FUZZY LOGIC BASED HYBRID RECOMMENDER OF MAXIMUM YIELD CROP USING SOIL, WEATHER AND COST

Aadithya U1, Anushya S2, Bala Lakshmi N3 and Rajeswari Sridhar4
Department of Computer Science and Engineering, Anna University, Chennai, India
E-mail: 1aadi1194@gmail.com,2anushya1995cool@gmail.com,3bala94.n@gmail.com and 4rajisridhar@gmail.com

Abstract
Our system is designed to predict best suitable crops for the region of farmer. It also suggests farming strategies for the crops such as mixed cropping, spacing, irrigation, seed treatment, etc. along with fertilizer and pesticide suggestions. This is done based on the historic soil parameters of the region and by predicting cost of crops and weather. The system is based on fuzzy logic which gets input from an Artificial Neural Network (ANN) based weather prediction module. An Agricultural Named Entity Recognition (NER) module is developed using Conditional Random Field (CRF) to extract crop conditions data. Further, cost prediction is done based on Linear Regression equation to aid in ranking the crops recommended. Using this approach we achieved an F-Score of 54% with a precision of 77% thus accounting for the correctness of crop production.

Keywords:
Fuzzy, Agricultural NER, Crop Recommendation, Weather Prediction, ANN

1. INTRODUCTION

Recommendation systems use a number of different technologies. These systems can be classified into two broad groups: Content-based systems and collaborative systems. Content-based systems examine properties of the items for recommendation while collaborative filtering systems recommend items based on similarity measures between users and/or items [1].

Information extraction is the task of automatically extracting structured information from unstructured and/or semi-structured machine-readable documents. In most of the cases this activity concerns processing human language texts by means of natural language processing.

In this project, we have developed a hybrid recommendation by including the advantages of content and collaborative systems to suggest a crop and farming strategy based on regional historic parameters [1]. We will be looking at the suitability of parameters, similarity of parameters with the historic weather and soil data collected using fuzzy logic. Thus the proposed crop recommender falls under the collaborative system of recommendation. On the other hand, the cost prediction, weather predictions and agricultural NER modules are content based systems. We have integrated content based and collaborative based system to form a new hybrid recommendation system, wherein the content based system comes from the consideration of weather and soil prediction systems while incorporation of cost contributes to the collaborative system of recommendation.

Given a region as input, the system should recommend the most suitable crops based on weather prediction for the region, soil parameters of the region, cost predicted for the crops to maximize yield. The system should also suggest fertilizer, pesticide, other farming strategies for the crops recommended. The crops should match the historic data and the weather, soil of the region. With increasing lack of information about agriculture among farmers’ major disasters for farmers has been on the increase [2]. This project will yield as a complete support for farmers and also small scale garden farmers for choosing crops for harvesting.

This paper is organized as follows: Section 2 discusses about the existing works in the field. Section 3 discusses about the methodology followed in our approach. Section 4 presents evaluation of the project and the results obtained. Section 5 briefs about the overall work and highlights the future scope of this project.

2. RELATED WORKS

In existing approaches, given a region and a crop, the suitability level for a crop is shown for different sub-regions within the region [1]. Many geo environmental factors like soil, climate, slope, flood and erosion hazards are considered. But it is limited to very few crops. Results on other environmental factors were not good.

Earlier, the multi-criteria land suitability was assessed more non-spatially, assuming the spatial homogeneity over the area under consideration. This, however, is unrealistic in cases like land suitability studies, where decisions are made using criteria which vary across in space (Malczewski, 2003) [3]. Non-spatial conventional Multi-criteria Decision making (MCDM) techniques average or total the impacts that are judged appropriate for the whole area under consideration. To address the spatial decision making, Multi-criteria Evaluation (MCE) and Geographic Information Systems (GIS) can be integrated [4].

The inability of the normal decision making methods to address the imprecision and the uncertainty paved the path for the fuzzy decision making techniques. There are some approaches which takes uncertainty of data into account [5] (like weather and nutrient data).

