Effects of Shannon entropy and J-Measurement for making rules by using decision list method

Debaprasad Misra, Arindam Giri

Abstract- Expert system makes lot of changes in our daily life. For making an expert system we need perfect, efficient and concise knowledge base system (KBS). The backbone of any KBS is the finest, optimum and exact 'rules' for any particular application that makes the success of the expert system. In this paper, we generate a few rules that come from a Soybean data set with different effect of Shannon entropy and J measurement by using as the rule evaluation parameter. The dissimilar out comes makes differentiate of the effect by using rule evaluation parameter. The experimental results are also focused the different error rate from Shannon entropy and J measurement and other effects that make the changes in the output. After that, we compare and evaluate the results and outputs from two cases, that represent in graph based layout and tabular representation. More over, a short review of KBS, entropy, decision list are also paying attention in this paper.

Keywords - Decision List, Entropy, KBS, Rule extraction, Rule evaluation, Shannon entropy, J-Measurement

1. INTRODUCTON

The ability of an expert system is to perform the particular task without directly interact with internal operations and functionalities. The knowledge base system (KBS) can be defined as a system that draws upon the knowledge of human expert captured in a knowledge base to solve problem that normally required human expertise. The key components of the knowledge base system are Rule, Objects, Attributes, Relationship, Definition, Events, Process, Facts, Hypothesis, Heuristics and least time of execution. Decision list in data mining areas focus on the order list of conjunctive rules to balance the overlap of the order of rules structure and ranking the rules. Rule evaluation is one of the major tasks in data mining areas. It is basically combinational approach of robustness, comprehensive, error vigilance and conciseness. Rule extraction is another method and process in data mining for the application of our real world. The KBS depends on the efficient, genuine and exact rules that come from particular data set. The main target is that to trained network to the corresponding rules. Entropy is appropriately associated with the lack of information, uncertainty and identification. The rules and the KBS has lots of different due to the effect of Shannon entropy and J-measurement that are passing as rule evaluation parameter.

In this paper we take Soybean data set which has more than 650 examples and 35 attributes. We apply Shannon and J-measurement passing as rule evaluation parameter. The out put make differentiate between the rules that generated from the Soybean data set. The rule helps to make the Knowledge Base System (KBS) for the data set. The input parameter for both cases (Shannon entropy and J-measurement) are effective and valuable in the whole process. The classifier performances are also jugged through the outputs.

2. KNOWLEDGE BASE SYSTEM (KBS)

We can define the Knowledge Base System (KBS) as a system that draws upon the knowledge of human experts captured in a knowledge-base to solve problems that normally require human expertise. A computerized system that uses knowledge about some domain to arrive at a solution to a problem from that domain. This solution is essentially the same as that concluded by a person knowledgeable about the domain of the problem when confronted with the same problem. *(By Gonzalez and Dankel)*

The main features of KBS are

i. Heuristic rather than algorithmic
ii. General vs. domain-specific
iii. Highly specific domain knowledge
iv. Knowledge is separated from how it is used

It is combinational approach of knowledge base and inference engine. The main key components are diagrammatically given below

• Debaprasad Misra, M.Tech Final Year Student,
Haldia Institute of Technology, W.B.U.T
Department of Computer Science and Engineering,
Haldia, West Bengal, India
dpmdeb.asn@gmail.com

• Arindam Giri, Assistant Professor
Haldia Institute of Technology, W.B.U.T
Department of Computer Science and Engineering,
Haldia, West Bengal, India
ari_giri111@rediffmail.com

