₹ 200 ISSN - 2249-555X

Volume : 1 Issue : 6 March 2012



Journal for All Subjects

www.ijar.in

Listed in International ISSN Directory, Paris.

ISSN - 2249-555X



Indian Journal of Applied Research

Journal for All Subjects

Editor-In-Chief

Dr A Kumar
Director, College Development Council (CDC)
Director, Internal Quality Assurance Cell (IQAC)
Professor in Management,
Department of Business Administration, Faculty of Management,
Bhavnagar University,

Editorial Advisory Board

Dr. S. N. Pathan
Maharastra

Dr. SM. Ramasamy
Gandhigram

Dr. M. M. Goel
Kurukshetra

Dr. S. Ramesh
Tamil Nadu

Dr Ramesh Kumar Miryala
Nalgonda.

Dr. B. Rajasekaran
Tirunelveli

Dr. A. R. Saravankumar
Tamilnadu

Dr. Roy M. Thomas
Cochin

Dr. G. Selvakumar
Salem

Dr. Apurba Ratan Ghosh
Burdwan

Dr. Shrawan K Sharma
Uttarakhand

Dr. Sudhanshu Joshi
Uttarakhand

Prof. (Dr.) B Anandampilai
Pudhukottai

Advertisement Details

Position	B/W (Single Color)	Fore Color
Full Inside Cover	₹ 6000	₹ 12500
Full Page (Inside)	₹ 5000	-

Subscription Details

Period	Rate	Discount	Amount Payable
One Year (12 Issues)	₹ 2400	Nil	₹ 2400
Two Year (24 issues)	₹ 4800	₹ 200	₹ 4600
Three Year (36 issues)	₹ 7200	₹ 300	₹ 6900
Five Year (60 issues)	₹12000	₹ 600	₹ 11400

You can download the Advertisement / Subscription Form from website www.ijar.in. You will require to print the form. Please fill the form completely and send it to the **Editor, INDIAN JOURNAL OF APPLIED RESEARCH** along with the payment in the form of Demand Draft/Cheque at Par drawn in favour of **INDIAN JOURNAL OF APPLIED RESEARCH** payable at Ahmedabad.

1. Thoughts, language vision and example in published research paper are entirely of author of research paper. It is not necessary that both editor and editorial board are satisfied by the research paper. The responsibility of the matter of research paper/article is entirely of author.
2. Editing of the Indian Journal of Applied Research is processed without any remittance. The selection and publication is done after recommendations of atleast two subject expert referees.
3. In any condition if any National/International University denies accepting the research paper published in IJAR, then it is not the responsibility of Editor, Publisher and Management.
4. Only the first author is entitle to receive the copies of all co-authors
5. Before re-use of published research paper in any manner, it is compulsory to take written permission from the Editor-IJAR, unless it will be assumed as disobedience of copyright rules.
5. All the legal undertaking related to Indian Journal of Applied Research is subject to Ahmedabad Jurisdiction.
7. The research journal will be send by normal post. If the journal is not received by the author of research papers then it will not be the responsibility of the Editor and publisher. The amount for registered post should be borne by author of the research paper in case of second copy of the journal.

Editor,
Indian Journal Of Applied Research
8-A, Banans, Opp. SLU Girls College, New Congres Bhavan, Paldi,
Ahmedabad-380006**,** Gujarat, INDIA
Contact.: +91-9824097643 E-mail : editor@ijar.in

INDEX

Sr. No.	Title	Author	Subject	Page No.
1	Impact of Liquid biofertilizers, Chemical fertilizers and Vermicompost on the Growth and Yield of Chenopodium album (L.)	Dr. Sivgami Srinivasan, A. Sabitha	Biochemistry	1-3
2	Production Of Biodiesel From Microalgae Using Sodium Methoxide	M. R. Ananadhi Padmanabhan, Dr. Shaleesha A.Stanley	Biotechnology	4-6
3	Profitability Analysis Of Bharath Sanchar Nigam Limited (Bsnl)	Dr. M.Dhanabhakyam, Swapna Kurain	Commerce	7-9
4	Water Resource Management Through Community Initiatives	Dr. Mahalaxmi Krishnan	Commerce	10-11
5	IPO Grading – An Insight	Dr. Bheemanagouda	Commerce	12-14
6	Public Private Partnership (PPP) in West Bengal: A Study	Suvarun Goswami	Commerce	15-17
7	Factors Influencing Impulse Buying of Personal Care Products in Salem City	Dr. A. Vinayagamoorthy, P. kannan	Commerce	18-21
8	Employees Savings and Investment Behaviour of Selected Chemical Industries in Cuddalore District	Dr. G. Sasikumar, R. Punniyaseelan	Commerce	22-28
9	Object Oriented Mutation Applied to Java platform and C++ classes	Ravindra K. Gupta, Dr. Shailendra Singh,Titu Singh Arora	Computer Science	29-32
10	Estimating The Intravenous Fluid Resuscitation In Patients In The Intensive Care Units (A Fuzzy Logic Based Expert Advisory System)	Kamal Kumari Dadhirao, M. Surendra Prasad Babu, Mr. Suresh Kumar Maddila	Computer Science	33-36
11	An Evaluation of Women Construction Workers	T. Christy Chanchu, T. Christy Cresida	Economics	37-39
12	Carlson's Tropic State Index(CTSI) Used in Mandya Lakes	Nagarathna, Hosmani S P	Economics	40-42
13	A water pollution of metropolitan cities in India	Dr. Pawar A.S.	Economics	43-44
14	Urbanization & Pollution in India	Dr. Pawar Ashok Shankarrao	Economics	45-46
15	Motivational Strategies on Enhancing Competencies of LIC Professionals	Dr. K. Gunasekaran	Education	47-49
16	Awareness On Socio - Educational Rights Among Students Studying in Secondary School	Dr. A. Selvan, Dr. P. Paul Devanesan	Education	50-52
17	Major Approaches in Instiling Values Related to Human Rights	Dr. A. Selvan, P. Paul Devanesan	Education	53-54
18	A Study On Achievement In Mathematics of IXth Standard Students in Relation to Locus of Control and Test Anxiety	Dr. N .Rajashekara, Prathima .H.P, Narayana Swamy.Y.R	Education	55-58
19	Impact of value-orientation on Leadership Behaviour of Secondary School Teachers in Tumkur District	Dr. Umme kulsum, G.P.Nagaraja	Education	59-62
20	Performance Comparison between LMS and NLMS Algorithm	Baljit Kaur	Engineering	63-65
21	Overview of Speech Enhancement Techniques for Various Applications	N B Umashankar, Anand Jatti	Engineering	66-67