The system by Prakash T.N [5] uses Analytic Hierarchy Process (AHP), Ideal Vector Approach and Fuzzy AHP. A multi-criteria decision making technique is developed using fuzzy logic and land suitability (current suitability) is analyzed for agricultural crops. Much more factors like soil, climate, irrigation, infrastructure and socio-economic factors are considered. But limited to a very small area (594 sq. km) and restricted to a single crop (rice).

There are systems [3] that use Artificial Neural Network to predict crop yields in different climatic zones based on daily weather data and to predict crop suitable for particular soil.
It uses meteorological daily weather report. We utilized the ANN approach for periodic weather prediction for a long range of period [4].

Existing works are largely scoped on evaluating the land suitability for a specific crop. Very few works exist on crop recommendation for a very small region. All works require the farmer to extensively take soil sample tests for the land. Unlike other works we aim at providing a prediction oriented approach to recommending. Our work also takes the uncertainty in soil and weather into account by the use of fuzzy logic. This system will aim at crop recommendation from the list of 44 crops that have been considered. Further, a data set of 576 districts covering all farming districts in India is considered, which has not been done by any previous works. We aim to keep the knowledge requirement of the farmer as minimum as possible to gain benefits from the system.

3. APPROACH

The block diagram of the entire system is shown in Fig.1. To aid in recommendation, regional weather prediction and regional soil data preprocessing have been done. Weather prediction has been implemented using ANN. Fuzzy rules for crops are created from crawled data using relation extraction. A separate agricultural NER module has been developed to aid in relation extraction as existing NER tools do not work for agriculture requirements. Rules generated consist of crop name and the suitable rainfall, temperature, soil and pH. So far 44 such crop rules have been generated. Using these rules as fuzzy rules and the regional weather prediction, soil data as input a fuzzy logic based crop recommendation is implemented.

For these recommended crops farming strategy is retrieved using SVM and hand labeled evidences with crawled data as dataset using IEPY toolkit. Further fertilizer, pesticide selection modules have been added to the crops. To shortlist the crops recommended, we have implemented a cost prediction module. We have developed both ANN and regression equation approaches for cost prediction to compare the accuracy. The end result would be most suitable, sustainable crops for the given district along with farming strategies to help cultivate the crop.

In this work, the following contributions have been made:
- It is the first system that handles 44 crops and 576 districts across India.
- It is the first system that recommends crops by considering multiple features like soil, weather and cost as against a single feature.
- Cost factor is very important parameter considering the disasters farmers are facing now. This is the first system that had considered this feature for crop recommendation.
- We have developed machine learning based NER system specifically for agricultural domain that focuses on tagging our own set of 12 entities.
- A farming strategy suggestion algorithm is also designed in this project which gives pointers for seed sowing, seed spacing, irrigation, harvesting and interleaved farming, which is one of its kind.
- Algorithm for strategy data cleaning using statistical methods has been devised for removing and segregating farming strategies.
3.1 WEATHER PREDICTION

Month wise weather data from 1901-2002 is obtained from Indian Meteorological Department site [6] and is processed to remove unnecessary data. Average rainfall and precipitation is taken for each year. Ten cross fold validation method is used and 90% of data is used for training while 10% for testing. After creation of neural structure, the structure is stored in dictionary format for easy retrieval. When the region is mentioned, weather prediction is called for the region, which returns the average temperature and precipitation for next 12 months in the given region.

3.2 COST PREDICTION

Cost data for each crop from 2005-2015 is obtained from data.gov.in Government data website and is processed to remove unnecessary data. Average cost for each crop across all areas is calculated and stored for each year. Ten cross fold validation method is used and 90% of data is used for training while 10% for testing. After creation of neural structure, the structure is stored in pickle format for easy retrieval. When the crop is mentioned, cost prediction is called for the region which returns the crop cost for next year.

Linear regression is also used to predict crop cost given the same data. Cost data acquired is processed to construct the graph in linear regression with year in X-axis and price in Y-axis. The mathematical equation for existing data is calculated to predict the price for next year. Cost prediction is based on [7].

