International Journal of Information Sciences and Application.
ISSN 0974-2255 Volume 5, Number 2 (2013), pp. 45-56

http://www.irphouse.com

Issues and Techniques in Data Mining and Cluster Techniques

Pankaj Kumar, Paritosh Kumar and Gajraj Singh

Ramjas College, University of Delhi
pkumar240183@gmail. com
CEM Kapurthala, Punjab(India)
paritosh200623@gmail. com
Ramjas College, University of Delhi
gajraj76@gmail. com

Abstract

Theobjectiveofthepaperistoshowthe issuesto befacedinDatamining andvariousclustering Techniques. Clusteringisbeingwidely usedinmany application includingmedical, finance and etc. Clustering may beappliedondatabaseusingvarious approaches, based upon distance, density, hierarchy, andpartition. Datamining is becominganincreasingly importanttoolto transformthis dataintoinformation. It iscommonlyusedinawiderange ofprofilingpractices, suchas marketing, surveillance, frauddetectionand scientificdiscovery.

Keywords Clustering, KnowledgeDiscovery, Noise, Data mining,

IIntroduction

Dataminingis theprocessofextracting hiddenpatterns fromdata. As more dataisgathered, withtheamountof datadoublingevery threeyears, [1]dataminingis becominganincreasinglyimportanttooltotransform thisdataintoinformation. Itiscommonlyusedin awiderangeofprofilingpractices, suchasmarketing, surveillance, fraud detection andscientific discovery. Datamining, theextractionof hiddenpredictive information fromlargedatabases, isa powerful newtechnologywithgreatpotentialtohelpcompaniesfocus onthe mostimportantinformationintheirdata warehouses. Dataminingtoolspredictfuture trendsand behaviours, allowingbusinessesto makeproactive, knowledge-driven decisions. The automated, prospectiveanalysesofferedby dataminingmove beyondthe analysesofpasteventsprovidedby retrospective toolstypicalof decisionsupportsystems.

Datamining toolscananswerbusinessquestionsthat traditionallyweretoo timeconsuming to resolve. They scourdatabasesfor hiddenpatterns, finding predictive information thatexperts may miss because it lies outside theirexpectations. Mostcompaniesalready collect andrefine massive quantities of data. Data miningtechniquescanbe implementedrapidly on existingsoftwareandhardwareplatformstoenhance thevalueof existinginformationresources, andcanbe integratedwithnewproductsandsystemsasthey are brought on-line. When implementedon high performance client/serveror parallelprocessing computers, dataminingtoolscananalyzemassive databases to deliver answers to questions such as, "Whichclientsaremostlikely torespondtomynext promotionalmailing, andwhy?"Datamining is the processof discoveringmeaningfulnewcorrelations, patterns, andtrendsby siftingthroughlarge amountsof datastoredin repositories, usingpatternrecognition technologiesaswellasstatisticaland mathematical techniques". However, reallydataminingturns databases intoknowledge bases which isoneofthe fundamental components of expert systems. Data miningistheuseof automateddataanalysistechniques to uncoverpreviouslyundetectedrelationshipsamong dataitems. Data miningofteninvolvesthe analysisof datastoredinadatawarehouse. Threeofthe majordata miningtechniquesare regression, classificationand clustering. DataMining, alsopopularly knownas KnowledgeDiscoveryinDatabases(KDD)[2], refers tothe nontrivialextractionofimplicit, previously unknownandpotentiallyusefulinformationfromdata indatabases. While dataminingandknowledge discoveryindatabases(orKDD)arefrequentlytreated assynonyms, datamining isactually partofthe knowledgediscovery process. Thefollowingfigure (Figure 1. 1)showsdataminingasastepinaniterative knowledge discovery process. The KnowledgeDiscoveryinDatabasesprocess comprisesofafew stepsleadingfromraw data collections to some form of new knowledge. The iterativeprocessconsistsofthefollowing steps:

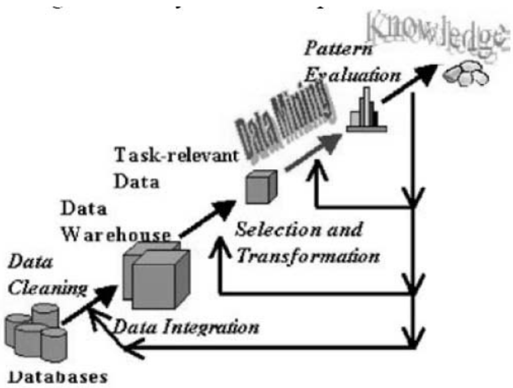

Fig 1: Aniterative Knowledge Discovery Process

a. Datacleaning: alsoknownasdatacleansing, itis a phase inwhichnoisedataandirrelevantdataare removedfromthecollection.
b. Data integration: atthis stage, multiple data sources, oftenheterogeneous, maybecombinedin acommonsource.

c. Dataselection: at thisstep, thedatarelevanttothe analysisisdecidedonand retrievedfromthedata collection.
d. Datatransformation: also known as data consolidation, it isa phaseinwhichtheselected dataistransformedintoforms appropriateforthe miningprocedure.
e. Datamining: itisthe crucialstepinwhichclever techniques are applied to extract patterns potentiallyuseful.
f. Pattern evaluation: in this step, strictly interestingpatterns representingknowledgeare identifiedbasedongivenmeasures.
g. Knowledgerepresentation: is thefinalphase in whichthe discoveredknowledge isvisually represented totheuser. Thisessential stepuses visualizationtechniquesto helpusersunderstand andinterpretthedatamining results.

IIIssuesinDataMining

Dataminingalgorithmsembody techniquesthathave sometimesexistedformany years, but haveonly lately beenappliedas reliableandscalabletoolsthattimeand againoutperform olderclassicalstatistical methods. Whiledataminingisstillinitsinfancy, itisbecoming a trendandubiquitous. Beforedataminingdevelopsinto aconventional, matureandtrusteddiscipline, many still pending issueshavetobeaddressed. Someofthese issuesareaddressedbelow. Notethattheseissuesare notexclusiveandarenotorderedinany way.

Security andsocial issues: Securityisan important issuewithany datacollectionthatissharedand/oris intendedtobeusedfor strategicdecision-making. In addition, whendatais collectedforcustomerprofiling, userbehaviourunderstanding, correlating personaldata withotherinformation, etc., large amountsofsensitive andprivateinformationaboutindividualsorcompanies is gatheredand stored. This becomescontroversial giventheconfidentialnatureofsomeofthisdataand the potential illegal access to the information. Moreover, dataminingcoulddisclosenew implicit knowledgeaboutindividualsor groupsthatcouldbe againstprivacy policies, especiallyifthereispotential disseminationof discoveredinformation. Anotherissue thatarisesfromthisconcernistheappropriateuseof data mining. Duetothevalueof data, databasesofallsortsofcontentareregularly sold, andbecauseof the competitive advantagethatcanbe attainedfromimplicitknowledge discovered, some importantinformationcouldbewithheld, while other informationcouldbewidely distributedandusedwithoutcontrol.

Userinterfaceissues: Theknowledgediscoveredby dataminingtoolsisusefulaslongasitisinteresting, andaboveallunderstandableby theuser. Gooddata visualizationeasestheinterpretationof datamining results, aswellashelpsusersbetterunderstandtheir needs. Many dataexploratoryanalysistasksare significantlyfacilitatedby theabilitytoseedatainan appropriatevisualpresentation. Thereare many visualizationideas andproposalsforeffectivedata graphicalpresentation. However, there isstill much researchto accomplishinorderto obtaingood visualizationtoolsforlargedatasetsthatcouldbeused todisplayandmanipulateminedknowledge. Themajor issuesrelatedtouser interfacesandvisualization are"screen real-estate", information rendering, and interaction. Interactivitywiththe dataanddatamining results iscrucialsinceitprovidesmeansfortheuserto focusand refinetheminingtasks, aswellasto picture the discoveredknowledgefromdifferentanglesandat differentconceptuallevels.