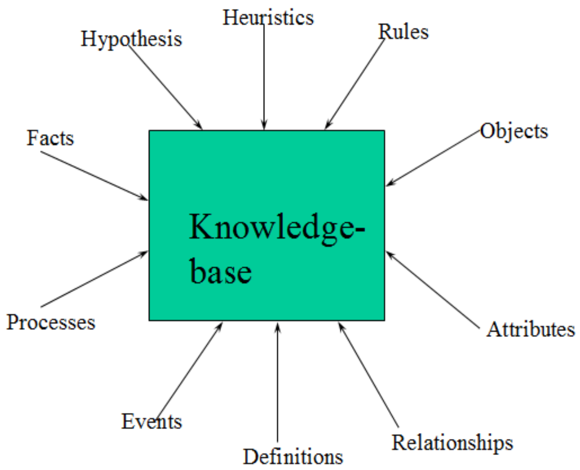

Fig 1:- Components of knowledgebase

The main tasks performed by KBS are given below

i) Diagnosis - To identify a problem given a set of indication or malfunctions.
e.g. diagnose reasons for engine failure

ii) Interpretation - To provide an understanding of a situation from available information. e.g. DENDRAL

iii) Prediction - To predict a future state from a set of data or observations. e.g. Drilling Advisor, PLANT

iv) Design - To develop configurations that satisfy constraints of a design problem. e.g. XCON

v) Planning - Both short term & long term in areas like project management, product development or financial planning. e.g. HRM

vi) Monitoring - To check performance & flag exceptions.
e.g., KBS monitors radar data and estimates the position of the space shuttle.

vii) Control - To collect and evaluate evidence and form opinions on that evidence.
e.g. control patient's treatment

viii) Instruction - To train students and correct their performance. e.g. give medical students experience diagnosing illness

ix) Debugging - To identify and prescribe remedies for malfunctions.
e.g. identify errors in an automated teller machine network and ways to correct the errors

Main advantages of KBS included are

i. Increase availability of expert knowledge
ii. Expertise not accessible
iii. Training future experts
iv. Efficient and cost effective
v. Consistency of answers
vi. Explanation of solution
vii. Deal with uncertainty
viii. Wide distribution of scarce expertise
ix. Ease of modification
x. Consistency of answers
xi. Perpetual accessibility
xii. Preservation of expertise
xiii. Solution of Problem involving incomplete data
xiv. Explanation of solution

3. DECISION LIST

Decision lists correspondent to simple case statements. Classifier consists of a series of tests to be applied to each input example/vector which returns a word sense, persist only until the first applicable test satisfied. Default test returns the majority sense Decision lists are a representation for Boolean functions Single term decision lists are more expressive than disjunctions and conjunctions; however 1-term decision lists are less expressive than the general disjunctive normal form and the conjunctive normal form

A decision list (DL) of length r is of the form,

if f_1 **then** output b_1
else if f_2 **then** output b_2
...
else if f_r **then** output b_r

where f_i is the ith formula and b_i is the ith boolean for $i \in \{1...r\}$. The last if-then-else is the default case, which means formula f_r is always equal to true. A k-DL is a decision list where all of formulas have at most k terms. Sometimes "decision list" is used to refer to a 1-DL, where all of the formulas are either a variable or its negation. [Wikipedia]

A simple DL:-
If X1=v11 && X2=v21 then c1
If X2=v21 && X3=v34 then c2

Term: conjunction ("and") of literals Clause: disjunction ("or") of literals.
CNF (conjunctive normal form): the conjunction of clauses.
DNF (disjunctive normal form): the disjunction of terms.
A decision list is a list of pairs
(f1, v1), …, (fr, vr),
fi are terms, and fr=true.

Building DL

i. For a de-accented form w, find all possible accented forms
ii. Collect training contexts
iii. collect k words on each side of w
iv. strip the accents from the data
v. Measure collocation distributions
vi. use pre-defined attribute combination:
   Ex: "-1 w", "+1w, +2w"
vii. Rank decision rules by log-likelihood
viii. Optional pruning and interpolation

We can summarized the decision list as, rules are easily understood by humans (but remember the order factor). DL tends to be relatively small, and fast and easy to apply in practice. DL is related to DT, CNF, DNF, and TBL. For learning greedy algorithm and other improved algorithms.