3.3 NER FOR AGRICULTURAL DOMAIN

Conventional NER system can identify person names, location, organization, date, etc. But there exists no NER system to identify crop names, their growing period and other conditions suitable for their growth which is required by our system. Hence we develop an NER system to extract this information from raw text.

Huge datasets containing crop growing conditions are obtained from National Bureau of Soil Survey and Land Use Planning (ICAR). The input is preprocessed and tokens and Part of Speech (POS) tags are generated. We use a Conditional Random Field (CRF) approach to tag the agricultural entities. CRF approach is chosen, as agricultural related words need not be in sequence. The features considered for the CRF approach are Word Feature (Names of crops), Numerical Features (Rainfall amount, soil pH), class feature (CRF uses a Begin, Intermediate and other occurrence to indicate the position of occurrence of a particular entity) and dictionary feature (soil type like red, black, crop type like kafir, rabi, etc.).

As CRF is a supervised approach, the Named Entities like crop, temperature, rainfall, etc. are annotated to form the training data. A lot of other features like gazette feature, word feature, etc. are appended to the training data. Using this training data, the input document is tagged to extract relations between the various tagged entities in the next stage.

The output of NER with entity mentioned is shown in Fig.2.

The NER tags designed and generated are given in Table.1.

<table>
<thead>
<tr>
<th>Tag Label</th>
<th>Tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>OTHERS</td>
<td>O</td>
</tr>
<tr>
<td>B-CRP, I-CRP</td>
<td>CROP</td>
</tr>
<tr>
<td>B-CLZ, I-CLZ</td>
<td>CLIMATIC ZONE</td>
</tr>
<tr>
<td>B-TMP, I-TMP</td>
<td>TEMPERATURE</td>
</tr>
<tr>
<td>B-RNF, I-RNF</td>
<td>RAINFALL</td>
</tr>
<tr>
<td>B-STX, I-STX</td>
<td>SOIL TEXTURE</td>
</tr>
<tr>
<td>B-STY, I-STY</td>
<td>SOIL TYPE</td>
</tr>
<tr>
<td>B-PH, I-PH</td>
<td>PH</td>
</tr>
<tr>
<td>B-GRW, I-GRW</td>
<td>LENGTH OF GROWING PERIOD</td>
</tr>
<tr>
<td>B-SD, I-SD</td>
<td>SOIL DEPTH</td>
</tr>
<tr>
<td>SOD</td>
<td>SODICITY</td>
</tr>
<tr>
<td>SAL</td>
<td>SALINITY</td>
</tr>
</tbody>
</table>

Fig.2. Output of NER

3.4 FUZZY RULE GENERATION

The result of NER is further processed with Relation Extraction to recognize suitable crop condition. The relation we try to identify is “suitable”. Stanford CoreNLP Relation extractor is used for this purpose. The structured data from Relation Extractor is parsed and the entities that qualify the suitable condition are extracted and stored in the form of dictionary. The input contains crop and the suitable conditions for that crop. Rules are framed using this information and given as output.

3.5 FUZZY CROP RECOMMENDATION

The module uses weather prediction data, soil dataset to know the parameters for the district given as input. We have used Gaussian membership function to fuzzify the parameters and have designed the following rules for crop recommendation.
3.5.1 Rules:

It consists of fuzzy IF-THEN rules. Let U and V be universe of discourses for antecedent and consequent of the rules, then the rule of if x is A, then y is B, where x belongs to U, and y belongs V, represents a relation between A and B, and extension to multiple rules and multiple antecedents can be easily done by specifying both composition and inference methods. The rules are formulated based on the crawled, NERed and relation extracted data for the various crops. Temperature range, Rainfall range and Soil pH range that are necessary for crop growth are used to formulate the rules. As these parameters for crop growth cannot be discrete, we claim that fuzzy approach is more apt for crop recommendation system using multiple features.

3.5.2 Inference:

After the fuzzy matching step, a fuzzy step is invoked for each of the relevant rules to produce a conclusion based on their matching degree. There are two methods: (1) the clipping method and (2) the scaling method. Both methods generate an inferred conclusion by suppressing the membership function of the consequent. The extent to which they suppress the membership function depends on the degree to which the rule is matched. Lower the matching degree, more severe the suppression of membership functions.

