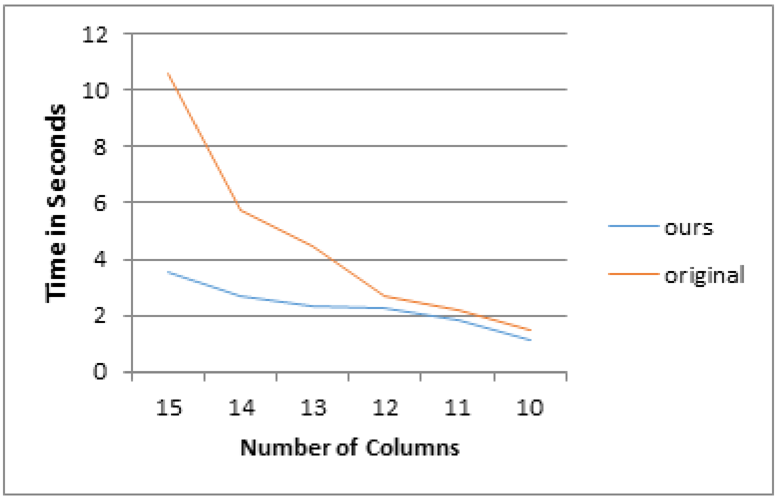


Figure 11 Graph between execution time and number of columns (items)

It can be observed that as we reduce the number of items the execution time decrease. Also at less number of columns our algorithm and original algorithm perform approximately same but as we increase the number of items our algorithm perform far better than the existing original algorithm.

4. CONCLUSION

The paper presents Spatial Association Rules mining techniques to extract interesting correlations, frequent patterns, and associations among sets of items in the spatial databases by integrating data mining and GIS techniques to extract patterns and rules from the data. It focuses on a hybrid approaches to generate association rules from Spatial Databases. It helps in developing a Spatial Decision Support System by discovering the possible influence of selected geographical objects on yellow rust of wheat crop. Further the Knowledge miner algorithm has been improved to take less number of scans and thus, improve the execution time of the developed system for implementation, GIS tools have been used. Remotely Sensed Data of Uttarakhand state of India has been used as the data sets. The Association Rule mining algorithms have been developed in JAVA and MATLAB to extract the Strong Association Rules.

REFERENCES

[1] Sashi Shekhar and Huang.Y., "Discovering Spatial Collocation Patterns: A summary of results", Advances in spatial and temporal database proceedings, vol.2121 of Lecture Notes in Computer Science, Springer-verlag, Berliln, pp. 236-256, 2001.

[2] Srikant .R and Agarwal R., "Mining Generalized Association Rules", Proceedings of the International Conference on Very Large Databases, San Francisco, CA: Morgan Kanfman Press, 1995.

[3] Zhang X.W., Su, F.Z., Shi, Y.S. and Zhang, D.O., "Research on Progress of Spatial Association Rule Mining", Journal of Progress in Geography, vol.26(6), pp.119-128 ,2007.

[4] Chawla Sanjay, Shekhar Shashi, Wu Weili and Ozesmi Uygar, "Modelling Spatial Dependencies for Mining Geospatial Data"

[5] D. Rajesh, "Application of Spatial Data Mining for Agriculture", International Journal of Computer Applications, vol.15 (2), 2011.

[6] S.Suriya and R.Deeplalshmi, "A Complete Survey On Association Rule Mining With Relevance To Different Domain", International Journal of Advance Scientific And Technical Research vol. 1, 2012.

[7] Report on Strategies to Reduce the Emerging Wheat Stripe Rust Disease, International Centre for Agricultural Research in the Dry Areas, 2011.

[8] F.Annevelink, "AgroInformatics In NetherLands,Institute of Agricultural Engineering", AgroInformatica, Jargang 3, no.5.

[9] Z.Yueling, and C.Liying, "The Application of the Spatio Temporal Data Mining Algorithm in Maize Yield Reduction", Elsevier, Science Direct, pp .7, 2011.

[10] C. Hong and Wang Xiaoyan, "Research on GIS- based Spatial Data Mining, IEEE, 2011.

[11] Koperski K., and Adhikary J., Han J. "Knowledge Discovery in Spatial Databases: Progress and Challenges", Proc. SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, Technical Report 96-08, University of British Columbia, Vancouver, Canada., 1996

[12] Koperski, K. and Han, J. "Discovery of Spatial Association Rules in Geographic Information Databases", In Advances in Spatial Databases (Lecture Notes in Computer Science 951), by Egenhofer, M. J. and Herring, J. R. (Eds.), Berlin: Springer-Verlag., 1995

[13] Koperski K., and Han J. "Discovery of Spatial Association Rules in Geographic Information Databases", Proc. 4th Int. Symp. on Large Spatial Databases (SSD '95), Portland, ME, pp 47-66., 1995