22	Hanford and Radioactive Tumbleweeds: A Case Study	Rajesh Gopinath	Engineering	68-69
23	Construction of a Knowledge Test for Tennis Players	Sanjay Singh, Sani Kumar Verma	Engineering	70-72
24	A Noval Switch Mode Power Supply for Flash Lamp Pumped Nd: Yag Laser	Kireet Semwal, S. C. Bhatt	Engineering	73-75
25	Uniform Price Electricity Market Model in Deregulated Power System	T. Nireekshana, Dr. G. Kesava Rao, Dr. S. Siva Naga Raju	Engineering	76-79
26	Effect of Notch Orientation, Temperature and Filler Material on Impact Toughness of GFRP Composites	M.C.Murugesh, Dr. K. Sadashivappa	Engineering	80-83
27	Prediction of Heat transfer rate for Indian coal slurry based on regression analysis	C.M. Raguraman, A. Ragupathy, L. Sivakumar	Engineering	84-86
28	A study of antimicrobial activity of the flavonoid-Kaempferol-3-rutinoside- from Gliricidia sepium (Jacq.) kunth.ex.Walp.	Anandhi, B., Sukumar. D, Arivudainambi. R	Engineering	87-88
29	Amitav Ghosh: Transfiguration of Memory and Imagination in the Shadow Lines	S.Kokila	English	89-90
30	Shareholder Activism in the U.S. and India	Ch. Chaitanya	Finance	91-93
31	Growth and Trends of Population in Nanded City, Maharashtra	Dr. Rathod S. B.	Geography	94-96
32	Regional Disparities of Ground water irrigation in Solapur District: A Geographical Analysis	Dr. G.U.Todkari, Shri B.D. Patil	Geography	97-99
33	Women and Employment in Textile Industries	Dr. S. S. Vijayanchali, Dr. E. Arumuga Gandhi	Home Science	100-102
34	Transition in Nutrition and Marketing of Agro Products	Dr. Anjali Rajwade	Home Science	103-104
35	Assessment of Knowledge About Effects of Inadequate Dietary Pattern on Mother and Child Health	Dr. Anjali Rajwade, Dr. Charuta Rajwade	Home Science	105-106
36	Job Enrichment	Kapil Dev Upadhyay, Dr. (Prof.) Vijay Kumar Soni	Human Resource Management	107-108
37	Quality of Work Life	Himani Avasthi, Dr. (Prof.) Vijay Kumar Soni	Human Resource Management	109-110
38	Understanding Security Requirement Engineering	Velayutham Pavanasam, Chandrasekaran Subramaniam	Information Technology	111-112
39	GENE Prediction	Er. Pankaj Bhambri,Dr. O.P. Gupta,Er. Meenakshi	Information Technology	113-114
40	Predictive Analysis In Data Mining Using Weighted Associative Classifier	Suwarna Gothane	Information Technology	115-119
41	Exchange Rate Forecasting	Dr. T. Koti Reddy	International Trade	120-124
42	Mass Media in India – New Trends	J. Ramanjaneyulu	Law	125-127
43	Right to Education Under the Indian Constitution	Mannava Muni Hemachand	Law	128-130
44	Criminal Law: A Tool for Consumer Protection in India	Aneesh V. Pillai	Law	131-132

