GaussianNoiseMultiplicativePrivacyForData PerturbationUnderMultiLevelTrust

RANJEET KUMARRAI*

Research Scholar, Department OfComputer Science & Engg., MUIT Lucknow. U.P.

DR.MANISHVARSHNEY Prof.**

Department OfComputer Science & Engg., School OfEngineering&Technology, MUIT, Lucknow, U.P.

Abstract

Dataminingistheprocessofexploringandanalyzinglargeblocksofinformationinordertouncovermeaningfulpatternsandtrends.Perturbationisamechanismthathasbeenintroducedinthefieldsofcelestialmechanicsandmathematicalphysics.Eachattributehasaweightassociatedwithit,whichindicateshowaccurateandcompleteitis.Databaseanddatasecurityadministratorsareforcedtoperformadifficultbalancingactwhenitcomestograntingemployeesaccesstoorganizationaldata.Forthistousemultiplicativedataperturbationinconjunctionwithsinglelevelandmultilayertrustthegeometrictypeofmultiplicativedataperturbationwillbecarriedoutinthismethod,aswell.Whengeneratingtheperturbedcopy,geometricperturbationinvolvestheorthonormalmatrix,translationalmatrix,andarandomgeneratedGaussiannoisevector,amongotherthings.Inthebeginning,theorthonormalmatrixwillbeusedtoperformtherotationperturbation,andthenthetranslationalmatrixandGaussiannoisecomponentswillbeaddedtoitforthefinalperturbedcopy.Wecansaythatundersingleleveltrust,additiveGaussiandataperturbationproducesperturbedcopiesusinguniformGaussiannoise.Regardlessoftheirtrustratings,alldataminersreceivethesameperturbedcopy.AdditiveGaussiandataperturbationatmulti-leveltrustisstudiedfordataminersatvarioustrustlevels.Dataperturbationisapopularrandomizationapproachthatensures bothaccurate data mining results and privacy.

Keywords: ***Gaussian,noise, data perturbation,multi-level,trust,*** **etc.**

DOI Number: 10.14704/nq.2022.20.12.NQ77008 **NeuroQuantology 2022; 20(12): 79-83**

1. INTRODUCTION

Inthebusinessworld,dataminingisaprocessby whichrawdataistransformedintousefulinformation.Businessescanlearnmoreabouttheircustomersthroughtheuseofsoftwarethatsearchesfor patternsinlargeamountsofdata.Thisallowsthemtodevelopmoreeffectivemarketingstrategies,increasesales,andreducecosts.Effectivedatacollection,warehousing,andcomputerprocessingarerequiredfordataminingtobesuccessful.Dataminingistheprocessofexploringandanalyzing largeblocksofinformationinordertouncovermeaningfulpatternsandtrends.Databasemarketing,creditriskmanagement,frauddetection,spam Emailfilteringandevendeterminingthesentimentoropinionofusersareallpossibleapplicationsforthistechnology.Itispossibletobreakdownthedataminingprocessintofivesteps.Datacollectinformationandloaditintoinformationwarehousesinthefirststageofthedatacollectionprocess. Following that, theyeither storeor managethedataon their own servers or in thecloud.

Businessanalysts,managementteams,andinformationtechnologyprofessionalsallhaveaccesst o the dataand can decide how they want toorganize and presentit totheir clients.

eISSN 1303-5150

- **Perturbation:**

Perturbationisamechanismthathasbeenintroducedinthefieldsofcelestialmechanicsandmathematical physics.Each attribute has a weight associated with it,which indicates howaccurateandcompleteitis.Everyconstraintinvolvingthisattributeisassociatedwithaweightthatrepresentstheimportance of theviolation of thatconstraint. The higherthelevel of trustinadataminer,thelessperturbedacopyofthedataitisallowedtoaccess.Inthisscenario,amaliciousdataminermaygainaccesstovariouslyperturbedcopiesofthesamedatathroughavarietyofmeans,andheorshemaycombinethesediversecopiestojointlyinferadditionalinformationabouttheoriginaldatathatthedataownerdoesnotintendtomakeavailabletothepublic.TheprimarychallengeofprovidingMLT-PPDMservicesispreventingsuchdiversityattacksfromoccurring.Toaccuratelyestimatethedistributionoforiginaldatavalues,anovelreconstructionprocedurehasbeendeveloped.Itispossibletoconstructclassifierswithaccuracythatiscomparabletoorbetterthanthatofclassifiersconstructedwiththeoriginaldatabyutilisingthesereconstructeddistributions.Asaresult,whencomparedtoothertechniques,perturbation mechanisms are themost suitable formaintaining privacy.

