ISSN: 2229-6956 (ONLINE)
DOI: 10.21917/ijsc.2016.0173

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

Aadithya U[1], Anushya S[2], Bala Lakshmi N[3] and Rajeswari Sridhar[4]

Department of Computer Science and Engineering, Anna University, Chennai, India

E-mail: [1]aadi1194@gmail.com,[2]anushya1995cool@gmail.com,[3]bala94.n@gmail.com and [4]rajisridhar@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.

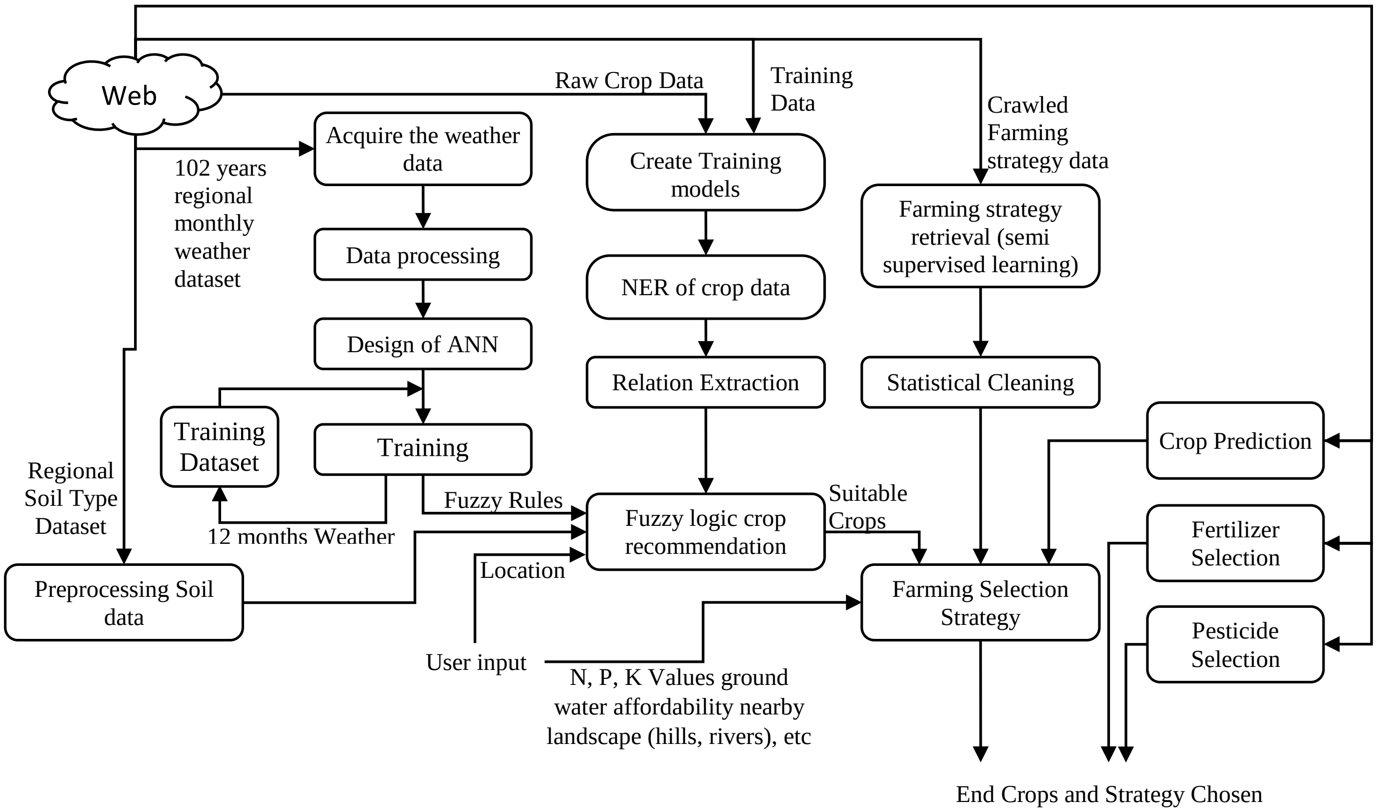

Fig.1. Block diagram of overall system

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 faming, 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].