45	Public Trust Doctrine: A Repository of Governmental Obligations to Protect the Environment	Dr. Kaumudhi Challa	Law	133-134
46	Application of Biometrics in the Investigation of Crime	Dr. R. Thilagaraj, Mr. G. Deepak Raj Rao	Law	135-137
47	Factors and Situations of Language Change	Dr. Syed Mohammed Haseebuddin Quadri	Literature	138-139
48	Omprakash Valmiki ke kahanio me chitrit dalitoki samajik stithi	Dr. V. P. Mahalinge	Literature	140-141
49	Fund Management in Pharmaceutical Industry in India	Dr. R. Perumal	Management	142-145
50	A study of Relationship management with customers in hospitality sector	Afreen Nishat A. Nasabi	Management	146-147
51	Outsourcing of Geriatric Services	Dr. V. Sreedevi	Management	148-150
52	Factor Analysis and Business Research	Dr. Kakali Majumdar	Management	151-154
53	Faculty Turnover Intention in Educational Institutions	K. R. Sowmya, Dr. N. Panchanatham	Management	155-156
54	Corporate Governance	Kalola Rimaben A., Chauhan Lalit R.	Management	157-159
55	Investment Avenues and Tax Implications for Small Investors	P. Menakadevi, Dr. V. Kumaravel	Management	160-163
56	Analyzing the IT Industry Progressions and Comprehensive Market Downfall Consciences through Deliberate Thinking	Dr. Venkatesh. J, Ms. Aarthy. C	Management	164-166
57	A Study On Problems and Prospects of Farmers in Cultivation of Paddy in Mahabubnagar District, A.P	Mr. M. Madhu Mohan	Management	167-170
58	An Analytical Study of Cash Management of IT Industries in Global Era	Bhavesh P Chadamiya, Mital R Menapara	Management	171-173
59	The Status & Performance of Registered Small Enterprises; A Study in Karnataka	Dr. D. Kumuda	Management	174-179
60	Malnutrition – A Business Opportunity for State Cooperative Dairy Federations (Scdfs)	Dr. N. Ramanjaneyalu	Management	180-181
61	Financial Performance of Indian Corporate Sectors During Pre and Post Mergers and Acquisitions	Dr. Shailesh N. Ransariya, Dr. Butalal C. Ajmera	Management	182-184
62	A Study on Influence of Pedagogical Effectiveness on Academic Achievement of Students of B Schools	Dr. Viral Shilu	Management	185-186
63	Stimulation of Entrepreneurship through Venture Capital in India	Vijayakumar, Subrahmanya K C	Management	187-189
64	Performance Analysis of Indian Banks	Ajay K. Shah	Management	190-192
65	An Emprical Study on Effect of Combination of Sun with Different Zodiac Signs on Return of it Sector of Indian Stock Market	Dr. Jaydip Chaudhari, Pinkal Shah	Management	193-197
66	Internet Banking Customer Satisfaction and Online Banking Service Attributes	Md. Mahtab Alam, Dr. Umesh R. Dangarwala	Management	198-199
67	Changing Customer expectations & preferences in Indian Banking	Ms. Smita S. Shetty	Management	200-202
68	Current e-Governance Scenario in Healthcare sector in India	Samir Roy	Management	203-204

69	Liver function abnormality & Vestibular symptoms with Azithromycin	Dr. Parveen Kumar Sharma, Dr. Rekha Bansal	Medical Science	205-206
70	Unicornuate Uterus with Functional Noncommunicating Rudimentary Horn - A Rare Mullerian Anomaly	Anju Kumari Rani, Anand Prakash, Deepa Kapoor	Medical Science	207-208
71	Antenatal Diagnosis and Successful Delivery of a Fetus with Complete Heart Block	Anju Kumari Rani, Aditya Kapoor, Deepa Kapoor	Medical Science	209-210
72	Kangaroo mother care diminishes pain from heel lance in preterm neonates:	Dr.S.Valliammal, Dr. Ramachandra, Mr. Raja Sudhakar	Nursing	211-213
73	Experiential Domain and Extentionalistic Vocabulary : A Critical Consideration of Behaviourism	Dr. Jatinder Kumar Sharma	Philosophy	214-215
74	The Effect of a Mindfulness-Based Stress Reduction (MBSR) Program on Stress and Well-Being in Adolescents: A Pilot Study	Urvashi Anand, Dr. Mahendra P. Sharma	Psychology	216-218
75	Personality Profile of Patients with Somatic Complaints at a Rural Health Centre	Dr. Mahendra P. Sharma,B. Padmasri and,Dr. Ramchandra	Psychology	219-220
76	Psychological Well-being: A study of the institutionalized aged	Dr. Pankaj S. Suvera	Psychology	221-223
77	Crystal Growth, Thermal and Optical Studies on L-arginine Based Nonlinear Optical Material	J. Ramajothi	Science	224-226
78	Encapsulation of Lactic Acid Bacteria in Calcium Alginate Beads for Bacteriocin Production	Vijay S. Wadhai, Shweta G. Kayarkar, Archana N. Dixit	Science	227-229
79	Synthesis Of Nonatitnate Based Dielectric Ceramic Using Various Grain Size of Starting Materials	Parul, Anand K Tyagi	Science	230-232
80	Separate Telangana Movement and Related Literary Works: A Study In Sociology of Literature	Dr. M. Sreedevi Xavier	Sociology	233-234
81	Changing Interests In Fine Arts as a Profession: A Study In Sociology of Art	Dr. Poonam Gandhi-Moirangthem	Sociology	235-236
82	Modeling Fertility and Growth of Mising Community of Assam, India	Deba Kr. Baruah, S. C. Kakaty	Statistics	237-239
83	Fort Vasota an Adventure Tourism Destination in Satara District	Dr. Ananadrao S. Patil	Tourism	240-242