1.1 Protecting Data through ‘Perturbation’ Techniques

Organizationstodaycollectmassiveamountsofdataabouttheircustomers,competitors,supply chainpartners,andinternalprocesses.Itisaconstantstrugglefororganizationstomakefulluseoftheirdata,anduncovering"unknown"bitsofknowledgewithintheirmassivedatastorescontinuestobeahighlysoughtaftergoal.Databaseanddatasecurityadministratorsareforcedtoperformadifficultbalancingactwhenitcomestogranting employeesaccesstoorganizationaldata.Sophisticatedorganizationsthatdomakeuseofdataminingandknowledgediscoveryalgorithms(e.g., inductivelearningalgorithms,neuralnetworks, etc.)todiscoverpreviouslyunknown‘patterns'intheirdatabenefitgreatlyfromhavingaccesstolargedatastorescontainingindividualrecords.Anotherimportantissuethatthedatabaseadministratormustdealwithistherequirementtoprotect individual'confidential'dataelementsinanorganisationaldatabasefrombeingimproperlydisclosedbyotherparties.Thescopeofthisprotectionincludesnotonlytraditionaldataaccessissues(e.g.,hackersandillegalentry),butalsomaskingindividualconfidentialrecordattributesinordertopreventindividualrecordsfrombeing identified even by authorized users.

2. LITERATURE REVIEW

Sulekh, V. Jane Varamani(2018) Datamining hasbeenexploredand implementedextensivelyinavarietyofdisciplines,includingtheInternetofThings(IoT),themedicalindustry,andcommercialdevelopment.However,duetoprivacyviolationsandincreased sensitiveinformationsharing,thesedatamining approachesconfrontmajorhurdles.PrivacyPreservingDataMining(PPDM)isasubsetofdata miningthat triestoprotectindividuals'personal informationfromunwantedorillegalpublication.Privacyissuesinfringeonaperson'srighttoprivacy and cause the study participant to lose dignity. It would also causesocialembarrassment,shame,anddishonour,aswellasharmtoone'ssocialandeconomicstanding.Severaldataminingmethodsthatcombineprivacy-preservingstrategiestohidesensitiveitem setsorpatternshavebeendevelopedinrecentyears.Akeyquestionhereiswhichprivacy-preservingstrategyprovidesthebestsecurityforsensitivedata.It'salsocrucialtocheckthequalityoftheoutcomeaswellasthealgorithm'sperformanceafterusingprivacy-preservingstrategies.Weexaminevarious noise-basedPrivacyPreservationstrategies in this research.

ThangarevathiS(2017) Dataminingistheprocessofextractingusefulinformationfromenormousamountsofdata.Intoday'sworld,dataminingisdoneondynamicdataratherthanstaticdata.Theprivacyofcrucialandsensitivedataisamajor considerationintoday'sDataMiningapproaches.Theprivacyofessentialdataisprotectedusing avarietyofways.DataPerturbationisanimportantstrategyformaintainingdataprivacy.Dataperturbationisadatasecurityapproachthatinvolvesalteringadatabaseinordertomaintainprivacyandsecrecy.It'susedfordataprivacyaswellasaccuracy.Wewillexaminethenumerousperturbationstrategiesthatcanbeusedtoprotectdataprivacyinthispaper,aswellastheramificationsofthet

eISSN 1303-5150

echniques.

Srijayanthi Srijayanthi Srijayanthi SrijayanthiS(2017)inrecentyears,oneofthekeyconcernsforminingmeaningfulinformationfromsensitivedatahasbeentheprivacyprotectionoflargescaledatasetsinbigdataapplicationssuchasphysical,biological,andbiomedicalsciences.Intermsofdataanalysis,validation,andpublishing,dataminingprivacyhasbecomeanabsoluterequirementforcommunicatingconfidentialinformation.Privacy-PreservingDataMining(PPDM)assistsinthemininggofinformationandthediscoveryofpatternsfromhugedatasetswhileprotectingprivateandsensitivedata.Numerousprivacypreservationapproacheshavebeendevelopedasaresultoftheadvancementofvarioustechnologiesindatagathering,storage,andprocessing.Wepresentareviewofthemostup-to-dateprivacypreservationapproaches in this study.