Miningmethodologyissues: Theseissuespertainto the data mining approaches applied and their limitations. Topics such as versatility of the mining approaches, the diversity of data available, the dimensionalityofthedomain, thebroadanalysisneeds (when known), the assessment of the knowledge discovered, theexploitationofbackgroundknowledge andmetadata, thecontrolandhandlingofnoiseindata, etc. areall examples that can dictate mining methodologychoices. Forinstance, itisoftendesirable tohavedifferentdataminingmethodsavailablesince differentapproachesmayperformdifferentlydepending uponthedataathand. Moreover, differentapproaches may suitand solve user's needs differently. Most algorithmsassumethedatatobenoise-free. Thisisof course astrong assumption. Most datasets contain exceptions, invalid or incomplete information, etc. which may complicate, if not obscure, the analysis processandinmanycasescompromisetheaccuracyof theresults. Asaconsequence, datapreprocessing (data cleaningandtransformation)becomesvital. Itisoften seenaslosttime, butdatacleaning, astime-consuming and frustrating as it may be, is one of the most importantphasesintheknowledgediscoveryprocess. Dataminingtechniquesshouldbeabletohandlenoise indataorincompleteinformation. Morethanthesizeof data, thesizeofthesearchspaceisevenmoredecisive fordataminingtechniques. Thesizeofthesearchspace isoftendependinguponthenumberofdimensionsin the domainspace. The searchspaceusuallygrowsexponentially when the number of dimensions increases. Thisisknownasthecurseofdimensionality. This"curse"affectssobadlytheperformanceofsome dataminingapproachesthatitisbecomingoneofthe mosturgent issuesto solve.

Performance issues: Manyartificial intelligence andstatisticalmethodsexistfordata analysisand interpretation. However, these methodswere oftennot designed fortheverylargedata sets data miningis dealing withtoday. Terabytesizesarecommon. This raisestheissuesof scalability andefficiencyof thedata miningmethodswhenprocessingconsiderably large data. Algorithms withexponentialandevenmedium orderpolynomialcomplexity cannotbeofpracticaluse fordata mining. Linearalgorithmsareusuallythenorm.

Insametheme, samplingcanbe usedformininginstead of thewholedataset. However, concernssuchas completenessandchoiceofsamplesmay arise. Other topicsintheissueof performanceareincremental updating, andparallelprogramming[3]. Thereis no doubt thatparallelismcanhelpsolvethesizeproblemif the datasetcanbesubdividedandthe resultscanbemerged later. Incrementalupdatingis importantformerging results fromparallelmining, orupdatingdatamining results whennew data becomes available without havingtoreanalyzethecompletedataset.

Datasourceissues: Therearemanyissuesrelatedto thedatasources, somearepracticalsuchasthediversity ofdatatypes, whileothersarephilosophicallikethe dataglutproblem. Wecertainly haveanexcessof data sincewealreadyhavemoredatathanwecanhandle andwe arestillcollectingdataatanevenhigherrate. If thespreadofdatabasemanagementsystemshas helped increasethegatheringofinformation, theadventofdata miningiscertainlyencouragingmore dataharvesting. The currentpractice is to collect as much data as possiblenowandprocessit, ortry toprocessit, later. Theconcerniswhetherwearecollectingtherightdata atthe appropriateamount, whetherweknow whatwe wanttodo withit, andwhetherwedistinguishbetween whatdataisimportantandwhatdata isinsignificant. Regardingthepracticalissuesrelatedto datasources, thereisthe subject ofheterogeneous databases and thefocuson diverse complexdatatypes. Weare storingdifferent typesofdatainavarietyofrepositories. Itisdifficultto expect a data mining system to effectively and efficiently achievegoodminingresultsonallkindsof dataandsources. Differentkindsofdataandsources may requiredistinctalgorithmsandmethodologies. Currently, thereis afocuson relationaldatabasesand datawarehouses, butotherapproaches needtobe pioneeredforotherspecificcomplexdata types. A versatiledataminingtool, forallsortsofdata, maynot berealistic.

