Research article Available online www.ijsrr.org ISSN: 2279–0543

International Journal of Scientific Research and Reviews

Spatial Data Mining: An Effective Approach in Agriculture Sector

Goyal Hemlata[*], Sharma Chilka[1] and Joshi Nisheeth[2]

[2*]Department of Computer Science, Apaji Institute of Mathematics and Applied Computer Technology (AIM and ACT), Banasthali University, Tonk (Raj.) India
[1]Department of Remote Sensing, School of Earth sciences, Banasthali University, Tonk (Raj.) India

ABSTRACT

Spatial Data Mining (SDM) is a multifaceted dealing with dataset having spatial and nonspatial relations in geospatial databases. SDM is an emerging technique to extract latent and implicit trends in geospatial data to acquire learning and expert decision making. A geospatial technology is an integration of remote sensing, global positioning system and geographical information system which is rapidly becoming important to facilitate interpretation, analysis, mapping, presentation, planning and monitoring massive amounts of data related to sustainable agriculture. Traditional data mining could not determine the potential knowledge. Geographic Information System serves the platform for agricultural monitoring resulting in map information and transforms reaming tables into graphic maps, deals with spatial geo-referenced database. SDM technology provides the platform to acquire implicit rules, trends, and relevance in real world spatial and nonspatial dataset which is the most significant issue in agriculture sector. The objective of this paper is to focus light on various SDM approaches and their usage in agriculture sector to carry out spatial analysis of GIS data.

KEYWORDS: GIS, Spatial and Nonspatial, SDM, SDM tools and techniques, SDM applications in agriculture

***CORRESPONDING AUTHOR**

HEMLATA GOYAL

Research Scholar, Department of Computer Science, Apaji Institute of Mathematics and Applied Computer Technology (AIM and ACT), Banasthali University-Niwai-Tonk

Email: singhalhem@gmail.com

1. INTRODUCTION

Spatial data deals with real world objects in 3-d geometrical space and give the information about to existence of where. The Spatial objects are distinctively recognized by its geo-referenced latitude and longitude coordinates and these real earth objects, represented in geodatabase in the form of point, line, polygon, surface and network.. In general a GIS is used to accumulate data and information with respect to geographical locations on the earth surface, retrieval and manipulation of it[1].

SDM is the process to discover the knowledge from huge geospatial dataset means to extract the unknown, valuable spatial relationship, patterns, or trends, not explicitly stored in spatial databases.

The agriculture is extremely branch out and varied in terms of its crops management practices, climate scenarios, soil, water, temperature, rainfall etc with both space and time. Integrated GIS and SDM has performed a huge range of applications in agriculture sector for e.g. yield prediction, precision agriculture, soil mapping etc. Detailed information about SDM in agriculture sector will be discussed in rest of the sections. Section 2 throws light on real world agriculture applications of SDM. Various spatial mining techniques were discussed in Section 3. The details of SDM tools in Section 4 and SDM results, conclusion and future directions were discussed in Section 5.

2. APPLICATIONS OF SDM IN AGRICULTURE

Many real world applications used SDM in agriculture sector for prediction, productivity, weed control etc. The following subsections provide such information in detail.

2.1 Crop yield prediction

Crop yield prediction has necessary and helpful task for sustainable agriculture management and gives the benefit to increase the performance of agriculture system. Towards this direction a knowledge inference system was explored for prediction of wheat crop using remote sensing dataset. Neural network approach was applied on input variables such as rainfall, temperature, VCI etc. to generate forecast pertaining to wheat yield[2,3]. To estimate the crop condition of Iowa state, US, the non-linear optimization technique was used on GIS and satellite data[4]. A twofold model[5], in first stage used association rule mining on the past agriculture data and generate rules from frequent item sets to generate the best association and in next stage they applied genetic algorithm to predict output as an optimized agriculture crop. Density based clustering technique used to collect rainfall (x) and yield(y) as independent and dependent variable respectively to draw the relationship in between for correlation and regression analysis to develop yield prediction model for some crops[6]. The Extensible Crop Yield Prediction Framework about SDM was developed to achieve precision for rice and sugarcane crops to make strategic decisions[7]. To categorize the cotton seeds quality depends on germination period contain soil, weather parameters was analyzed by different classifier and proposed J48 classifier is best for it[8]. Crop yield prediction for Ghana district had used simple, double, damped-trend linear exponential smoothing and ARIMA model of time-series[9].