O	O	2709	2711	pH	pH	NN	B-NP	O	O
B-PH	B-PH	2712	2713	4	4	CD	I-NP	O	O
I-PH	I-PH	2713	2714	.	.	.	O	O	O
I-PH	I-PH	2714	2715	5	5	CD	B-NP	O	O
O	O	2716	2718	to	to	TO	I-NP	O	O
B-PH	B-PH	2719	2720	8	8	CD	I-NP	O	O
I-PH	I-PH	2720	2721	.	.	.	I-NP	O	O
I-PH	I-PH	2721	2722	0	0	CD	I-NP	O	O
O	O	2722	2723	)	)	)	O	O	O
O	O	2725	2728	The	The	DT	B-NP	O	O
O	O	2729	2734	soils	soil	NNS	I-NP	O	O
O	O	2735	2739	most	most	RBS	B-VP	O	O
O	O	2740	2746	suited	suit	VBN	I-VP	O	O
O	O	2747	2749	to	to	TO	B-PP	O	O
O	O	2750	2761	cultivation		cultivation		NN	B-NP
O	O	2762	2764	of	of	IN	B-PP	O	O
O	O	2765	2768	the	the	DT	B-NP	O	O
O	O	2769	2773	crop	crop	NN	I-NP	O	O
O	O	2774	2777	are	be	VBP	B-VP	O	O
O	O	2778	2783	heavy	heavy	JJ	B-NP	O	O
O	O	2784	2789	soils	soil	NNS	I-NP	O	O
O	O	2790	2791	(	(	(	O	O	O
B-STX	B-STX	2791	2795	clay	clay	NN	B-NP	O	O
O	O	2796	2798	or	or	CC	O	O	O
B-STX	B-STX	2799	2803	clay	clay	NN	B-NP	O	O
I-STX	I-STX	2804	2808	loam	loam	NN	I-NP	O	O
O	O	2809	2812	and	and	CC	O	O	O
B-STX	B-STX	2813	2817	loam	loam	NN	B-NP	O	O
O	O	2818	2823	soils	soil	NNS	I-NP	O	O
O	O	2823	2824	)	)	)	O	O	O
O	O	2826	2829	The	The	DT	B-NP	O	O
O	O	2830	2835	broad	broad	JJ	I-NP	O	O
O	O	2836	2840	soil	soil	NN	I-NP	O	O
O	O	2841	2846	types	type	NNS	I-NP	O	O
O	O	2847	2852	under	under	IN	B-PP	O	O
B-CRP	O	2853	2857	rice	rice	NN	B-NP	O	B-CRP
O	O	2858	2869	cultivation		cultivation		NN	I-NP

Fig.2. Output of NER

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 karif, 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.1. Entity Classes

Tag	Label
O	OTHERS
B-CRP, I-CRP	CROP
B-CLZ, I-CLZ	CLIMATIC ZONE
B-TMP, I-TMP	TEMPERATURE
B-RNF, I-RNF	RAINFALL
B-STX, I-STX	SOIL TEXTURE
B-STY, I-STY	SOIL TYPE
B-PH, I-PH	PH
B-GRW, I-GRW	LENGTH OF GROWING PERIOD
B-SD, I-SD	SOIL DEPTH
SOD	SODICITY
SAL	SALINITY

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

r ← region farmer
rules ← FUZZY RULE GENERATION()
W ← WEATHER PREDICT(r)
rs ← region soil dataset(r)
params ← fuzzify($rs[r]$, $W[r]$)
for month ← 1 to 6 do
  for all possible crop growing duration windows, x do
    res ← MAX MIN INFERENCE(params, rules)
    crop ← DEFUZZIFICATION (*res*)
  end for
  RecommendedCrops.append(*crop*, *month*)
end for
return RecommendedCrops

Algorithm 2: MaxminInference

IN ← input_region_parameters
for all in ∈ IN do
  for all s ∈ fuzzy sets do
    $\mu[in][s]$ ← y axis readings of in, s intersection
  end for
end for
for all r ∈ fuzzy rules do
  $\mu[r]$ ← evaluate antecedents of r
  if *op* is OR then
    $\mu_a S b (x) \leftarrow \max(\mu_a, \mu_b)$
  end if
  if *op* is AND then
    $\mu_a T b (x) \leftarrow \min(\mu_a, \mu_b)$
  end if
  $\mu[r]$ ← clipping top($\mu[r]$)
end for
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.