Volume : 1 | Issue : 6 | March 2012 ISSN - 2249-555X

Research Paper Information Technology



Predictive Analysis In Data Mining Using Weighted Associative Classifier

* Suwarna Gothane

* Asst Prof ATRI, Hyderabad

ABSTRACT

Association rule mining is well researched techniques of data mining for expressive task, initially used in market analysis. Based on existing rules in transactional database those satisfy some minimum support and confidence constraints. Classification using Association rule mining is another major Predictive analysis technique that aims to discover a small set of rule in the database that forms an accurate classifier. Associative Classifier the combined approach that integrates association rule mining and classification rule mining called Associative Classification(AC). There are many AC approaches that have been proposed CBA, CMAR, CPAR. Proposed system introducing w-support(a new measure of itemsets) in databases. W-support introduced as frequency of itemset may not be important as it appears, because the weights of transactions are different. These weights are derived from internal structure of database based on the assumption that good transactions consist of good items. This assumption is exploited by extending Kleinberg's HITS model(Bipartite graphs).

Keywords : Associated Rules', 'Predictive analysis', 'Weighted associative classifier

1. Introduction

Data mining is defined as "the nontrivial extraction of implicit, previously unknown, and potentially useful information from data". Association rules represent a data mining technique that has great potential in the medical domain to improve disease prediction. It can be said that 'knowledge is power' only to the extent that it is useful in making such predictions as accurately as possible. Classification and prediction are two forms of data mining functionalities that can be used to extract predictive models. A new predictive modeling approach known as associative classification, integrating association Mining and classification inside into single system is being discussed as better alternative for predictive analytics[1]. Some of the classification techniques presented are Class Base Association (CBA)[2], Classification Based on Multiple Class-Association Rules (CMAR)[3], Classification based on Predictive Association Rule (CPAR)[4]. As discussed in it achieves higher classification accuracy than do traditional classification approaches such as C4.5, FOIL, RIPPER. According to[2] these traditional classifiers are faster but in many cases accuracy is not so high and many of the rules found by associative classification method cannot be discovered by traditional classification algorithm. Medical field is good examples where such applications may appear. Let us consider an example were a physician has to examine a patient. There is a considerable amount of information associated with the patient (e.g. personal data, medical tests, etc.). A classification system can assist the physician in this process. The system can predict if the patient is likely to have a certain disease or present incompatibility with some treatments. Considering the output of the classification model, the physician can make a better decision on the treatment to be applied to the patient[5].

This paper proposes target to develop weighted associative classifiers that classifies the rules based on weighted support and weighted confidence. In section 2, we have discussed the concept of data mining, association rules. In section 3 we described association classification. In section 4; Weighted Associative Classifiers is being discussed. In section 5, Extending the process to weighted associative classifier where association rule mining will be performed based on transaction and attribute weights. Here attribute weights will be measured based on attribute and transaction relation & identified using HITS algorithm. In section 6, conclusion and future work of this paper is given.

2.Data Mining and Association rules

2.1 Data Mining

Data mining (knowledge discovery from data) is extraction of interesting (nontrivial, implicit, previously unknown and potentially useful) patterns or knowledge from huge amount of data. Knowledge discovery (mining) in databases (KDD), knowledge extraction, data/pattern analysis, data archeology, data dredging, information harvesting, business intelligence are alternative names of data mining. Classification and prediction are the data mining functionalities which can be used to extract predictive models describing important data classes or to predict future data trends.

Data mining or knowledge discovery in databases (KDD) is a collection of exploration techniques based on advanced analytical methods and tools for handling a large amount of information. The techniques can find novel patterns that may assist an enterprise in understanding the business better and in forecasting. Data mining is a collection of techniques for efficient automated discovery of previously unknown, valid, novel, useful and understandable patterns in large databases. The patterns must be actionable so that they may be used in a decision of an enterprise making process.

