Perturbation method For Effective Data Privacy In Data Mining

Bhavesh M. Patil
Anita A. Patil
Sujata R. Patil
Gayatri S. Pawar

SSBT's College of Engineering & Technology,
Bambhori, Jalgaon - 425 001, Maharashtra, India

September 28, 2016

Outline

Outline
Introduction
Literature survey
Proposed System
Architecture
System Design
Algorithms
Advantages

1 Introduction
2 Literature survey
3 Proposed System
4 Architecture
5 System Design
6 Algorithms
7 Advantages
8 Disadvantages
9 Conclusion
10 Reference

Outline

1 Introduction
2 Literature survey
3 Proposed System
4 Architecture
5 System Design
6 Algorithms
7 Advantages
8 Disadvantages
9 Conclusion
10 Reference

Introduction

- Data mining technology has been developed with the goal of providing tools for automatically and intelligently transforming large amount of data in knowledge relevant to users.
- knowledge discovery process, however, can also return sensitive information about individuals, compromising the individuals right to privacy.
- For this reason, several data mining techniques, incorporating privacy protection mechanisms, have been developed based on different approaches.

Outline

Outline
Introduction
Literature survey
Proposed System
Architecture
System Design
Algorithms
Advantages

1 Introduction
2 Literature survey
3 Proposed System
4 Architecture
5 System Design
6 Algorithms
7 Advantages
8 Disadvantages
9 Conclusion
10 Reference

Literature survey

- The typical additive reconstruction technique is column-based additive randomization.
- This type of techniques relies on the facts that column-based value distortion can be applied to reconstruct some sensitive columns.
- Some existing techniques addresses the problem of building a decision tree classifier in which the values of individual records have been perturbed using randomization method.

Outline

Outline
Introduction
Literature survey
Proposed System
Architecture
System Design
Algorithms
Advantages

1 Introduction
2 Literature survey
3 Proposed System
4 Architecture
5 System Design
6 Algorithms
7 Advantages
8 Disadvantages
9 Conclusion
10 Reference

Proposed System

- The proposed approach, aim to achieved better result for privacy on database than existing system.
- Proposed algorithm perturbs the data by normalization, average and multiplying a noise.
- The Proposed approach uses value distortion method for data perturbation.

Outline

1 Introduction
2 Literature survey
3 Proposed System
4 Architecture
5 System Design
6 Algorithms
7 Advantages
8 Disadvantages
9 Conclusion
10 Reference

Architecture

Outline
Introduction
Literature survey
Proposed System
Architecture
System Design
Algorithms
Advantages