3.5.3 Defuzzifier:

The function of this module is to transform the fuzzy output into crisp output. Defuzzification process requires the most computational complexity in fuzzy system, and center-of-gravity or height defuzzification method is common. In this work center of gravity method is used for defuzzification. In this method, the crisp value of the output variable is computed by finding the variable value of the center of gravity of the membership function for the fuzzy value [8].

The above steps are computed for each sowing month and for the different crop growing durations possible in order to generate a range of suitable crops with corresponding growth windows. The crops generated are given as the recommended crops to the farming strategy selection module for short-listing. The modified algorithms for fuzzy crop recommendation are given below.

Pyfuzzy toolkit has been used to aid in fuzzy logic algorithms. Fuzzy crop recommendation is explained in algorithm 1 and algorithm 2.

Algorithm 1: Fuzzy Crop Recommendation

1: function FUZZY CROP RECOMMENDATION()
2:   region farmer
3:   Weather～ WEATHER PREDICT(region)
4:   Crop～ region soil dataset(region)
5:   params～ fuzzify(region, Weather)
6:   for month in 1 to 6 do
7:     for possible growing durations do
8:       max～ MAX MIN INFERENCE(plant, rules)
9:       crop～ DEFUZZIFICATION(max)
10:      RecommendedCrops.append(crop, month)
11:   end for
12: return RecommendedCrops

Algorithm 2: MaxminInference

1: function MAXMININFERENCE(params)
2:   Input～ input region parameters
3:   for all rules in IN do
4:     µ[rule]～ evaluate antecedents of rule
5:     if µ[rule] is OR then
6:       µx y～ max(µa, µb)
7:     else if µ[rule] is AND then
8:       µx y～ min(µa, µb)
9:     end if
10:    µ～ clipping top(µ)
11: return µ

3.6 SUPERVISED FARMING STRATEGY RETRIEVAL

After identifying the crop for a particular region, based on the predicted weather for the region and the soil type of the region, a crop can be grown by following a particular strategy to have a high yield. This module of our work aims at collecting a huge set of web documents and books and uses them as input to generate two datasets for agriculture. A farming strategy dataset is developed that consists of seed treatment, plant spacing, irrigation, harvesting and growth period tips to aid the farmer in growing the recommended crops.

A mixed cropping dataset that consists of strip cropping, crop rotation, intercropping and other mixed cropping possibilities for the recommended crops. It has been identified that growing mixed crops is useful, as the companion crop helps in nutrient source, pest repellent, cover crop, space usage and many other factors. In this work, we have suggested the mixed crop growing which is not done in any other work. A semi supervised active learning core is used. IEPY open source library is used for information retrieval utility. The architecture of farming strategy retrieval is given in the Fig.3.

The textual analysis phase does preprocessing steps such as POS tag, NER etc. The classifier is trained using sample evidences hand labeled from the documents. The classifier then classifies other input sample evidences to obtain text segments as result. These text segments are stored in the dictionary for later access. The dataset consists of roughly 150 documents crawled from web. It can be further extended by processing using the trained classifier.

Fig.3. Architecture Diagram of Farming Strategy Retrieval

Sentences from Crawled Documents → Textual Analysis → Classifier (SVM) → Text Segment (E1, E2, F)
3.6.1 Textual Analysis:

Text tokenization and sentence splitting, Text lemmatization, Part-Of-Speech (POS) tagging, Named Entity Recognition (NER), Syntactic parsing, Text Segments creation. In this work, we have suggested and used the following features for suggesting farming strategy which is one of its kind

3.6.2 Features used in Classifier:

- number of tokens
- symbols in between
- in same sentence
- verbs_count
- verbs_count_in_between
- total_number_of_entities
- other_entities_in_between
- entity_distance
- entity_order_bag_of_word_bigrams_in_between
- bag_of_words_in_between
- bag_of_word_bigrams
- bag_of_wordpos
- bag_of_words

The algorithm is devised using handpicked relations, their possible entities and features to detect them generated by analysis of possible farming strategies [9]. Support Vector Machine is used to classify the rest of the sentences. This is explained in algorithm 3.