2.2Association Rules

Classically, given a transactional dataset D and a set of data items I={I1, I2 …, Im}, an association rule (AR) is of the form: $X \Rightarrow Y$, expressing the semantics that "occurrence of X is associated with occurrence of Y", where X and Y are disjoint collections of data items (i.e., X and Y are called itemsets, $X, Y \subset I$, and $X \cup Y = \varnothing$). Also, X is referred to as the antecedent of an AR and Y as consequent of the AR. Let |D| be the number of transactions (interchangeably referred to as records, otherwise indicated where necessary) in D, and ||X||(||XY||) be the number of transactions containing X (XY), the degrees of support and confidence (denoted as Dsupp and Dconf respectively) for $X \Rightarrow Y$ are defined as: $Dsupp(X \Rightarrow Y) = ||XY||/|D|$ and $Dconf(X \Rightarrow Y) = ||XY||/||X||$[6]. An example of an association rule in supermarket basket transactions is "Fruit $\Rightarrow$Vegwtable, with Dsupp=20% and Dconf=80%" meaning that "20%

of the customers bought Fruit and Vegetable simultaneously, and 80% of the customers who bought Fruit also tended to buy Vegetable". Such association rules are also referred to as Boolean ARs, since the association concerned is the correspondence of the occurrence states, each being a binary value 0 or 1. There have been many efforts proposed to discover Boolean ARs in various ways[7] among which the Apriori algorithm[8] is usually deemed as a typical one. In addition to binary data values in Boolean association rules, categorical or generally quantitative data values are commonly encountered in many real applications. As shown in Table 1, for instance, Age and Height are quantitative attributes, whereas Bald is a categorical attribute.

Table 1: Database(D) with Continuous Domains

D	Age	Height	Bald	….
ID1	35	170	1	….
ID2	24	180	0	….
ID3	15	182	0	….
ID4	57	165	1	….
….	….	….	….	….

While the Apriori algorithm is effective and efficient, it can hardly be used directly in discovery of quantitative association rules. A commonly used way is to partition the domains and therefore introduce new attributes with intervals. In doing so, the original transactional dataset D could be transformed into a binary database D′ which is exemplified in Table 2. Accordinglv. quantitative ARs such as "Age(65,100) & Height(180,190)⇒Bald" can be discovered using a Apriori-type algorithm.

Table 2: Database(D') Transformed from D.

D'	Age (0,25)	Age (25,60)	Age (60,100)	Bald	….
ID1	0	1	0	1	….
ID2	1	0	0	0	….
ID3	1	0	0	0	….
ID4	0	1	0	1	….
….	….	….	….	….	….

3.Associative Classifications

Associative classification (AC) mining is a promising approach in data mining that utilizes the association rule discovery techniques to construct classification systems is a three-step process.

1. Generate the set of association rules from the training set with certain support and confidence thresholds as candidate rules.
2. Pruning the set of discovered rules to weed out those rules that may introduce over fitting.
3. Classification Phase is the step to make a prediction for test data and measure the accuracy of the classifier.

An association rule is an implication of the form A=>B, where A, B C I, where I is set of all items, and A ∩ B=Ø. The rule A=>B has a support s in the transaction set D if s% of the transactions in D contain A U B .The rule A=>B holds in the transaction set D with confidence c if c% of transactions in D that contain A also contain B. The problem of discovering all association rules from a set of transactions D consists of generating the rules that have a support and confidence greater than given thresholds. These rules are called strong rules, and the framework is known as the support confidence framework for association rule mining.

The AC is performed on relational databases that contain records of entities of particular domain. Each record consists of m attribute a1, a2 …..am and one important attribute known as predictive Attribute (Class Label). The attributes are continuous or categorical. Continuous attributes are required to be converted into categorical by using discritization technique. An example of such database from medical field is shown in table 3.

The AC is different from association rule mining. Following are the terminology for AC analogous to Association Rule Mining.

i. Attribute which is pair (attribute, value), is used in place of Item. For example (Age, 62) is an attribute.
ii. Attribute set is equivalent to Itemset for example ((Age, old), (BP, high)).
iii. Support count of Attribute (Ai, vi) is number of rows that matches Attribute in database and Support count of Attribute set (Ai, vi)…(Am, vm) is number of rows that matches Attribute set in data base.

Table3: Sample Database for Heart Patient

R_ID	Age	Smoking Habits	Hypertension	BMI	Heart_Disease
1	40	Yes	Yes	40	Yes
2	64	Yes	No	28	No
3	55	No	Yes	40	Yes
4	62	Yes	Yes	50	Yes
5	44	No	Yes	30	No

iv. An Attribute set $((A_i, v_i)…(A_m, v_m))$ passes the minsup threshold if support count $((A_i, v_i)…(A_j, v_j))$, ≥minsup.
v. CAR Rules are of form $((A_i, v_i), …,(A_j, v_j))\rightarrow c$ where c Class-Label. Where Left hand side is itemset and right hand sigh is class. And set of all attribute and class label together i.e $((A_i, v_i), …,(A_j, v_j),c)$ is called rule attribute.
vi. Support count of rule attribute $((A_i, v_i), …,(A_j, v_j), c)$ is number of rows that matches item in database.
vii. Rule attribute $((A_i, v_i), …,(A_j, v_j), c)$ passes the minsup threshold if support count of $((A_i, v_i), …,(A_j, v_j), c)$ ≥minsup.

A special subset of association rules whose right-hand-side is restricted to the class attribute is used for classification. This subset of rules is referred as the Class Association Rules (CARs).Using association rule for classification is advantageous over traditional classifiers is that the simple if -then -else rules are used for classification which makes it easy for the end user to understand and interpret it. Also unlike decision tree approach one can easily update the rule set without affecting the complete rule set where as in decision tree approach it requires reshaping of complete tree.