ISSN 2229-5518

Extension: probabilistic DL
Ex: if A & B then (c1, 0.8) (c2, 0.2) (by Fei Xia , 2006)

4. ENTROPY

Entropy is a compute of disorder or mess, or more precisely unpredictability. The use of probabilities to describe a situation implies some uncertainty. If we toss a fair coin, we don't know what the outcome will be. We can, however, describe the situation with a probability distribution: { fPr(Coin = Heads) =1=2; Pr(Coin = Tails) = 1=2}. If the coin is biased, there is a different distribution {fPr (BiasedCoin = Heads) = 0:9;Pr(BiasedCoin = Tails) =0:1}

It is important to realize the difference between the entropy of a set of promising outcomes, and the entropy of a particular outcome. A single toss of a fair coin has entropy of one bit, but a particular result (e.g. "heads") has zero entropy, since it is entirely "predictable".

Named after Boltzmann's H-theorem, Shannon denoted the entropy H of a discrete random variable X with possible values $\{x_1, ..., x_n\}$ as,

$$H[X] = E(I(X)) \quad (1)$$

Here E is the expected value, and I is the information content of X.

$I(X)$ is itself a random variable. If p denotes the probability mass function of X then the entropy can explicitly be written as

$$H(X) = \sum_{i=1}^{n} p(x_i)\, I(x_i) = \sum_{i=1}^{n} p(x_i) \log_b \frac{1}{p(x_i)} = -\sum_{i=1}^{n} p(x_i) \log_b p(x_i), \quad (2)$$

where b is the base of the logarithm used. Common values of b are 2, Euler's number e, and 10, and the unit of entropy is bit for $b = 2$, nat for $b = e$, and dit (or digit) for $b = 10$.

We define the Shannon entropy of a random variable X by

$$H[X] \equiv -\sum_{x \in \mathcal{X}} \Pr(x) \log_2(\Pr(x)) \ . \quad (3)$$

5. RULE EXTRACTION AND RULE EVALUATION

Due to some important characteristic of data mining such as fast, robustness, powerful, independence of prior assumption, classification that are making the wide range of role in the real world. The knowledge discovery in database (KDD) with the help of some rule evaluation and extracting method performed in the data mining areas successfully. The basic steps are

1) Data preparation (It includes data cleaning, data options, data preprocessing and data expression)
2) Rule extracting (by using algorithm and training).
3) Rule evaluation.
4) Gaining knowledge and
5) Make KBS (by using some specific tools).

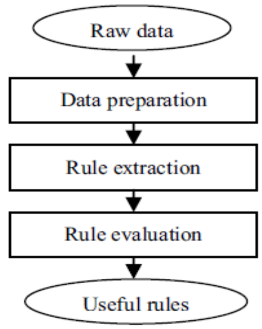

Fig 2:- Data to rules

Rule extraction method focused on various key factors that may causes effects to rule considering the all parameters. For making KBS or an expert system the exact and perfect rules are very essential. In data mining there are many extracting methods are available for clutch the rules from particular data set. The main three broad techniques are

i) Rule extraction based on performance analysis.
ii) Rule extraction based on configuration analysis
iii) Rule extraction based on target system.

Rule evaluation technique is mainly depends on the desire system that we wish to make, but the key factor of the rule evaluation is:-

i) Error performance and vigilance
ii) Strength and robustness
iii) Comprehensive and broad
iv) Input-output parameters
v) Related to specific application

The rules can be evaluated according to the following goals.

i) Find the best sequence of rules
ii) Test the accuracy of the rules
iii) Detect how much knowledge in the network is not extracted.

So these are the main key points that rule extracting and rule evaluation that has performed their individual's tasks in data mining for making expert system or KBS. These are some broad step that we discussed here but in depth study in rule extraction and evaluation method has lots of other interrelated steps and tasks that meet the success of the two methods. Due to huge amount of data across the globe now a days increasing rapidly so we have to very careful and concuss for have chosen the data set specifically the data, for avoid producing wrong rules that incorporate with the expert system or KBS.