2.2 Precision agriculture

Precision farming is most of the important activities in agriculture. It is considered to be a group of technologies consisting of vegetation, climate, pest control, market planning etc. In order to this a cross-validation learning approach was used to overcome the problem with traditional mining methods and increase the predictive performance of yield prediction[10]. Geospatial outlier detection techniques were used for filtering erroneous crop yield data, and assessed their performance via quantitative methods for yield and soil mapping[11]. The spatial simulation based SDSS namely Apollo DSS was designed for precise agriculture in Central Iowa, United States to improve corn grain yield[12]. A decision making system for various locales and harvests was produced to choose the sort of seed to grow, when and the amount to flood, rate of utilization of fertilized and product yield[13]. The web based SDSS called as "RiceCheck" was developed for Paddy Precision farming in Tanjung Karang, Malaysia to improve the productivity and efficiency of rice production[14]. The web based SDSS is developed for farm level decision making for regular and irregular fields in Guangzhou, China, which helped farmers to select suitable fertilizer with quantity that should be used as per the crop requirement[15]. The SDSS was also developed for cotton crop precision farming using soil type, soil fertility index, weather, nitrogen and site specific water application data. The after-effects of progressive simulation models demonstrated that there is potential increment in the yield with ideal nitrogen and water rates[16].

2.3 Crop productivity improvement

Soil fertility is an important parameter in crop productivity improvement. The SDSS named as “SIGAA” (Advanced Agricultural Geographical Information System) was developed in Romania, which helped to consider the interactions between different environmental factors to decide the crop suitability and increase in crop yield[17]. The world's largest project on SDSS named as "SDSS to manage and analyse citrus production information management" was conducted at “Three Gorges area”, of south-western part of China, which is one of the optimum citrus productions in world. The project used spatial and non-spatial data to increase agriculture production and improve crop management decisions[18]. The SDSS named as "Sugarcane Information and Management System", was developed in Thailand for decision making and planning to improve sugarcane production based on the parameters such as planting date, crop type, water resource, fertilizer application rate[19]. The biophysical parameters are one of the important parameters for the crop yield in agriculture. The SDSS named as "AGROLAND" was developed in Germany to evaluate the biophysical agricultural land resources. It helped decision makers to calculate present and the prospect agricultural land resources based on marginality index (MI)[20]. K means algorithm for cluster analysis applied on spatial dataset rainfall and temperature to enhancement of crop yield in the field of agriculture[21]. Genetic algorithm used to find out the best association rules and their optimization to increase the crop productivity[22]. Studies conducted by researchers[23] in Karminagar-AP analysed major crop yields of Karminagar like paddy, soyabean, turmeric etc. by generating a hybrid model of clustering and classification data mining techniques.

2.4 Suitable area Identification to crop cultivation

Identification of suitable areas is very important to increase the crop productivity in agriculture sector. In order to this a farmer prediction system was developed to classify the crop appropriate type of soil using nonlinear neural networks mapping structure[24]. The multi-Criteria Evaluation (MCE) based SDSS was developed in china to identify the appropriate areas for the cultivation of different types of oats based on climatic and topographic attributes. The web based support for soil and area suitability improved oat crop production for that area[25].

2.5 Weed control, fertilizer and pest management

To streamline the yield production, Agribusiness Pest Management is a vital range of concern. Hemlock Looper DSS was designed in Newfoundland, Canada for managing of eastern hemlock looper pest[26]. The decision tree based SDSS helped to decide management of spraying action particular farm maintaining environmental balance[27]. Better crop yield is highly affected on inputs of fertilizer and herbicides. How regression analysis and cluster analysis used for assessment of consumption of fertilizer nutrients on the yield[28] and how data mining techniques like Hierarchical Clustering, Density-based Clustering and Association rule mining incorporated with agriculture together with pest, pest scouting, and climate parameters were valuable for optimization of pesticide usage and better management[29].

2.6 Soil mapping and classification

Soil mapping is the geo-encoding based process for classifying soil types and other soil properties in a given area. Several rule induction techniques[30] were applied to enhance the quality of soil map. A study[31] had discussed about several issues affect the reliability and usefulness of traditional soil survey process and its products and used fuzzy logic for their research. Agricultural soil profiles accuracy is measured by data mining Expectation-Maximization algorithm and Farthest- First clustering algorithms[32].