4. Weighted Associative Classifiers

Weighted Associative Classifiers is another concept that assigns different weights to different features and can get more accuracy in predictive modeling system like medical field etc[9]. In any prediction model all attributes do not have same importance in predicting the class label then different weights can be assigned to attributes according to their predicting capability. A weighted associative classifiers consists of training dataset $T=\{r_1, r_2, r_3…. r_i…\}$ with set of weight associated with each {attribute, attribute value} pair. Each i^{th} record r_i is a set of attribute value and a weight w_i attached to each attribute of r_i tuple / record. In a weighted framework each record is set of triple$\{ a_i, v_i, w_i \}$ where attribute ai is having value v_i and weight w_i, $0 < w_i <= 1$. Weight is used to show the importance of the item. Using Weighted Associative Classifiers, weighted rules like "medium Income→young Age", "{(Age,">62"), (BMI,"45"), (Boold_pressur,"95-135")},→Heart Disease, (Income[20,000-30,000]→Age[20-30]) can be used for

Table4: Relational Database with record weight

ID	Age	Smoking Habits	Hyper-Tension	BMI	Record Weight
1	40	Yes	Yes	40	(0.2+0.8+0.6+0.8) /4=0.60
2	64	Yes	No	28	(0.3+0.8+0.5+0.3) /4=0.42

3	55	No	Yes	40	(0.2+0.8+0.6+0.5) /4=0.52
4	62	Yes	Yes	28	(0.3+0.8+0.6+0.8) /4=0.67
5	44	No	Yes	30	(0.2+0.7+0.6+0.3) /4=0.45

Table5: Weight of symptoms for heart disease (attribute weight)

Sr.No	Symptoms	Weights
1	Age<40	0.1
2	40<age<58	0.2
3	age>58	0.3
4	Smoking_habits=yes	0.8
5	Smoking_habits=no	0.7
6	Hypertension=yes	0.6
7	Hypertension=no	0.5
8	BMI<=25	0.1
9	26<=BMI<=30	0.3
10	31<=BMI<=40	0.5
11	BMI>=40	0.8

classification. For the synthetic database given in table2, Weight of each record is calculated in Table4, using Weight of different attribute in predicting the probability of Heart Disease from Table5.

Defining Support and confidence measure:
New formulae of support and confidence for classification rule X→Class Label, where X is set of weighted items, is as follows:

Weighted Support: Weighted support WSP of rule X→Class label, where X is set of non empty subsets of attribute value set, is fraction of weight of the record that contain above attribute-value set relative to the weight of all transactions.

$$\mathrm{WSP}(X \rightarrow \text{Class_label}) = \frac{\sum_{i=1}^{|X|} \text{weight}(ri)}{\sum_{k=1}^{|n|} \text{weight}(ri)}$$

Example: The Weighted support of the rule R(Hypertension="yes")→Heart_Disease="yes" can be calculated as:

$$\mathrm{WSP}(R) = \frac{\text{Sum of Record Weight having the condition Hypertension} = \text{'yes'' true and also given class label Heart_Disease}}{\text{Sum of Weight of all transactions}}$$

Weighted Confidence: Weighted Confidence of a rule X→Y where Y represents the Class label can be defined as the ratio of Weighted Support of (X U Y) and the Weighted Support of (X).

$$\text{Weighted Confidence} = \frac{\text{Weighted Support } (X \cup Y)}{\text{Weighted Support } (X)}$$

Example: The Weighted confidence of the rule R(Hypertension="yes")→Heart_Disease="yes" can be calculated as

$$\mathrm{WC}(R) = \frac{\text{Sum of Record Weight having the condition Hypertension} = \text{'yes'' true and also class label Heart_Disease}}{\text{Sum of record Weight having the condition Hypertension} = \text{"yes"}}$$

Algorithms for Frequent Itemset Discovery:Apriori Algorithm:
Apriori uses the downward closure property of itemset support that any subset of a frequent itemset must also be frequent. Thus during each iteration of the algorithm only the itemsets found to be frequent in the previous iteration are used to generate a new candidate set, Ck. Before inserting an itemset into Ck, Apriori tests whether all its subsets are frequent. This pruning step can eliminate a lot of unnecessary candidates. The candidates are stored in a hash tree to facilitate fast support counting. An internal node of the hash tree at depth d contains a hash table whose cells point to nodes at depth d+1. All the itemsets are stored in the leaves. The insertion procedure starts at the root, and hashing on successive items, inserts a candidate in a leaf. For counting Ck, for each transaction in the database, all k-subsets of the transaction are generated in lexicographical order. Each subset is searched in the hash tree, and the count of the candidate incremented if it matches the

```
L1 = {frequent 1-itemsets};

for (k=2; Lk ≠ Ø; k++)

Ck= Set Of New Candidates;

for all transaction t ∈ D
for all k-subsets of t

if (s Ck) s.count++;

Lk = {c ε Ck|c.count ≥ minimum support};
Set of all frequent itemsets = Uk Lk;
```

Figure 1: Apriori Algorithm

subset. This is the most compute intensive step of the algorithm. The last step forms Lk by selecting itemsets meeting the minimum support criterion. The complete algorithm is shown in figure1.