Table 1:- Few examples of attributes in Soybean data set

Attribute	Category	Information
Month	Discrete	8 values
Plant-stand	Discrete	3 values
Precip	Discrete	4 values
Temp	Discrete	4 values
Hail	Discrete	3 values
Crop-hist	Discrete	5 values
Leaves	Discrete	2 values
Leaf spot info	Discrete	4 values
Leaf-shared	Discrete	3 values
Class	Discrete	19 values

6. Experimental Results

We take Soybean data set which has more than 650 entries and 35 attributes. Few of attributes are given below

We choose and apply Shannon entropy and J-Measurement respectively in the data set. The different out put that we got in two cases are discussed below:-

Case 1:- Shannon entropy

First we choose Shannon entropy and apply it by means of decision list method in the data set. Here we use, significance level of pruning=1.00000, rule evaluation parameter as Shannon entropy and minimum no of support of rules =12.

After applying this we got 39 rules that has generated from the soybean data set. Few rules are like this

IF crop-hist in [?]
class in [2-4-d-injury]
ELSE IF temp in [?]
class in [cyst-nematode]
ELSE IF leaf-mild in [upper-surf]
class in [powdery-mildew]
ELSE IF leaf-mild in [lower-surf]
class in [downy-mildew]
ELSE IF external-decay in [watery]
class in [phytophthora-rot]
ELSE IF int-discolor in [black]
class in [charcoal-rot]
ELSE IF int-discolor in [brown]
class in [brown-stem-rot]
ELSE IF fruit-pods in [?]
class in [phytophthora-rot]
ELSE IF leaves in [abnorm]
class in [phyllosticta-leaf-spot)
ELSE IF fruit-pods in [diseased] -- leaf spots-halo in [absent]
class in [anthracnose]
ELSE IF fruit-pods in [diseased] -- crop-hist in [same-lst-two-yrs]
class in [frog-eye-leaf-spot]
ELSE IF fruit-pods in [diseased] -- crop-hist in [same-lst-sev-yrs]
class in [frog-eye-leaf-spot]
ELSE IF fruit-pods in [diseased] -- germination in [90-100]
class in [frog-eye-leaf-spot]
ELSE IF canker-lesion in [dk-brown-blk] -- plant-stand in [lt-normal]
class in [phytophthora-rot]
ELSE IF fruiting-bodies in [present] -- plant-stand in [lt-normal]
class in [brown-spot]
ELSE IF leafspots-halo in [yellow-halos] -- precip in [norm] -- seed-tmt in [none]
class in [bacterial-blight]
ELSE IF fruiting-bodies in [present] -- leafspots-halo in [absent]
class in [diaporthe-stem-canker]

ELSE (DEFAULT RULE)
class in [brown-spot]

So, these are the few rules that have come from the soybean data set after applying the Shannon entropy, the error rate for this method we get that is, 0.1083. The values prediction table is given below,

Table 2 Values prediction for Shannon entropy

Value	Recall	1-Precision
diaporthe-stem-canker	1.0000	0.0000
charcoal-rot	1.0000	0.0000
rhizoctonia-root-rot	0.95000	0.0000
phytophthora-rot	1.0000	0.0000
brown-stem-rot	1.0000	0.0000
powdery-mildew	1.0000	0.0000
downy-mildew	1.0000	0.0000
brown-spot	0.8587	0.1413
bacterial-blight	0.9500	0.2083
bacterial-pustule	0.8000	0.2727
purple-seed-stain	1.0000	0.0000
anthracnose	0.8864	0.0000
phyllosticta-leaf-spot	0.3000	0.6667
alternarialeaf-spot	0.8901	0.2703
frog-eye-leaf-spot	0.7473	0.0811
diaporthe-pod-&-stem-blig	1.0000	0.0000
cyst-nematode	1.0000	0.0000
2-4-d-injury	1.0000	0.0000
herbicide-injury	0.6250	0.2857

Case 2:- J-Measurement

Now we choose J-Measurement and apply it by means of decision list method in the data set. Here we use significance level of pruning=1.00000, rule evaluation parameter as J-Measurement and minimum no of support of rules =12.
After applying this we got 5 rules that has generated from the soybean data set. The rules are like this