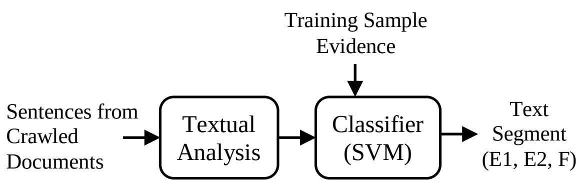

Fig.3. Architecture Diagram of Farming Strategy Retrieval

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.

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_wordpos_bigrams
bag_of_wordpos
bag_of_word_bigrams
bag_of_pos
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 sentences as positive and negative. Part of the input documents is used as training dataset. This is used by SVM to classify the rest of the sentences. This is explained in algorithm 3 and 4.

Algorithm 3: Text Analysis

```
s ← tokenize(sentence)
n ← No.of words(s)
W ← Words in s
for i ← 1 to n do
    LEM [i] ← lemmatize (W[i])
    POS[i] ← POStag (W[i])
end for
E ← StanfordNER(s)
t ← StanfordParser(s)
Segments ← textsegmentation(s, t)
return Segments
```

Algorithm 4: Classifier

```
dataset ← WebCrawler(agridata)
D ← TEXT ANALYSIS(dataset)
r ← inputstrategy
(< X, Y >, r) ← training(D)
clf ← SVM()
clf.fit (< X, Y>,r)
for all s ∈ D do
    T ← clf.predict(s, features(s), (< X, Y>, r))
    for all< x, y, segment > ∈ T do
        if positive then
            result.append (x,y,segment)
        end if
    end for
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