Algorithm 3: Text Analysis

1: s ← tokenize(sentence)
2: n ← No.of words(s)
3: W ← Words in s
4: for i ← 1 to n do
5:   LEM [i] ← lemmatize ([W[i]])
6:   POS[i] ← POStag ([W[i]])
7: end for
8: E ← StanfordNER(s)
9: t ← StanfordParser(s)
10: Segments ← textsegmentation(s, t)
11: return Segments

Algorithm 4: Classifier

1: dataset ← WebCrawler(agridata)
2: D ← TEXT ANALYSIS(dataset)
3: r ← inputstrategy
4: (<X, Y>, r) ← training(D)
5: clf ← SVM()
6: clf.fit (<X, Y>, r)
7: for all s ∈ D do
8:   T ← clf.predict(s, features(s), (<X, Y>, r))
9:   for all x, y, segment ∈ T do
10:     if positive then
11:       result.append (x,y,segment)
12:     end if
13: end for
14: end for

3.7 STATISTICAL CLEANING

The data sets generated by farming strategy retrieval are dependent on the data crawled. Hence, it was required to develop a smart cleaning algorithm using our designed probabilistic and NLP approaches. The crop names also differed from document to document. Hence, a synonym resolution algorithm was developed to unify all the crop data in the project. Character bi-gram of each crop name is taken and is associated with the previous crops. Example: Cashew and Cashewnut are same. Character n-gram between them will match the cashew part. Only 3 letters varies and common letters are continuous, so they point to same crop. Such a strategy is maintained to have a common name.

Intersection between farming strategy dataset and mixed cropping dataset is removed by following a new text analysis algorithm. Frequency of occurrence of crops is calculated and scaled to 0 to 1 range in order to utilize for ranking later. Farming strategy data is cleaned by using keywords to rank and eliminate useless data. Keywords include some commonly identified agriculture related words like “farm”, “yield”, “land”, “harvest”, “sow”, etc., name of the months, measuring units like “kg/ha”, “acre”, “cm”, etc. Mixed crops are ranked and low frequency crops are eliminated. This is explained in algorithm 5.

Algorithm 5: Statistical Strategy Cleaning

1: DS ← CLASSIFIER()
2: mc ← ExtractMixedCropping(DS)
3: fs ← ExtractFarmingStrategy(DS)
4: mc ← ngramSimilarityElimination(mc)
5: synonym ← group similar crop name(mc)
6: keyword ← list of keywords
7: for all strategy ∈ eachcrop.strategy do
8:   strategy ← clean strategy(strategy)
9: end for
10: for all crop ∈ crops do
11:   rank ← Rank using keywords(fs)
12: end for
13: return mc, fs

3.8 FARMING STRATEGY SELECTION

In order to shortlist the crops recommended and to combine farming strategy and mixed cropping data to them this algorithm had to be developed. Weights used for short-listing are cost prediction for crop, weather and soil parameters suitability for companion crop, frequency of occurrence in farming strategy dataset generated by the system, budget deviation and sowing month. The top ranked crops are retained as recommendations. They are combined with fertilizer and pesticide selection algorithms. The farming strategy data are rearranged using keywords to give better clarity in the UI. The detailed algorithm is given below.