IF canker-lesion in [dk-brown-blk]
    class in [phytophthora-rot]
ELSE IF leafspots-halo in [absent]
    class in [brown-stem-rot]
ELSE IF int-discolor in [none] -- fruit-pods in [norm] -- lodging in [yes] -- leaf-mild in [absent]
    class in [alternarialeaf-spot]
ELSE IF seed in [abnorm]
    class in [downy-mildew]
ELSE IF leaves in [abnorm]
    class in [frog-eye-leaf-spot]
ELSE (DEFAULT RULE)
    class in [diaporthe-pod-&-stem-blig]

The error rate for J-Measurement is 0.6369. The value prediction is like this

Table 3 Values prediction for J-Measurement

Value	Recall	1-Precision
diaporthe-stem-canker	0.0000	1.0000
charcoal-rot	0.0000	1.0000
rhizoctonia-root-rot	0.0000	1.0000
phytophthora-rot	1.0000	0.5028
brown-stem-rot	0.7955	0.7727
powdery-mildew	0.0000	1.0000
downy-mildew	1.0000	0.6429
brown-spot	0.0000	1.0000
bacterial-blight	0.0000	1.0000
bacterial-pustule	0.0000	1.0000
purple-seed-stain	0.0000	1.0000
Anthracnose	0.0000	1.0000
phyllosticta-leaf-spot	0.0000	1.0000
alternarialeaf-spot	1.0000	0.6592
frog-eye-leaf-spot	0.1209	0.5769
diaporthe-pod-&-stem-blig	0.2000	0.0000
cyst-nematode	0.0000	1.0000
2-4-d-injury	0.0000	1.0000
herbicide-injury	0.0000	1.0000

So these are few out put that we got after applying the two methods i,e Shannon entropy and J-Measurement in the data set. The no of by are 39 and 5 by case 1 and case 2 respectively and error rate for shows the differentiates of the approaches and performances.

7. DICUSSION

In the previous section we can see the different out put getting from the two rule evaluation parameter that is Shannon and J-Measurement. Although the input parameter like level of pruning, minimum no of supported rule etc are some but due to operation and characteristic of the two parameters makes lots of different outcome for their individual's performances in the same data set. The different values that we got for the two cases for values prediction make another evaluation that represented in following table

Value	Recall		1-Precision	
	Shannon entropy	J-Measurement	Shannon entropy	J-Measurement
diaporthe-stem-canker	1.0000	0.0000	0.0000	1.0000
charcoal-rot	1.0000	0.0000	0.0000	1.0000
rhizoctonia-root-rot	0.95000	0.0000	0.0000	1.0000
phytophthora-rot	1.0000	1.0000	0.0000	0.5028
brown-stem-rot	1.0000	0.7955	0.0000	0.7727
powdery-mildew	1.0000	0.0000	0.0000	1.0000
downy-mildew	1.0000	1.0000	0.0000	0.6429
brown-spot	0.8587	0.0000	0.1413	1.0000
bacterial-blight	0.9500	0.0000	0.2083	1.0000
bacterial-pustule	0.8000	0.0000	0.2727	1.0000
purple-seed-stain	1.0000	0.0000	0.0000	1.0000
anthracnose	0.8864	0.0000	0.0000	1.0000
phyllosticta-leaf-spot	0.3000	0.0000	0.6667	1.0000
alternarialeaf-spot	0.8901	1.0000	0.2703	0.6592
frog-eye-leaf-spot	0.7473	0.1209	0.0811	0.5769
diaporthe-pod-&-stem-blig	1.0000	0.2000	0.0000	0.0000
cyst-nematode	1.0000	0.0000	0.0000	1.0000
2-4-d-injury	1.0000	0.0000	0.0000	1.0000
herbicide-injury	0.6250	0.0000	0.2857	1.0000

Table 4 Differentiate value prediction for Shannon and J-Measurement

The execution time also differ from one to another, Shannon method takes much times to execute than J-Measure method but in the context of generating rule Shannon makes better performances (generating 39 rules) than J-Measurement (only 5 rules). The error rates are significantly changes between two looms (0.1083 and 0.6369).