2.7 Water resource planning

SDM has also played very important role in water resource management. WebGIS based SDSS is developed in northwest of China provides support to send out water resources by monitoring vegetation, crops in Shule River Basin. The Service DSS part of it analyzes, the zone contained how much water staged and choose the path of infrastructure construction for water services and decision making in crop cultivation[33]. GIS based generalized framework is developed in assessment of groundwater resources in Godavari Irrigation Project-Andhra Pradesh, India for choosing the crop pattern according to the groundwater accessibility to expand the crop production[34].

2.8 E-agriculture information management system

E-Agriculture Information Management System was developed for farmers via delivering SMS alerts regarding market and weather news to their mobile phones on daily basis to the registered farmers. And other agriculture parameter inform about fertilizers, price in market etc information sent to that farmer through SMS via SMS Gateway on seasonal basis[35].

3. SDM TECHNIQUES

SDM techniques is nothing new, just an expansion of data mining techniques applied on domain specific GeoSpatial dataset. Spatial data using traditional data mining methods such as association, classification, clustering, trend detection, outliers generates interesting facts in agriculture domain. SDM techniques deals with spatial and non-spatial objects, attributes of neighbouring objects and their spatial relations to find class, making clustering rules to detect outliers and deviations of trends, find association to extract multilevel topological relations.

3.1 Spatial rules

Spatial rules can be generated to depict the relations in between the structure of spatial object. Spatial data contains these following rules:

Spatial characteristic rules*:* It describes the dataset, for eg. in Jaipur city the mean vegetation condition index value is 50.

Spatial discriminate rules: It describes the difference between different classes of the dataset. It describes the features that separate the different classes. For eg. in Jaipur city the mean vegetation condition index value is 50 while in village, it is 70.

Spatial association rules: It is the implications of one set of data by another set. The mean vegetation condition index value is approximately 60 near to rural area in Jaipur city.

Characterization: It is the process of finding a description for a dataset or some subset thereof.

Trend detection: It is viewed as regular changes in one or more non-spatial attribute values of spatial object as move away from another spatial object. For eg. The mean vegetation condition index value is increase as the proximity from Jaipur city to the rural increases.

3.2 Spatial association rule

It discovers uncovering relationship from spatially related dataset and used to describe patterns of the database. At least one of the antecedent or the consequent should contain some spatial predicate in the rule and support and confidence are defined identically for regular association rules.

3.3 Spatial classification

It is used to partitions the classes of spatial objects. Spatial objects could be classified using nonspatial attributes, spatial predicated, or spatial and nonspatial attributes. Concept hierarchies, sampling, generalization and progressive refinement techniques can be used to improve efficiency.

3.4 Spatial clustering

It is used by groups to discover similarity in between spatial dataset according to characteristics found in the actual spatial data. Set of like elements are grouped into one clusters and elements from different clusters are not alike.

This section concentrates on strategies or procedures utilized for SDM by the researchers prior. Table 1 summarizes SDM techniques used in agriculture sector.

Table1. Summary of SDM Techniques

Author	Year	Research Area	Technique	Remarks
[1]	2017	Association Rule Mining in big satellite dataset	MRPrePost algorithm	Satellite dataset
[5]	2015	Agriculture recommender	Association rule mining and genetic algorithm	GIS data
[7]	2015	Crop yield prediction	Data discretization with hybrid model	Satellite dataset used for rice and sugarcane crop
[9]	2014	Yield prediction	Time series	Ghana district
[23]	2014	Analysis of Various Crop Yields in different Spatial Locations of Karimnagar District in AP	Hybrid technique- Clustering and classification	Spatial dataset
[35]	2014	E-Agriculture Information Management System	Birch clustering Hierarchical tree	SMS Gateway
[8]	2011	Categorize the quality of cotton seeds according to germination duration	Classification technique	Cotton seeds
[10]	2010	Precision agriculture	Cross-validation technique	Spatial dataset
[2][3]	2008	Crop yield prediction for wheat	Neural Network	Experiments are made on satellite images
[19]	2006	Management of sugarcane	GIS – Spatial query technique	Spatial dataset, Sugarcane crop
[4]	2005	Yield estimation model	non linear, multi-variate optimization method	Data collected from GIS
[6]	2014	Yield estimation	correlation and regression analysis	GIS data