Moreover, theproliferationof heterogeneousdata sources, at structural and semantic levels, poses important challengesnot onlytothe database communitybutalsotothedataminingcommunity.

IIIDataMiningTechniques

Thethreedataminingtechniquesare:

a. Decisiontrees
b. Neuralnetworks
c. Clustering

Decisiontree

A decisiontree [4] (or tree diagram) is a decision support tool that usesa tree-like graph or model of decisions and theirpossibleconsequences, including chanceeventoutcomes, resourcecosts, and utility. Decision trees are commonly used in operations research, specificallyin decision analysis, to help identify astrategy mostlikely to reachagoal. Another useof decisiontreesisasadescriptivemeansfor calculatingconditionalprobabilities. Indata miningand machinelearning, adecisiontreeisapredictivemodel; thatis, amappingfromobservationsaboutanitemto conclusions about its target value. More descriptivenames for such tree models are classification tree (discrete outcome) or regression tree (continuous outcome). Inthesetreestructures, leavesrepresent classificationsandbranchesrepresentconjunctionsof featuresthatleadtothoseclassifications. Themachine learning techniqueforinducinga decision tree from dataiscalleddecisiontreelearning, or(colloquially) decisiontrees. Indecisionanalysis, a"decisiontree"— andtheclosely-relatedinfluencediagram—

isusedasa visualandanalyticaldecisionsupporttool, wherethe expected values(or expected utility) of competing alternativesarecalculated. AdecisionTreeconsistsof3 typesofnodes: -

1. Decisionnodes-commonlyrepresentedby squares.
2. Chancenodes-representedbycircles.
3. Endnodes-representedbytriangles.

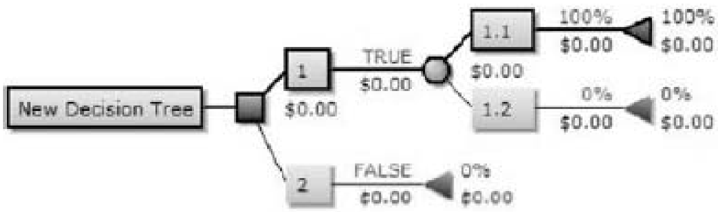

figure1. 2

Drawnfromlefttoright, adecisiontreehasonly burst nodes(splittingpaths)butnosink nodes(converging paths).

Influencediagram

Adecisiontreecanberepresentedmorecompactlyas aninfluencediagram, focusingattentionontheissues andrelationshipsbetweenevents.

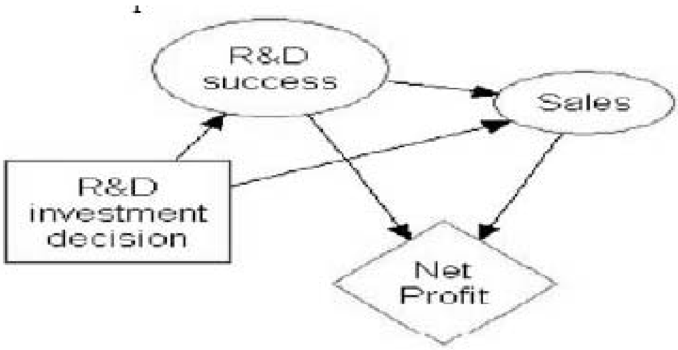

figure1. 3 influencediagram

Creationofdecisionnodes

Three popularrulesareappliedin theautomaticcreationof classificationtrees. TheGinirulesplitsoff a singlegroupof aslargeasizeaspossible, whereasthe entropy andtowing rulesfindmultiplegroups comprising asclose tohalfthesamples aspossible. Bothalgorithmsproceedrecursively downthetreeuntil stoppingcriteriaaremet.