The different values for Recall in a Chart layout is given below

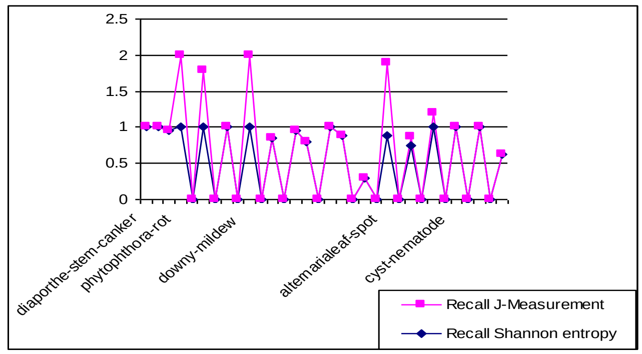

Fig 3 – Recall values for Shannon and J-Measurement

8. Conclusion

Now-a-days we need such type of process where less of time, less of error, speedy, user friendly environment are highly needed for making success of an expert system or KBS. The rules are so important for this activity. Here the two approaches are used as rule evaluation parameters that make lots key differences in their separate performances. We use soybean data set here with few examples but in future we can take large data bases that help to point reserve and much wider difference between two processes. Rule extraction some how depends on desired goals, methods and tools. Evaluation of rule major criteria for the framework of the KBS. The Shannon entropy and J-Measurement are two important rule evaluation parameter for generating and asses the rules which helps to lead the KBS. Decision list another vital method for data mining that makes decision by generating rules for a particular system.

There is some dissimilarity in the rule that generated between by the Shannon entropy and J-Measurement as the rule evaluation parameter we can see the effect of the corresponding methods and the process in the data set. The differences help to understand the effect and performances which including error rates, execution time, classifier performances, values prediction and the generated rules.

References

[1] Arash Ghorbannia Delavar,Mehdi Zekriyapanah Gashti,, Behroz nori Lohrasbi , Mohsen Nejadkheirallah," *RMSD: An*

ISSN 2229-5518

optimal algorithm for distributed systems resource whit data mining mechanism", Canadian Journal on Artificial Intelligence, Machine Learning and Pattern Recognition Vol. 2, No. 2, February 2011

[2] A.Ghorbannia Delavar , M.Zekriyapanah Gashti , B.Noori Lahrod ," *ERPASD: A Novel Algorithm forIntegrated Distributed Reliable Systems Using Data Mining Mechanisms "* , IEEE ICIFE 2010 , September 17-19, 2010, Chongqing, China

[3] Arash Ghorbannia Delavar , Narjes Rohani , Mehdi Zekriyapanah Gashti . "ERPAC*: A Novel Framework for Integrated Distributed Systems Using Data Mining Mechanisms* ", International Conference on Software Technology and Engineering. IEEE ICSTE 2010, October 3-5, 2010. San Juan, Puerto Rico, USA

[4] José C. Riquelme, Jesús S. Aguilar and Miguel Toro, *"Discovering Hierarchical Decision Rules With Evoluive Algorithm in Supervised Learning"* IJCSS, Vol.1, No.1, 2000

[5] Fariba Shadabi and Dharmendra Sharma, *"Artificial Intelligence and Data Mining Techniques in Medicine – Success Stories"*, 2008 International Conference on Bio Medical Engineering and Informatics

[6] Harleen Kaur and Siri Krishan Wasan, "*Empirical Study on Applications of Data Mining Techniques in Healthcare "*

[7] Data Mining: Concepts and Techniques Jiawei Han and Micheline Kamber, Morgan Kaufmann, 2001

[8] Portia A. Cerny "Data mining and Neural Networks from a Commercial Perspective "

[9] Antony Browne, Brian D. Hudson; , David C. Whitley , Martyn G, Philip Picton *"Biological data mining with neural networks: implementation and application of a flexile decision tree extraction algorithm to genomic problem domains"*

[10] K. Mumtaz1 S. A. Sheriff 2 and K. Duraiswamy , *"Evaluation of Three Neural Network Models using Wisconsin Breast Cancer Database"* Journal of Theoretical and Applied Information Technology P 37 -42

[11] A.Vesely **"***NEURAL NETWORKS IN DATA MINING*" AGRIC. ECON. – CZECH, *49*, 2003 (9): 427–431