Algorithm 6: Farming Strategy Selection

1: r ← region farmer
2: W ← WEATHER PREDICTION()
3: Cost ← COST PREDICTION()
4: Crops ← FUZZY CROP RECOMMENDATION()
5: MC, FS ← STATISTICAL STRATEGY CLEANING()
6: for all c ∈ Crops do
7:   CompanionCrops ← MC [c]
8: \( cc \leftarrow \text{MostSuitedToRegionParameters}(r, \text{CompanionCrops}) \)
9: \( co \leftarrow \text{Cost}[r][c] \)
10: \( f_q \leftarrow \text{FrequencyIn (STATISTICAL\_STRATEGY CLEANING)} \)
11: \( mo \leftarrow 12 – \text{growing month} \)
12: \( \text{rank}[c] \leftarrow \alpha \cdot co + \beta \cdot f_q + \gamma \cdot mo \)
13: \( \text{SingleCropStrategy}[c] \leftarrow \text{FS}[c] \)
14: \( \text{end for} \)
15: \( \text{res} \leftarrow \text{return TopNRank(Crops, rank, SingleCropStrategy)} \)
16: \( \text{return res} \)

3.9 Fertilizer and Pesticide Selection

The crop nutrient dataset is obtained from fao.org. The \( N, P, K \) values present in a region is obtained from region soil data set. Using sufficiency approach the minimal nutrient required to grow that crop in the given region is calculated and given as output for different yield of crops.

Crop pesticide data is collected from books by Integrated Pest Management. The tables are parsed and the information is extracted which is stored in dictionary format to retrieve in \( O(1) \) time.

4. Result and Evaluation

The test data set consists of district wise historic crop production data for all 576 districts in India. Thus 576 yearly input test cases were considered. Each module of the system was also tested separately. The results of this module testing as well as the testing of the entire system are summarized in the following subsections [10].

4.1 Evaluation Metrics Used

4.1.1 Confusion Matrix:

- Compare with historic data of crops grown in previous years in the region
- Cost of misclassification is different if you have a lot more test data of one class than the other
- A confusion matrix (or confusion table) shows a more detailed breakdown of correct and incorrect classifications for each class.
- The rows of the matrix correspond to ground truth labels, and the columns represent the prediction.
- Precision is the fraction of events where we correctly declared ‘i’ out of all instances where the algorithm declared ‘i’.
- Conversely, recall is the fraction of events where we correctly declared ‘i’ out of all of the cases where the true state of the world is ‘i’.
- Use each crop as a class of confusion matrix.

Sample output of the system is shown in the Fig.4.

![Sample Output of the system](image-url)
A part of the confusion matrix is shown in Fig. 5. It is evident that, high values are accumulated at the diagonal. Hence the recommendation of the system and the historic data match in most cases. Very few instances have deviated from historic data since most of the other values are zeros.

4.2 RESULTS

4.2.1 Weather Prediction:

The weather prediction is evaluated using tenfold cross validation where 90% data was used for training and 10% was used for testing and root mean square of the difference between actual and predicted values were found. The Fig. 6 shows the distribution of Average temperature error over all the districts. The legends represent the range of root mean square error rate while the pies represent the percentage of district falling under each range.

In case of average temperature, the temperature steadily increases over the time and hence the root mean square error value is very low than that of precipitation. In case of precipitation, since rainfall amount won’t fully depend on previous year data, the error rate is high. Taking cloud cover, atmospheric pressure can reduce the error rate for precipitation prediction.

4.2.2 Crop Prediction:

The crops recommended have been compared with the historical production statistics based on area and productivity in the region. The weightage given to cost of the crop in the final recommendation is compared by setting it to a proportion of 1/3 and 1/2. The corresponding results obtained are described in the graph shown in Fig. 7. Here, C denotes the weightage given to cost, Y axis denotes the precision, recall, f-score and accuracy obtained for cost weightage of 1/2 and 1/3.

![Confusion Matrix for Crops](image1)

![Distribution of Error Rate over Districts](image2)

![Result obtained for different Cost Weightage](image3)
It is evident that a high weightage for cost is not very efficient, as not everyone can afford to produce those crops. Hence historical data also shows less production for those crops. Hence the overall recall is low. The recall is hugely affected by the recall of Rice crop and Sugarcane. Even in dry districts such as Ramanathapuram historic data shows highest production for rice, whereas the system does not suggest rice. This has caused a low recall of 0.08, 0.16 for rice crop, for cost weightages 1/3 and 1/2 respectively. This shows the improper cultivation in many regions. Similarly several suitable crops such as cashew nut are not grown (according to the data). Hence the validation with historic datasets affects few specific crops, which results in the overall precision and recall being affected. Results distributed over crops are shown in Fig.8. The reason for several crops not having a good precision recall is also because the support evidences are less. This shows that they may have been suppressed by other crops during the fuzzy rule inference.