Amongst decisionsupporttools, decision trees(and influencediagrams)haveseveraladvantages:

- Decision trees are simple to understand and interpret.
- Peopleareabletounderstanddecisiontreemodels afterabriefexplanation.
- Have valueevenwithlittleharddata. Important insights can be generated based on experts describingasituation(itsalternatives, probabilities, andcosts)andtheirpreferencesforoutcomes.
- Use awhite box model. If a given result is provided by a model, the explanation for the result iseasilyreplicatedbysimple math. Canbe combined withotherdecisiontechniques. The followingexample usesNetPresentValue calculations, PERT3-pointestimations(decision #1) and a linear distribution of expected outcomes(decision#2):

b. Neuralnetwork

NeuralNetworks[5]areanalytictechniquesmodeled afterthe(hypothesized)processesof learninginthe cognitivesystemandthe neurologicalfunctionsofthe brainandcapable of predictingnew observations(on specificvariables)fromotherobservations(onthe same orothervariables)afterexecutingaprocessofso-called learningfromexistingdata. NeuralNetworksisone of theDataMiningtechniques. Thefirststepistodesigna specific networkarchitecture(thatincludesaspecific numberof"layers"eachconsistingofacertainnumber of"neurons"). Thesize andstructureof thenetwork needstomatchthenature(e. g., theformalcomplexity) of theinvestigatedphenomenon. Becausethelatteris obviously notknownverywellatthisearlystage, this taskisnoteasy andofteninvolvesmultiple"trialsand errors. " (Now, there is, however, neural network softwarethat appliesartificialintelligencetechniquesto aid inthattedioustaskandfinds"thebest"network architecture). The newnetworkisthensubjectedtothe processof"training. "Inthatphase, neurons apply an iterativeprocesstothenumberof inputs(variables)to adjusttheweightsof thenetworkinordertooptimally predict(intraditionaltermsonecouldsay, finda"fit" to) the sample data on which the "training" is performed. Afterthephaseof learningfromanexisting dataset, thenewnetworkisreadyanditcanthenbe usedto generatepredictions. Neuralnetworkhadbeen used torefer to a network or circuit of biological neurons. The modernusageof thetermoftenrefersto artificialneuralnetworks, whichare composedof artificial neurons or nodes. Thus theterm has two distinctusages: Biologicalneuralnetworksaremade up of realbiologicalneuronsthatareconnectedor functionally relatedintheperipheral nervoussystemor thecentralnervoussystem. Inthefieldof neuroscience, they areoftenidentifiedasgroupsofneuronsthat performaspecific physiologicalfunctioninlaboratory analysis. Artificialneuralnetworksare made upof interconnectingartificial neurons (programming constructs thatmimicthepropertiesofbiological neurons). Artificialneuralnetworksmayeitherbeused togainanunderstandingofbiologicalneuralnetworks, or forsolvingartificialintelligenceproblemswithout necessarilycreatingamodelofa realbiologicalsystem. Thereal, biologicalnervoussystemishighly complex andincludessomefeaturesthatmay seemsuperfluous basedonanunderstandingof artificialnetworks. In generalabiologicalneuralnetworkiscomposedof a grouporgroupsofchemicallyconnectedorfunctionally associatedneurons. Asingleneuronmaybeconnected to manyother neuronsandthetotalnumberof neurons andconnectionsinanetworkmay beextensive. Connections, calledsynapses, are usuallyformedfrom axonsto dendrites, thoughdendrodendriticmicrocircuits [1]andotherconnectionsare possible. Apart fromtheelectricalsignaling, there are other formsof signalingthatarisefromneurotransmitter diffusion, whichhaveaneffectonelectricalsignaling. Assuch, neural networksareextremely complex. Artificialintelligenceandcognitivemodelingtry to simulatesomepropertiesof neuralnetworks. While similarintheirtechniques, theformerhas theaimof solvingparticulartasks, while the latteraimstobuild mathematicalmodelsofbiologicalneuralsystems. In the artificialintelligencefield, artificialneural networks havebeenappliedsuccessfully tospeechrecognition, image analysis and adaptivecontrol, in order to construct software agents (in computer and video games) or autonomousrobots. Most ofthecurrently employed artificialneural networksforartificial intelligenceare basedonstatisticalestimation, optimization and control theory. The cognitivemodellingfield involvesthephysicalormathematical modelingof thebehaviourofneuralsystems;ranging from the individual neural level (e. g. modellingthespike responsecurvesofneuronstoastimulus), through the neuralclusterlevel(e. g. modellingthe releaseand effectsofdopamineinthebasalganglia)to the completeorganism(e. g. behavioralmodelingofthe organism'sresponseto stimuli).