4. SPATIAL DATA MINING TOOLS

There is a huge list of open source and paid SDM tools. DBMiner (DBlearn), GeoMiner, GeoDA (ESDA,STARS), Weka-GDPM, R language (sp, rgdal, rgeos), Descrates, ArcGIS (ArcView, ArcInfo, ArcEditor) are some of them. ArcGIS is paid tool and includes surface, network analysis, overlay, interpolation, analysis and geo statistical modeling techniques.

5. CONCLUSION AND FUTURE PROSPECTS

It is observed that GIS based methods are broadly connected in various zones of agriculture for e.g. in increasing crop yield, crop water supplies, precision farming, site suitability and to concentrate the effect of climatic parameters. This paper throws light on the essence of SDM, its real world applications in agriculture domain. The techniques discussed in literature for SDM were classification, association rule mining, , trend analysis, clustering, neural network, correlation, regression analysis for prediction, planning, decision making etc. specifically utilized for SDM. As a future heading we will make observational study of SDM on crop yield analysis, monitoring and forecasting.

6. REFERENCES

1. Goyal, H., Sharma, C., and Joshi, N., 2017. An Integrated Approach of GIS and Spatial Data Mining in Big Data. International Journal of Computer Application. 2017; 169-11, 0975-8887.
2. Stathakis, D., Savin, I., and Nègre, T. Neuro-fuzzy modeling for crop yield prediction. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 2013; 34, 1-4.
3. Ruß, G., Kruse, R., Schneider, M., and Wagner et al. Data mining with neural networks for wheat yield prediction. Advances in Data Mining. Medical Applications, E-Commerce, Marketing, and Theoretical Aspects, 2008; 47-56.
4. Prasad, A. K., Chai, L., Singh, R. P. et al. Crop yield estimation model for Iowa using remote sensing and surface parameters. International Journal of Applied Earth Observation and Geoinformation, 2006; 8(1): 26-33.
5. Bhaleraro O.B., Lobo L.M.R.J. Agriculture recommender using data mining techniques, International Journal of Applied Research, 2015; 5(4): 849-851.
6. Vardhan, B. V., Ramesh, D., and Goud, O. S. C. Density Based Clustering Technique on Crop Yield Prediction. International Journal of Electronics and Electrical Engineering, 2014; 2(1): 44-47.
7. Manjula, A., and Narsimha, G. XCYPF: A flexible and extensible framework for agricultural Crop Yield Prediction. In Intelligent Systems and Control (ISCO), IEEE 9th International Conference on, 2015; 1-5.
8. Revathi, P., and Hemalatha, M. Categorize the Quality of Cotton Seeds Based on the Different Germination of the Cotton Using Machine Knowledge Approach. International Journal of Advanced Science and Technology, 2011; 36: 9-14.
9. Choudhary A., Jones J. Crop yield prediction using time series models. Journal of Economic and Economic Education Research, 2014; 15-3.