```
DS ← CLASSIFIER()
mc ← ExtractMixedCropping(DS)
fs ← ExtractFarmingStrategy(DS)
mc ← ngramSimilarityElimination(mc)
synonym ← group similar crop name(mc)
keyword ← list of keywords
for all strategy ∈ eachcrop.strategy do
    strategy ← clean strategy(strategy)
end for
for all crop ∈ crops do
    rank ← Rank using keywords(fs)
end for
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

```
r ← region farmer
W ← WEATHER PREDICTION ()
Cost ← COST PREDICTION ()
Crops ← FUZZY CROP RECOMMENDATION ()
MC, FS ← STATISTICAL STRATEGY CLEANING()
for all c ∈ Crops do
    CompanionCrops ← MC [c]
```

```
cc ← MostSuitedToRegionParameters(r, CompanionCrops)
  co ← Cost[r][c]
  fq ← FrequencyIn (STATISTICAL_STRATEGY
  CLEANING)
   mo ← 12 − growing month
     rank[c] ← α.co + β.f q + γ.mo
     SingleCropStrategy[c] ← FS[c]
end for
res ← return TopNRank(Crops, rank, SingleCropStrategy)
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 sub-sections [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 of 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.


Cashewnut - Cashew

Mixed cropping

- Horse gram
- Coconut
- Groundnut
- Beans

Seed treatment

- : - Cashew seedlings are grown under shade -LRB- 45 % -RRB- and hardened off before planting in the orchard . It is very important

Spacing

- Farming : - Planting space of 8 m x 5 m is recommended for cashew farming . The trees grow vigorously
- in the first 3 years and as soon as the crowns touch each other alternate cashew trees should be removed until the permanent planting distance of 10 to 12 m is reached . Branches hanging on the

Irrigation

- : - Fresh seeds that sink in water are planted in an upright position in a planting bag containing a loose , sterilised soil mixture . 3 to 4 seeds
- Farming : - In Cashew Farming , irrigation is an important factor during establishment of young trees because it doubles the growth tempo of young trees in a dry season . Due to the deep

Growth period

- with a well-defined dry season of at least four months to produce the best yields . Coincidence of excessive rainfall
- Cashew Farming : - In Cashew Farming , during first year of planting , the sprouts coming from the rootstock should be removed frequently to ensure better health of the plant . These sprouts eat up
- by dusting with 2 % sulphur W.P. Flowering to Harvesting time of Cashew : - In Cashew Farming , flowering is affected by weather conditions and also varies from tree to tree , but continues for a period of 3 months . High tempera
- as soon as possible , especially under wet conditions and should be dried before storage Harvesting , Yield and Marketing of Cashew : - In Cashew Farming , cashew plants start bearing after three years of planting and reach full
remunerative yields for another 20 years . The cashew nuts are
- pruning of cashew plants during first 3-4 years is essential for providing proper shape to the trees . The trees are shaped
- is restricted to altitudes upto 700 m above mean sea level where the temperature does not fall below 20 C for prolonged period . Areas where the temperatures
- fertilizer doses for cashew Age Urea -LRB- g -RRB- SSP -LRB- g -RRB- MOP -LRB- g -RRB- 1 st Year 375 275 75 2 nd year 750\xa0525\xa0150 3 rd year 1100\xa0750\xa0200 The ideal time for application of fertilizer is immediat
Fertilisers should be applied
- removed and the nuts are dried in sun for 2-3 days to bring the moisture level from 25 per cent to 9 per cent . The maturity of the

Harvesting

- mosquito , flower thrips , stem and root borer and fruit and nut borer are the major pests , which are reported to cause around 30 % loss in yield of cashew nut . Basically cashew crop does
- harvested during February -- May . Normally , harvesting consists

Pesticide

Fertilizer

- Suggestion1: 21.59kg/ha of Nitrate; 134.99kg/ha of P2O5; 16.18kg/ha of K2O;

Fig.4. Sample Output of the system