Applications:

Theutility ofartificial neuralnetworkmodelsliesinthe factthatthey canbe usedto inferafunctionfrom observations and also to use it. This is particularly usefulinapplicationswherethecomplexityofthedata ortaskmakesthedesignofsuchafunctionby hand impractical.

Real lifeapplications

The taskstowhichartificial neuralnetworksareapplied tendtofallwithin thefollowingbroadcategories: Function approximation, or regression analysis, including time series prediction and modelling. Classification, including pattern and sequence recognition, noveltydetection andsequentialdecision making.

Data processing, includingfiltering, clustering, blind signalseparationand compression. Applicationareas include system identification and control (vehicle control, processcontrol), game-playingand decision making (backgammon, chess, racing), pattern recognition(radarsystems, face identification, object recognition, etc.), sequence recognition (gesture, speech, handwritten text recognition), medical diagnosis, financial applications, data mining(or knowledge discovery in databases, "KDD"), visualizationande-mailspamfiltering.

Clustering:

“Theprocessoforganizingobjectsintogroupswhose membersaresimilarinsomeway”. Clusteringisadata mining(machine learning)techniqueusedtoplacedata elements into related groups without advance knowledgeof thegroupdefinitions. Popularclustering techniquesincludek-meansclusteringand expectation maximization(EM)clustering.

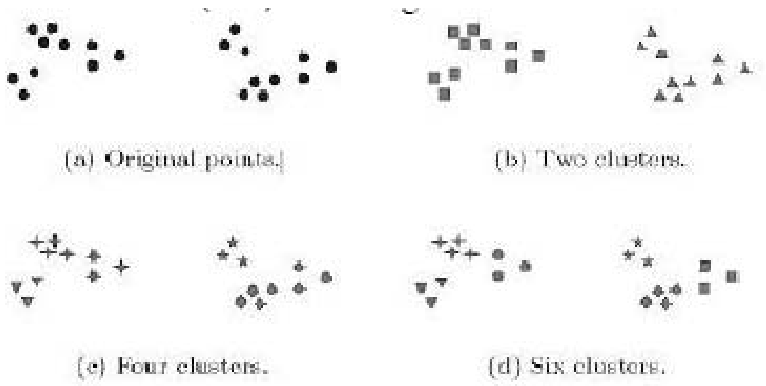

Figure4: clustering

Inthiscaseweeasilyidentifythe4clustersintowhich the data can be divided; the similarity criterion is distance: twoormoreobjectsbelongtothesamecluster ifthey are“close”accordingtoagivendistance(inthis case geometrical distance). This iscalled distance- basedclustering. Clusteringistheassignmentofobjects intogroups(calledclusters)sothatobjectsfromthe sameclusteraremoresimilartoeachotherthanobjects fromdifferentclusters. Setoflikeelements. Elementsfrom different clusters are notalike. Thedistance between pointsin a cluster isless than thedistance betweenapoint inthe clusterandanypoint outsideit. Problemsoccurringinclusteringare:

- Outlinehandlingisdifficult, theelementsdonot naturallylieintoanycluster.
- Dynamicdatainthedatabasemeansthatcluster membership maychangeovertime.
- Interpretingthesemanticmeaningofeachcluster canbedifficult.
- There isnot a single answer to a clustering problem.