A.T.RaviandS.Chitra(2015) Datacollectionandmonitoringusingdataminingforsecurityandbusiness-relatedapplicationshassparkedasurgeinprivacy concerns.PPDM(PrivacyPreservingDataMining)strategiesnecessitatedatamodificationinordertodisinfectthemofsensitiveinformationoranonymizethematareasonablelevelofuncertainty.TheimpactsofK-anonymizationforevaluationmetricsareinvestigatedusingPPDMwithanadultdataset.TheArtificialBeeColony(ABC)algorithmisusedinthisstudyforfeaturegeneralisationandsuppression,whichremovescharacteristicswithoutreducingclassificationaccuracy.Originaldataset generalisation also achieves k-anonymity.

Swapnil Kadam,(2015) Dataperturbation,acommonlyusedandapprovedPrivacyPreservingDataMining(PPDM)method,implicitlyimpliesdataminershaveasinglelevelof trust.Thedifficultyof constructingappropriate modelsregardingaggregateddatawithoutaccesstopreciseinformationororiginalrecordsinindividualdatarecordsisaddressedbyPrivacyPreservingDataMining.Beforedata is published, theperturbation-based PPDMtechniqueprovidesrandomperturbationtoindividualvaluestosafeguarddataprivacy.Previousapproachestothisproblemareunsuitablebecausetheyimplicitlyassumesingle-levelconfidenceindataminers.Weconsiderthisassumptioninthispapertobroadenthescopeofperturbation-basedPPDMtoMultilevelTrust(MLT-PPDM).Themoretrustworthyadatamineris,thelessperturbedacopyofthedataitcanaccess,accordingtoourmethod.Amalevolentdataminercanusenumerousmethodstogainaccesstoseveralperturbedcopiesofthesamedata,andthencombinethesecopiestoinferextrainformationabouttheoriginaldatathatthedataownerdoesnotwanttoshare.ThechallengeofofferingMLT-PPDMservicesispreventingdiversityassaults.Thisproblemissolvedbycorrectlyassigningperturbationacrosscopiesatvariouslevelsoftrust.Intermsofourprivacygoal,wedemonstratethatoursolutioniseffectiveagainstdiversityattacks.Thatis,ourtechniquepreventsdataminerswithaccesstoanycollectionofperturbedcopiesfromrecreatingtheoriginaldatamorepreciselythanthebesteffortusinganyindividualcopyinthecollection.Oursolutionenablesadataownertobuildperturbedcopiesoftheirdataondemand, based on trust levels. Thismethod provides themostflexibilityto data owners.

3. OBJECTIVES OF THE STUDY

- To investigateProtecting Data through‘Perturbation’Techniques.
- To evaluateGaussian noisedataperturbationundermulti-level trust.

4. RESEARCH METHODOLOGY

Thefirstpieceofresearchsuggestedmakesuseo fGaussiannoisetoperturbsensitivedatainbothsinglelevelandmultilayertrustsituations.Inthefirstinstance,additivedataperturbationwill be used to perturbthe data by using Gaussian noise. Under a single degree of trust,Gaussiannoisewillbeintroducedintothesensitivedata,andtheresultingperturbedcopywillbedeliveredevenlytoalldataminers,regardlessoftheirtrustlevels.Differentperturbedcopies willbegenerateddependingonthetrustlevelsofthedataminers,whichwillbeachievedbymultilayertrust.Whenthedataminerwilloperatingatalowertrustlevel,theamountofnoiseintroduced willdisproportionatelymorethanwhenthedataminerwilloperatingatahighertrustlevel.

eISSN 1303-5150

Tousemultiplicativedataperturbationinconjunctionwithsinglelevelandmultilayertrustthegeometrictypeofmultiplicativedataperturbationwillbecarriedoutinthismethod,aswell.Whengeneratingtheperturbedcopy,geometricperturbationinvolvestheorthonormalmatrix,translationalmatrix,andarandomgeneratedGaussiannoisevector,amongotherthings.Inthebeginning,theorthonormalmatrixwillbeusedtoperformtherotationperturbation,andthenthetranslationalmatrixandGaussiannoisecomponentswillbeaddedtoitforthefinalperturbedcopy.