Existing System: Apriori Using Preassighned Weights
The most important aspect of Apriori is to completely incorporate the subset frequency based pruning optimization that is, it does process any itemset whose subset is known to be infrequent. It utilizes data structure called hash tree to store the counters of candidate itemsets. Since, there are usually a large number of distinct single items in a typical transaction database and their combinations may form a very huge number of itemsets, it is challenging to develop scalable methods for mining frequent itemsets in a large transaction database. Here downward closure property, called Apriori is used, that is among frequent K-itemsets: A k-itemset is frequent only if all of its sub-itemsets are frequent. This implies that frequent itemsets can be mined by first scanning the database to find the frequent 1-itemsets, then using the frequent 1-itemsets to generate candidate frequent 2-itemsets and check against the database to obtain the frequent 2-itemsets. This process iterates until no more frequent k-itemsets can be generated for some k. This is the essence of Apriori algorithm. In existing system the attribute weights are already predefined. By using this weight the result is not accurate because the weights are predefined weights.

Limitations of Apriori
The main drawback in this algorithm is that it performs n passes over the database. Where n is the length of the longest frequent itemset. In Kth pass, the counts of candidate itemsets of length (called k-itemsets) are obtained.

Another drawback is that Apriori follows a tuple-by-tuple approach that is, it updates counters of candidate itemsets after reading in each transaction from the database. It hence suffers from the drawback that much redundant work (traversal

of the data structure holding the counters of itemsets) is performed after each and every transaction.

5Proposed System

5.1Apriori Using Automated Weights
Extending the process to weighted associative classifier where association rule mining will be performed based on transaction and attribute weights. Here these attribute weights will be measured based on attribute and transaction relation. This attribute and transaction relation will be identified using HITS algorithm.

Measuring weights/ranking using HITS
A database of transactions can be depicted as a bipartite graph without loss of information.

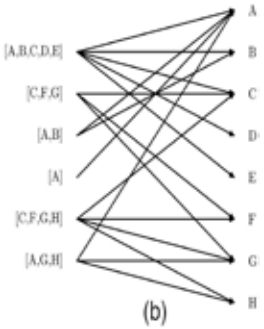
Let D be the, $D = \{T_1, T_2, ...T_m\}$ a list of transactions and $I = \{i_1, i_2, ..i_n\}$ be the corresponding set of items. Then, clearly D is equivalent to the bipartite graph $G = (D, I, E)$ where,

$$E = \{(T, i) : i \in T, T \in D, i \in I\}$$

Example 1: Consider the database shown in figure2a It can be equivalently represented as a bipartite graph, as shown in figure2b. The graph representation of

TID	Transaction
100	[A, B, C, D, E]
200	[C, F, G]
300	[A, B]
400	[A]
500	[C, F, G, H]
600	[A, G, H]

(a)



(b)

Figure2: The bipartite graph representation of a database (a) Database (b) Bipartite graph
the transaction database is inspiring. It gives us the idea of applying link-based ranking models to the evaluation of transactions. In this bipartite graph, the support of an item i is proportional to its degree, which shows again that the classical support does not consider the difference between transactions. However, it is crucial to have different weights for different transactions in order to reflect their different importance. The evaluation of itemsets should be derived from these weights. Intuitively, a good transaction, which is highly weighted, should contain many good items; at the same time, a good item should be contained by many good transactions. The reinforcing relationship of transactions and items is just like the relationship between hubs and authorities in the HITS model[10]. Regarding the transactions as "pure" hubs and the items as "pure" authorities, we can apply HITS to this bipartite graph. The following equations are used in iterations:

$$auth(i) = \sum_{T:i \in T} hub(T),\ hub(T) = \sum_{i:i \in T} auth(i) ....(1)$$

When the HITS model eventually converges, the hub weights of all transactions are obtained. These weights represent the potential of transactions to contain high-value items. A transaction with few items may still be a good hub if all component items are top ranked. Conversely, a transaction with many ordinary items may have a low hub weight.

5.2W-support - A New Measurement:
Itemset evaluation by support in classical association rule mining is based on counting[11]. In this section, introducing a link-based measure called w-support and formulate association rule mining in terms of this new concept. As the iteration converges, the authority weight

$$auth(i) = \sum_{T:i \in T} hub(T)$$

represents the "significance" of an item I, accordingly, generalize the formula of $auth(i)$ to depict the significance of an arbitrary itemset, as the following definition shows:

Definition 1: The w-support of an itemset X is defined as

$$w\sup p(X) = \frac{\sum_{T:X \subset T \wedge T \in D} hub(T)}{\sum_{T:T \in D} hub(T)} ....(2)$$

Where $hub(T)$ is the hub weight of transaction T. An itemset is said to be significant if its w-support is larger than a user specified value.