Anotherissueiswhat datashouldbeusedforclustering Wecanthensummarizesome basicfeaturesof clustering:

- Thenumberofclusters isnotknown.
- Theremay not be any periorknowledge concerningtheclusters.
- Clusterresultsaredynamic.

Different Typesofclusters

Clustering aims tofind useful groups of objects (clusters), whereusefulnessisdefinedbythegoalsof the dataanalysis. Notsurprisingly, thereare severalsdifferent notionsof a cluster that prove useful in practice.

Well-separated Aclusterisasetofobjectsinwhich eachobjectiscloser(ormoresimilar) to everyother objects in the cluster than to any object not in the cluster. Sometimesathresholdisusedto specifythatall the objectsinaclustermustbesufficientlyclose(or similar)toone another.

thisidealisticdefinitionofa clusterissatisfiedonly whenthedatacontainsnatural clustersthat arequitefarfromeachother. The distance betweenany twopointswithinagroup. well-separatedclusters donotneed tobeglobular, but can beany shape

Figure5: well–separatedclusters

prototype-BasedAclusterisasetofobjectsinwhich eachobjectiscloser(more similar)totheprototype thatdefinetheclusterthantotheprototypeof anyother cluster. fordata with continuous attributes, the prototypeof aclusteisoftenacentroid, i. e, theaverage (mean)of allthepointsinthecluster. whenacentroidis notmeaningful, suchaswhenthe datahascategorical attributes, the prototype isoftenamedoid, i. e., the most representativepointofacluster. Formany typesofdata, theprototypecanberegarded asthemost centralpoint, andinsuchinstances, wecommonly refertprototype-basedclusters ascenter-basedclusters.

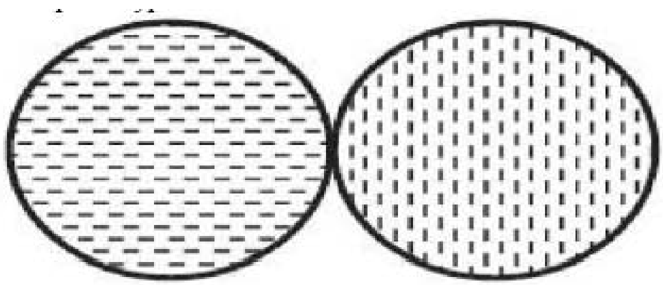

Figure6: prototypebasedclusters

Graph–basedifthedata isrepresentsasagraph, where thenodesareobjectsandthelinks represents connections amongsobjects then a cluster can be defined as a connected component, i. e., a graph of objectsthatareconnectedto oneanother, butthathave no connection to objects outside the group. An importantexampleof graph–basedclustersare contiguity-basedclusters, where twoobjectsare connectedonlyiftheyarewithinaspecifieddistanceof eachother. this impliesthateachobjectinacontiguity- based cluster is closer to some other objects in the clusterthantoanypointinadifferentcluster.

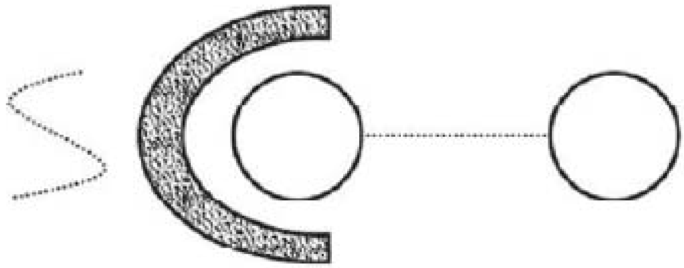

Figure7: Graph-basedclusters

Density–basedAclusterisadenseregionofobjects thatissurroundedbyaregionoflowdensity. Adensity baseddefinitionof aclusterisofenemployedwhenthe clusterare irregularorintertwined, andwhennoise and outliersarepresent.