5. RESULTANDDISCUSSION

Gaussian noise for perturbation of data

Gaussiannoiseisastatisticalnoisewithaprobabilitydensityfunctionthatiscomparabletothatofthenormaldistributioninstatistics.Gaussiandistributionisanothernamefornormaldistribution. Equation gives the probabilitydensity function ofGaussian noise (1)

$$gf(x) = \left(\frac{1}{\sqrt{2\pi\sigma}}\right) e^{\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)} \quad (1)$$

Themeanvalueisμandthevarianceisσ^2.Scalarsorvectorscanbeusedforthemeanvalueandvariance.Thelengthofthevariancevectormustbethesameasthelengthofthefirstseedvector.Thecovariancematrixinthissituationisadiagonalmatrixwhosediagonalmembersaredrawnfromthevariancevector.TheoutputGaussianrandomvariablesareuncorrelatedbecausethe off-diagonalelements arezero.

Privacy PreservationEstimation

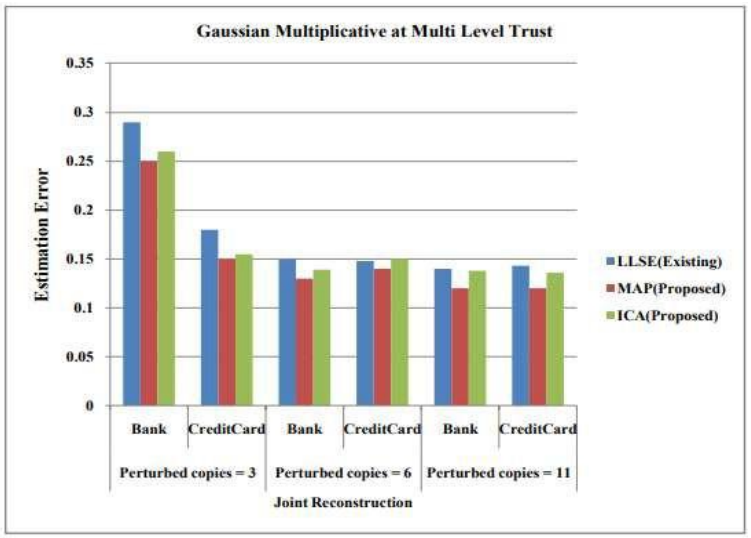
TheoriginaldataisreconstructedusingLinearLeastSquareError(LLSE)based estimation.Thehighertheerrorrateintheoriginaldatareconstruction,themoreprivacywillberetained.Graphical charts are used to show the evaluation outcomes.

ThenoisecomponentσZi^2issupposedto havevaryingvalues for Gaussian noisedataperturbationundermultileveltrust.Theclassifieraccuracyfor Gaussiandata perturbationat multileveltrustisshown inTable1, withthevalues of alltrustlevelsaveraged.The findingsforallthreeclassifier models under multi level trust are shown in Figure 1.

Table 1

Classifier accuracy for Gaussian dataperturbation at MultiLevel Trust

	Bank Dataset			Credit Card Dataset		
ClassifierAccuracy	DecisionTree	Naïve Bayes	kNN	DecisionTree	Naïve Bayes	kNN
OriginalData	90.76	89.24	84.92	77.73	54.20	75.36
GaussianAdditive	90.06	88.81	75.00	75.22	51.90	56.00
GaussianMultiplicative	85.05	84.76	84.21	72.30	50.34	71.00



eISSN 1303-5150

Figure 1: GaussianMultiplicative Privacy measure under Multi Level Trust

Figure1showstheprivacyaccuracyfortheGaussianmultiplicativeapproachbasedonthenormalizedestimationerror.Thenumberofperturbedcopiesrepresentsthenumberofdataminers withvarying levels of trust. Copies=3 denotes the reconstruction ofthe original data fromthreeperturbeddatasets(correspondingly forcopies6and11).Thenormalizedestimationerroroccurswhenattemptingtorecreatetheoriginaldatafromperturbeddata.Iftheestimationerrorisconsiderable,itsuggeststhattheoriginaldatawasnotrebuiltprecisely.Inallnoisefilteringsystems,itisobviousthatthemultiplicativeformofdataperturbationresultsinahighererrorrate whenreconstructingthe originaldata, resulting in higher privacy.