Example. Consider the database shown in Figure2 again. The HITS iteration gives the hub weight of each transaction and w-support of each 1-itemset. It is interesting to point out that the best hub (transaction 500[C FG H]) is not the one with the largest item number, and the most significant 1-itemset({C}) is not the one with the largest support. This shows the intrinsic difference between link-based and counting-based measurement. Transactions 200 and 500 and items C, F, and G form a complete bipartite graph, which implies that a strong cross-selling effect exists between the three items. These items should be highly evaluated because they not only occur frequently by themselves but also reinforce the value of each other by occurring together. On the other hand, although item A has the highest support, it seldom shows up with other valuable items[12]. Thus, A should be ranked somewhat lower. In essence, w-support introduces the cross-selling effect into the evaluation of itemsets.

For association rules, we give the following definition.

Definition2. The w-support of an association rule X=>Y is defined

wsupp (X=>Y) = wsupp(X U Y)

$$wconf(X \Rightarrow Y) = \frac{wsupp(X \cup Y)}{wsupp(X)}.$$

The w-confidence can be understood as the ratio of the hub weights received by X together with Y to the total hub weights received by X. Basically, w-support measures how significantly X and Y appear together; w-confidence measures how strong the rule is. If wconf(X=>Y) is large, it shows that many good hubs that vote X also vote Y, although the fraction of these hubs may be small. Accordingly, association rule mining is to discover all rules with w-support and w-confidence above some given thresholds.

6.Conclusion and Future Work
The HITS model is use to derive the weights of transactions from a database. Based on these weights, a new measure w-support is defined to give the significance of itemsets. It differs from the traditional support in taking the quality of transactions into consideration. Then, the w-support and w-confidence of association rules are defined in analogy to the definition of support and confidence. The proposed HITS algorithm based weight measurement model is significantly improving the quality of classifier.

This advanced AC method could be applied in future in many real time scenarios to get more accurate results. For example Fuzzy and Weighted AC can be combined to give more accurate results as most of the data in medical is quantitative and weight can be applied to each feature based on their prediction capability. The temporal aspect can be used to improve the prediction for the patients of different age group.

REFERENCES

[1]Ranjana Vyas, Lokesh Kumar Sharma, Om Prakash vyas, Simon Scheider Associative Classifiers for Predictive analytics: Comparative Performance Study, second UKSIM European Symposium on Computer Modeling and Simulation 2008 | [2]W. Li, J. Han, and J. Pei. CMAR: Accurate and efficient classification based on multiple class association rules. In ICDM'01, pp. 369-376, San Jose, CA, Nov.2001. | [3]Yin, X. & Han, J. CPAR: Classification based on predictive association rule. In Proceedings of the SIAM International Conference on Data Mining. SanFrancisco, CA: SIAM Press, 2003,pp. 369-376. | [4]Feng Tao, Fionn Murtagh and Mohsen Farid. Weighted Association Rule Mining using Weighted Support and Significance Framework Proceedings of the ninth ACMSIGKDD international conference on Knowledge discovery and data mining 2003, Pages:661-666 Year of Publication: 2003 | [5]Fadi Thabtah, A review of associative classification mining, The Knowledge Engineering Review, Volume 22, Issue 1 (March 2007), Pages 37-65, 2007. | [6]Agrawal R., Srikant R.. Fast Algorithms for Mining Association Rules, In: Proceedings of 20th International Conference onVery Large Databases, Santiago, Chile, pp. 487-499, 1994. | [7] G. Chen, Q. Wei, D. Liu and G. Wets, Simple association rules (SAR) and the SAR-based rule discovery, Computers and Industrial Engineering 43 (2002) 721–733. | [8]H. Toivonen, Sampling large databases for association rules, in Proc. of the 22nd Int. Conf. on Very Large Database (Bombay, India), pp. 134-145. | [9]Sunita Soni, Jyothi Pillai,O.P. Vyas An Associative Classifier Using Weighted Association Rule2009International Symposium on Innovations in natural Computing, World Congress on Nature & Biologically Inspired Computing (NaBIC 2009), 978-1-4244-5612-3/09/$26.00 c_2009 IEEE 1492-1496 | [10]J.M. Kleinberg, "Authoritative Sources in a Hyperlinked Environment,"J. ACM, vol. 46, no. 5, pp. 604-632, 1999. | [11]R. Agrawal, T. Imielinski, and A. Swami, "Mining Association Rules between Sets of Items in Large Datasets," Proc. ACM SIGMOD '93, pp. 207-216, 1993 | [12] Ke Sun and Fengshan Bai" Mining Weighted Association Rules without Preassigned Weights" Ieee Transactions on knowledge and data engineering vol. 20, No. 4, pp 489-495, April 2008

Sara Publishing Academy
Indian Journal Of Applied Research
Journal for All Subjects

Editor,
Indian Journal Of Applied Research
8-A, Banans, Opp. SLU Girls College,
New Congres Bhavan, Paldi, Ahmedabad-380006.
Contact.: +91-9824097643 E-mail : editor@ijar.in

Printed at Unique Offset, Novatsing Rupam Estate, Opp. Abhay Estate, Tavdipura, Shahibaug, Ahmedabad