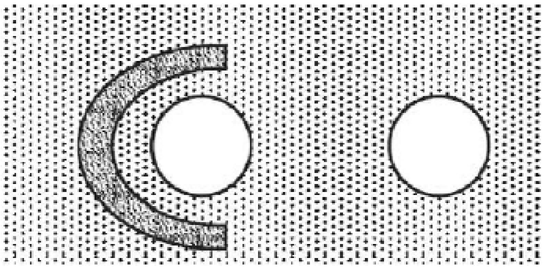

Figure8: Density-basedclusters

Shared- property(conceptual clusters) More generally, wecandefineaclusterasasetofobjectsthat sharesomeproperty. this definitionencompassesallthe previousdefinitionof acluster, e. g., objectsinacenter-basedclustersharethepropertythattheyareallclosest tothesamecentroidormedoid.

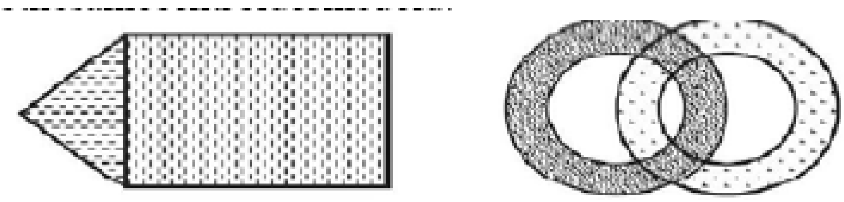

figure 9: shared-propertybasedclusters

Noise

Clustering isoftenconstructedonnoise-freedatasets. In real-worldapplications, itis inevitablethatthedatasets contain noises, which may result in unsatisfactory resultsof theclustering algorithms. Outliersaresample points withvalues much different from thoseofthe remainingsetof data. Outliersrepresenterrorinthedata orcouldbe correctdatavaluethataresimply muchdifferentfrom the remainingdata. A person3. 0meterstallis exceptionallytallthisvalueprobablywouldbeviewed as anoutlier. Some clusteringtechniques do not performwellwiththe presenceofoutliers. clusteringalgorithmsmayactuallysearchesandremoveoutliersto ensurethatthey performbetter. However, caremustbe takeninacctuallyremovingoutliers. forexample, supposethatthedataminingproblemisto predict flooding. Extremelyhighwaterlevelvaluesoccur infrequently, andwhencompared thenormal water levelvalues may seem to be outliers. However, removingthesevaluesmay notallowthedatamining algorithmtoworkeffectivelybecausetherewouldbe nodatathatshowedthat floods everactuallyoccurred.

Conclusion

This paper discusses about the performance of clusteringtechniqueinthepresenceofnoise. Noisecan appearinmany realworlddatasetsandheavilycorrupt thedatastructure. Theperformanceofmany existing algorithmsisdegradedbythepresenceofnoise.

Reference

[1] http: //sivra. in/en/datamining. html
[2] Varun Kumar, NishaRathee, ” Knowledge discovery from database Using anintegration of clustering andclassification” (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 3, March 2011
[3] http: //sivra. in/en/datamining. html
[4] http: //vserver1. cscs. lsa. umich. edu/~spage/ONLINECOURSE/R4Decision. pdf
[5] Raghavendra B. K., S. K. Srivatsa, ” Evaluation of logistic Regression and Neural Network ModelWith Sensitivity Analysis on Medical Datasets” International Journal of Computer Science and Security (IJCSS), Volume (5): Issue (5): 2011International Journal of Computer Science and Security (IJCSS), Volume (5): Issue (5): 2011