The experts suggested use of sensors to accurately obtain the soil conditions in the field of farmer. Since we are looking at very wide cover of districts sensors cannot be used. Also there were suggestions to concentrate on micro nutrients in soil and further be precise about \( N, P, K \) values. Data for micro nutrients were not completely available and \( N, P, K \) values solely depends on user’s input. The expert felt that most of the crops that were recommended are proper, with very few regularly grown crops missing and very few improper ones recommended. His reviews on strategy were equally good, with suggestion to include more strategies like irrigation details based on farmer’s land information. The scores obtained in expert evaluation is given in Table.3.

![Graph showing Precision, Recall, and F1-score](image)

**Fig.8. Results distributed over crop**

Expert evaluation is also done to get an overall idea of the validity of farming strategies suggested as historic data based evaluation is not completely dependable due to error in crop choices that have prevailed in agriculture. Crop correctness and strategy correctness are two measures used to evaluate the overall output.

To evaluate the crops and different strategies, the input and output of several districts were shown to experts from agricultural domain and an overall rating scale of 0-5 was used to rank the system. The description of crop correctness scale is given in Table.2.

<table>
<thead>
<tr>
<th>Score</th>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Perfect</td>
<td>Strategies are complete and precise</td>
</tr>
<tr>
<td>4</td>
<td>Fair</td>
<td>Strategies are correct and mostly complete</td>
</tr>
<tr>
<td>3</td>
<td>Acceptable</td>
<td>Strategies are correct but not complete</td>
</tr>
<tr>
<td>2</td>
<td>Bad</td>
<td>Few strategies are wrong</td>
</tr>
<tr>
<td>1</td>
<td>Poor</td>
<td>Most strategies are wrong</td>
</tr>
<tr>
<td>0</td>
<td>Nonsense</td>
<td>Strategies make no sense</td>
</tr>
</tbody>
</table>

**Table.2. Scale to Evaluate Strategy Correctness**

<table>
<thead>
<tr>
<th>Crop Correctness Score</th>
<th>Strategy Correctness Score</th>
<th>Correct Strategy %</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.0</td>
<td>3.0</td>
<td>77.77</td>
</tr>
</tbody>
</table>

**5. CONCLUSION AND FUTURE WORKS**

This is a crop recommendation (hybrid) system which utilizes fuzzy logic to choose from 44 crop rules. The large number of crop rules was possible, by devising a crop fuzzy rules generation algorithm that has been designed using NER and relation extraction. As standard NERs do not support agricultural usage an agricultural NER system has been proposed as part of the system. Further to aid in obtaining the weather parameters of the users region, the standard ANN method is used to predict weather. To aid in short-listing the crops recommended by fuzzy logic, cost prediction of crop has been done. Here, a comparison between ANN based and regression equation based approach was made to choose the best approach to predict cost. Along with the recommended crops, farming strategy suggestion and fertilizer, pesticide selection algorithms have been newly introduced by us. Farming strategy selection is an extensive information extraction algorithm which was presented in sections 3.6, 3.7 and 3.8. The results of performance evaluation show an F-score of 54%.

This research has a lot of scope for further developments. The efficiency of agriculture NER can be improved using more training data and rules, which will allow it to be utilized in farming strategy retrieval algorithm also, which will highly reduce even the need for statistical cleaning. Fuzzy rules can be extended to consider previous soil utility, and soil texture using remote sensing on agricultural land, which will increase the precision. More agricultural parameters can be identified to be included in the system either in fuzzy logic or as a separate module. The cross sectional and top view images of soil can be processed to get a better idea about the soil texture. The system can also be integrated with sensors which will give daily report of soil and weather to aid in strategy suggestion. As a budding domain there are still many requirements in agriculture that have not been explored.

**REFERENCES**