Crop	Arecanut	Arhar	Bajra	Banana	Barley	Castor seeds	Chilli	Coconut	Cotton	Green gram	Groundnut	Horse gram	Jowar	Jute	Maize	Mustard	Onion	Ragi	Rice	Sesame
Arecanut	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Arhar	2	13	0	0	0	2	0	1	1	0	1	1	1	1	0	1	0	1	0	0
Bajra	0	0	21	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0
Banana	0	0	0	8	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Barley	0	0	0	0	7	1	1	1	0	0	0	0	0	0	0	0	0	2	0	0
Castor seeds	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Chilli	2	0	0	0	1	0	18	0	0	0	0	0	0	0	0	1	0	0	0	0
Coconut	0	0	0	0	0	0	0	5	0	0	0	0	0	1	0	0	0	0	0	0
Cotton	0	0	0	0	0	0	1	1	19	0	0	0	0	1	0	0	1	0	0	0
Green gram	1	0	1	1	0	0	1	1	0	28	0	1	2	2	0	0	0	2	0	0
Groundnut	1	0	0	0	0	1	0	0	0	0	25	0	0	1	1	1	0	1	0	0
Horse gram	1	0	0	1	0	2	0	0	0	0	0	17	0	2	0	0	0	0	0	0
Jowar	1	0	0	0	0	0	0	0	0	0	0	0	40	1	0	0	0	2	0	0
Jute	0	0	0	0	0	0	0	2	0	0	0	0	0	12	0	0	0	0	0	0
Maize	1	0	0	0	0	0	0	3	0	0	0	0	0	0	59	0	0	0	0	0
Mustard	6	0	0	0	0	0	0	4	0	0	1	0	0	0	0	49	0	2	0	0

Fig.5. Confusion matrix for crops

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.

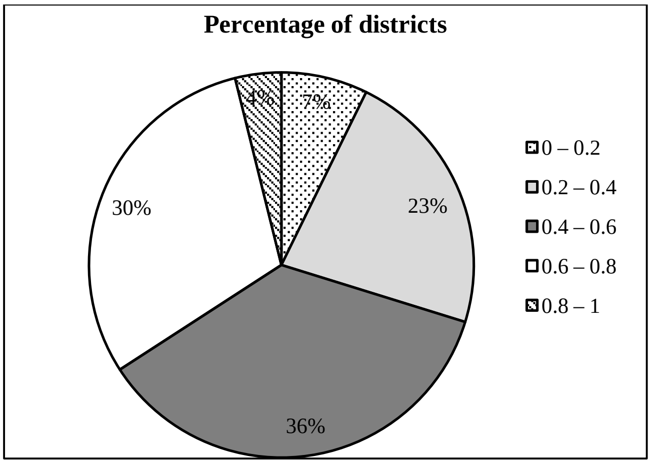

Fig.6. Distribution of Error Rate over Districts

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.

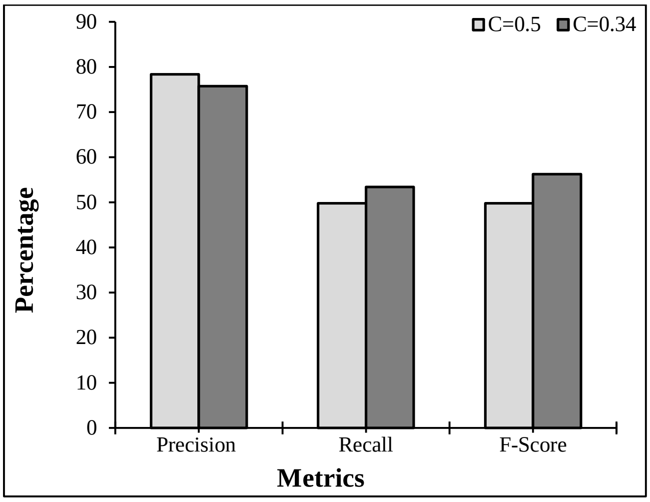

Fig.7. Result obtained for different Cost Weightage

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.

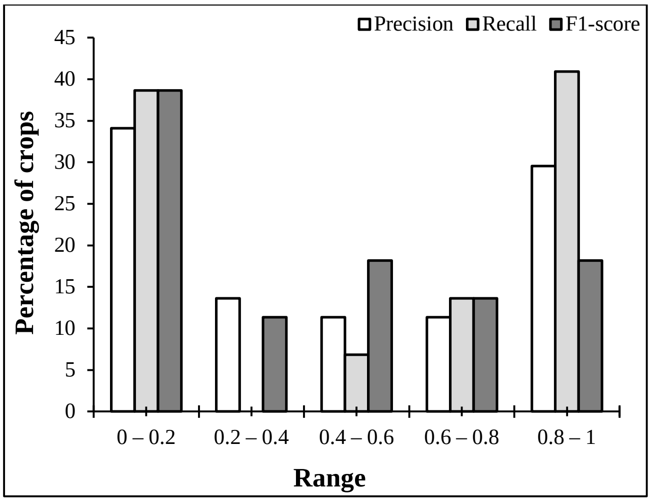

Fig.8. Results distributed over crop

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.2. Scale to Evaluate Strategy Correctness

Score	Level	Description
5	Perfect	Strategies are complete and precise
4	Fair	Strategies are correct and mostly complete
3	Acceptable	Strategies are correct but not complete
2	Bad	Few strategies are wrong
1	Poor	Most strategies are wrong
0	Nonsense	Strategies make no sense

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 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.

Table.3. Results obtained in Expert Evaluation

Crop Correctness Score	Strategy Correctness Score	Correct Strategy %
3.0	3.0	77.77

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

[1] Blaz Bahar, "A Comparison of Different Types of Recommender Systems", EnggD Thesis, Faculty of

Computer and Information Science, University of Ljubljana, 2012.

[2] "Agriculture in India", Available at: https://en.wikipedia.org/wiki/Agriculture in India, 2015.

[3] Jacek Malczewski, "GIS-Based Land Use Suitability Analysis: A Critical Overview", *Progress in Planning*, Vol. 62, No. 1, pp. 3-65, 2004.

[4] Piotr Jankowski, "Integrating Geographical Information Systems and Multiple Criteria Decision Making Methods", *International Journal of Geographic Information System*, Vol. 21, No. 3, pp. 251-273, 1995

[5] T.N. Prakash, "Land Suitability Analysis for Agricultural Crops: A Fuzzy Multi-Criteria Decision Making Approach", Ph.D Dissertation, Department of Geoinformatics, International Institute for Geo Information Science and Earth Observation, 2003.

[6] India Water Portal, "Meteorological Datasets: Download Entire Datasets for various Meteorological Indicators from 1901 to 2002", Available at: http://www.indiawaterportal.org/metdata/.

[7] Mayankkumar B. Patel and Sunil R. Yalamalle, "Stock Price Prediction using Artificial Neural Network", *International Journal of Innovative Research in Science, Engineering and Technology*, Vol. 3, No. 6, pp. 13755-13762, 2014.

[8] V.R. Thakare and H.M. Baradkar, "Fuzzy System for Maximum Yield from Crops", Proceedings of National Level Technical Conference, pp. 4-9, 2013.

[9] Nguyen Bach and Sameer Badaskar., "A Review of Relation Extraction", Literature review for Language and Statistics II, 2007.

[10] National Informatics Center, "District-wise, season-wise crop production statistics from 1998", Available at: District-wise, season-wise crop production statistics from 1998, 2015.

[11] Naoaki Okazaki, "Crfsuite: A Fast Implementation of Conditional Random Fields (CRFs)", Available at: http://www.chokkan.org/software/crfsuite/.