AnotherCreditcarddatasetisusedtotesttheapproach.Whentheestimationerrorishigh,theoriginaldataisreconstructedincorrectly.TheGaussianMultiplicativeperturbationhasagreaterestimationerrorinbothdatasets.Thisdemonstratesthatthisstrategyoutperformstheadditivemethodintermsofprivacyaccuracy.PerturbedcopiesforMlevelsaregeneratedusing anoisecomponentσZi^2takenfromarandomGaussiandistributionwithmultileveltrust.Themalevolentdataminersaresupposedtobeawareofthenoisedistribution,mean,andcovariance oftheoriginalandperturbeddata.Theresultsclearlyrevealthatjointestimationisgrowingforthe GaussianmultiplicativetechniqueGaussianmultiplicativetechnique,demonstratingthatthemultiplicativetechniqueachievestheprivacygoalmoreeffectively.Theestimationerrordoesnotchangemuchwhenthenumberofperturbedcopiesavailabletomaliciousdataminersrises,anditremainssteadyforvariedavailablecopies.IncomparisontoGaussianmultiplicativeperturbation,thisexhibitsanincreased privacy level.

6. CONCLUSION

Wecansaythatundersingleleveltrust,additive Gaussiandataperturbationproducesperturbed copiesusinguniform Gaussiannoise.Regardlessoftheirtrustratings, all dataminersreceive the sameperturbedcopy.AdditiveGaussiandataperturbationatmultileveltrustisstudiedfordataminers at various trustlevels. Data perturbation is a popular randomization approach that ensuresbothaccuratedataminingresultsandprivacy.Theadditiveandmultiplicativetypesofdataperturbationhavebeenusedinpreviousresearch.Theincreasingamountoferrorratedependingonvarioussortsofattacksisusedtoquantifyprivacy.Dataminingtechniquesthatautomaticallytranslatedataintoknowledgemayyieldconfidentialinformationaboutaspecificuser,puttingtheuser'srighttoprivacyatrisk.Theresultsclearlyrevealthatjointestimationisgrowingfortheadditivemethodintermsofprivacyaccuracy.Perturbedcopiesfor Mlevelsaregeneratedusing anoisecomponent σZi^2 takenfromarandomGaussiandistributionwithmultileveltrust.Themalevolentdataminersaresupposedtobeawareofthenoisedistribution,mean,andcovariance oftheoriginalandperturbeddata.Theresultsclearlyrevealthatjointestimationisgrowingforthe Gaussianmultiplicativetechnique,demonstratingthatthemultiplicativetechniqueachievesthe privacygoalmoreeffectively.Theestimationerrordoesnotchangemuchwhenthenumberofperturbedcopiesavailabletomaliciousdataminersrises,anditremainssteadyforvariedavailablecopies.IncomparisontoGaussianmultiplicative perturbation,thisexhibitsanincreased privacy level.

REFERENCES

1. S. Srijayanthi (2017)" A Comprehensive Survey on Privacy Preserving Big Data Mining" International Journal of Computer Applications Technology and Research Volume 6–Issue 2, 79-86, 2017, ISSN:-2319–8656
2. Ravi,A.T.& Chitra,S..(2015),PrivacyPreservingDataMining.ResearchJournalofApplied Sciences, Engineering and Technology.9. 616-621. 10.19026/rjaset.9.1445.
3. Mr.Swapnil Kadam(2015)"PreservingDataMiningthroughDataPerturbation"InternationalJournalofAdvancedResearchinComputerEngineering&Technology(IJARCET)Volume 4 Issue11
4. Luo,Zhifeng &Wen,Congmin.(2014).Achaos-basedmultiplicativeperturbationschemeforprivacypreservingdatamining.ProceedingsoftheIEEEInternationalConferenceonSoftwareEngineeringand Service Sciences,ICSESS.941-944.10.1109/ICSESS.2014.6933720.
5. Fares,Tamer&Khalil,Awad &Mohamed,Bensaada.(2008).PrivacyPreservationinDataMiningusingAdditiveNoise..21stInternationalConferenceonComputerApplicationsinIndustryand Engineering,CAINE 2008. 50-55.

eISSN 1303-5150