10. Ruß, G., and Brenning, A. Data mining in precision agriculture: management of spatial information. Computational intelligence for knowledge-based systems design, 2010; 350-359.
11. Chu Su, P. Statistical geocomputing: Spatial outlier detection in precision agriculture.
12. Thorp, K. R., DeJonge, K. C., Kaleita, A. L., et al. Methodology for the use of DSSAT models for precision agriculture decision support. Computers and electronics in agriculture,2008; 64(2), 276-285.
13. Jones, J. W., Hoogenboom, G., Porter, C. H., et al. The DSSAT cropping system model. European journal of agronomy, 2003; 18(3), 235-265.
14. Norasma Che’Ya S. N., Mohamed Shariff A.R., Mohd Amin Mohd Soom, et al. 2011. Generating Online GIS Decision Support System For Paddy Precision Farming. In proceedings of GSD11 World Conference Spatial Data Infrastructure Convergence-Building SDI Bridges to address Global Challenges.
15. Sha, Z., and Bian, F. An integrated GIS and knowledge-based decision support system in assisting farm-level agronomic decision-making. Knowledge Acquisition 1993.
16. McKinion, J. M., Jenkins, J. N., Akins, D., Turner, et al. Analysis of a precision agriculture approach to cotton production. Computers and Electronics in Agriculture, 2001; 32(3), 213-228.
17. Pirna, C. G., and Lache, S. Integration of Soil Specific Parameters in Designing Decision Support System in Precision Agriculture. In proceedings of 10th International Scientific Conference “Engineering for Rural Development”, 2011. Latvia.
18. Wu, W., Liu, H., Dai, H., Li, W., and Sun, P. A Spatial Decision Support System for Citrus Management A Case Study of the Three Gorges Area of China. In Environmental Science and Information Application Technology, 2009. International Conference on ESIAT 2009; 1: 601-603.
19. Weerathaworn, P., Saravanan, R., and Prabpan, M. Sugarcane information and management system for mitr phol sugar group, Thailand. Sugar tech, 2006; 8(1), 1-2.
20. Laudien, R., Pofagi, M., and Roehrig, J. Development and implementation of an interactive Spatial Decision Support System for decision makers in Benin to evaluate agricultural land resources—Case study: AGROLAND.International Journal of Applied Earth Observation and Geoinformation, 2010; 12: S38-S44.
21. D. Rajesh. Application of SDM for Agriculture, International Journal of Computer Applications, 2011; 15(2): 0975 – 8887.

22. Jaiswal, A., and Dubey, G. Identifying best association rules and their optimization using genetic algorithm.International Journal of Emerging Science and Engineering, 2013; 1(7): 91-96.
23. Rao, C. M., Rao, A. A., and Reddy, N. M. Analysis of Various Crop Yields in Different Spatial Locations of Karimnagar District in AP. International Journal of Computer Science Issues, 2014;11(4): 130.
24. Sharma G., Rudrakshi. Enhancing Back Propagation Neural N/w Algorithm for crop prediction, International Journal of Advanced Research in Computer Science and Software Engineering, 2014; 4(6): 1368-1372.
25. Wan, F., Wang, Z., Li, F., et al. GIS-Based Crop Support System For Common Oatand Naked Oat in China. In International Conference on Computer and Computing Technologies in Agriculture , Springer, Boston, MA 2008; 209-221.
26. Power, J. M., and Saarenmaa, H. Object-oriented modeling and GIS integration in a decision support system for the management of eastern hemlock looper in Newfoundland.Computers and electronics in agriculture, 1995; 12(1): 1-18.
27. Cohen, Y., Cohen, A., Hetzroni, A., et al. Spatial decision support system for Medfly control in citrus. computers and electronics in agriculture, 2008; 62(2): 107-117.
28. Ghante, V. N., Kumar, L. R., Benki, A. M., et al. Perceptions of Karnataka farmers on insect pests and pest management practices in Bt. cotton. International Journal of Plant Protection, 2011; 4(2), 271-275.
29. Tripathy A.K., Adinarayana J., Sudharsan D. Geo SDM for Agriculture Pest Management - a Framework, Centre of Studies in Resources Engineering Indian Institute of Technology Bombay, Mumbai, India.
30. Moran, C. J., and Bui, E. N. Spatial data mining for enhanced soil map modelling. International Journal of Geographical Information Science; 2002; 16(6): 533-549.
31. Zhu, A. X., Hudson, B., Burt, J., Lubich, et al. Soil mapping using GIS, expert knowledge, and fuzzy logic. Soil Science Society of America Journal, 2001; 65(5): 1463-1472.
32. Armstrong, L. J., Diepeveen, D., and Maddern, R. The application of data mining techniques to characterize agricultural soil profiles. In Proceedings of the sixth Australasian conference on Data mining and analytics-2007; 70: 85-100.
33. Zhang, H., Yi, S., and Wu, Y. Decision support system and monitoring of eco-agriculture based on WebGIS in Shule Basin. Energy Procedia, 2012; 14: 382-386.

34. Chowdary, V. M., Rao, N. H., and Sarma, P. B. S. GIS-based decision support system for groundwater assessment in large irrigation project areas. Agricultural water management, 2003; 62(3): 229-252.
35. Thankachan S., Kirubakaran S. E-Agriculture Information Management System, International Journal of Computer Science and Mobile Computing, 2014; 3(5): 599 – 607.