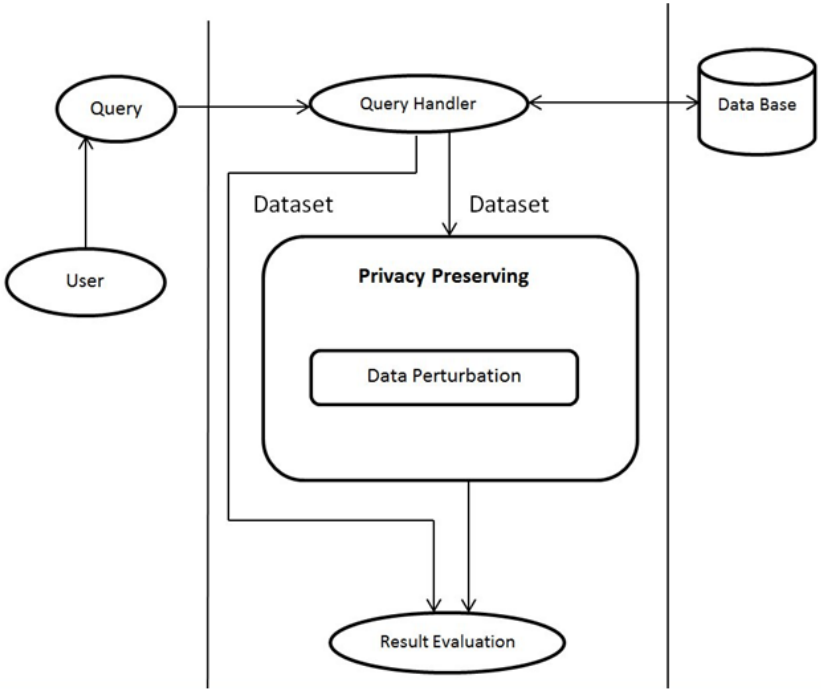


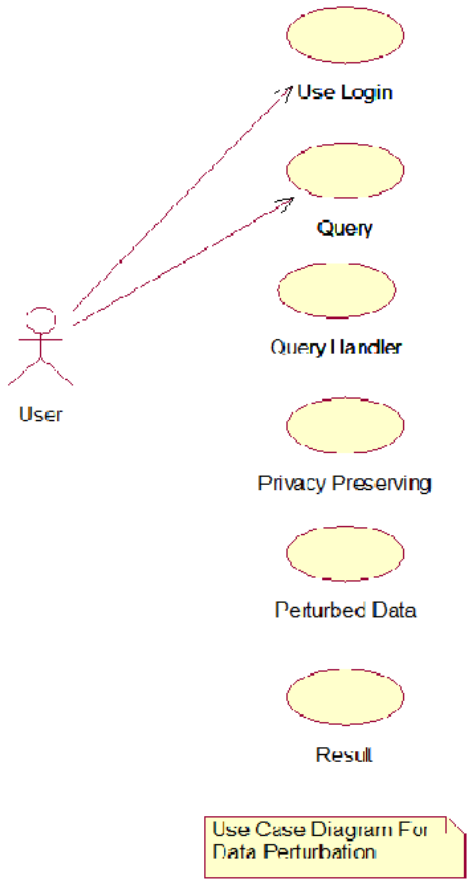
Fig.System architecture

Outline

Outline
Introduction
Literature survey
Proposed System
Architecture
System Design
Algorithms
Advantages

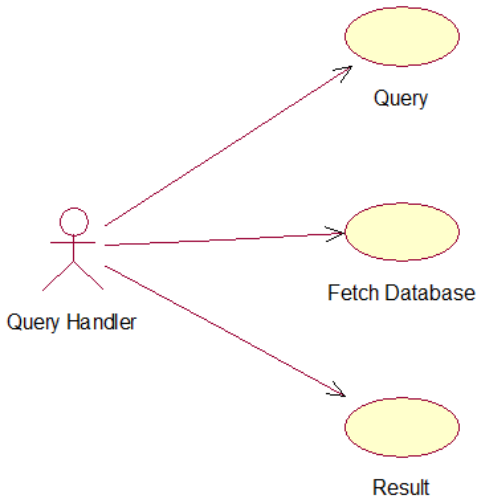
1 Introduction
2 Literature survey
3 Proposed System
4 Architecture
5 System Design
6 Algorithms
7 Advantages
8 Disadvantages
9 Conclusion
10 Reference

System Design



Use Case Diagram

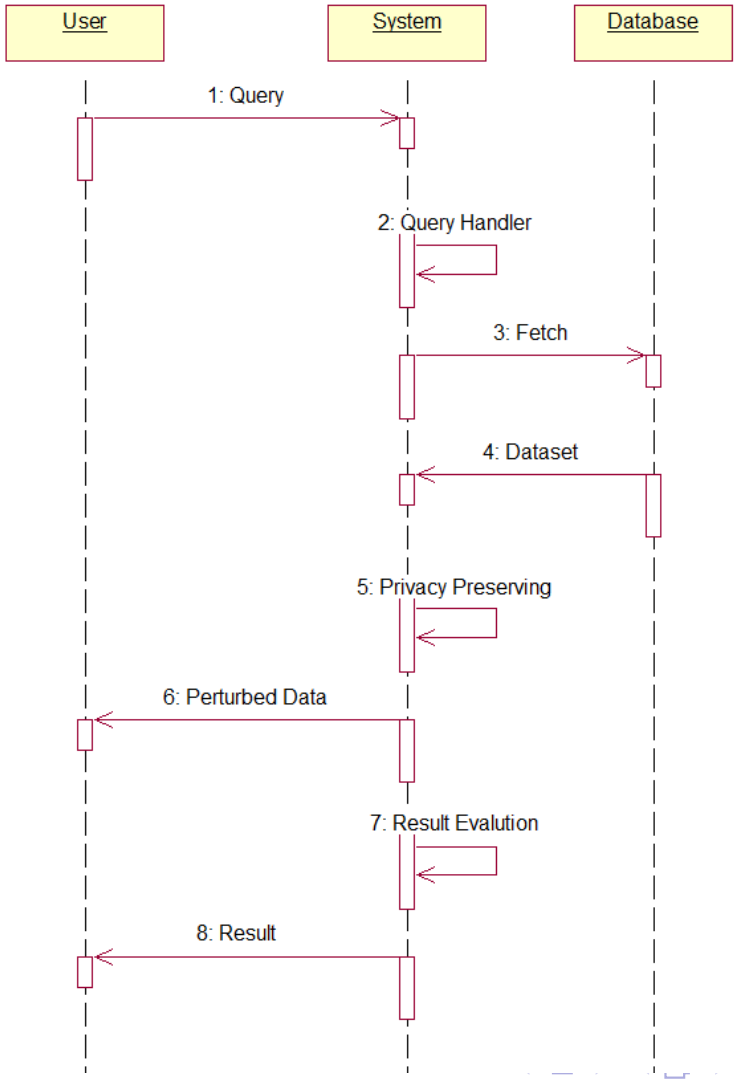
System Design



Use Case Diagram

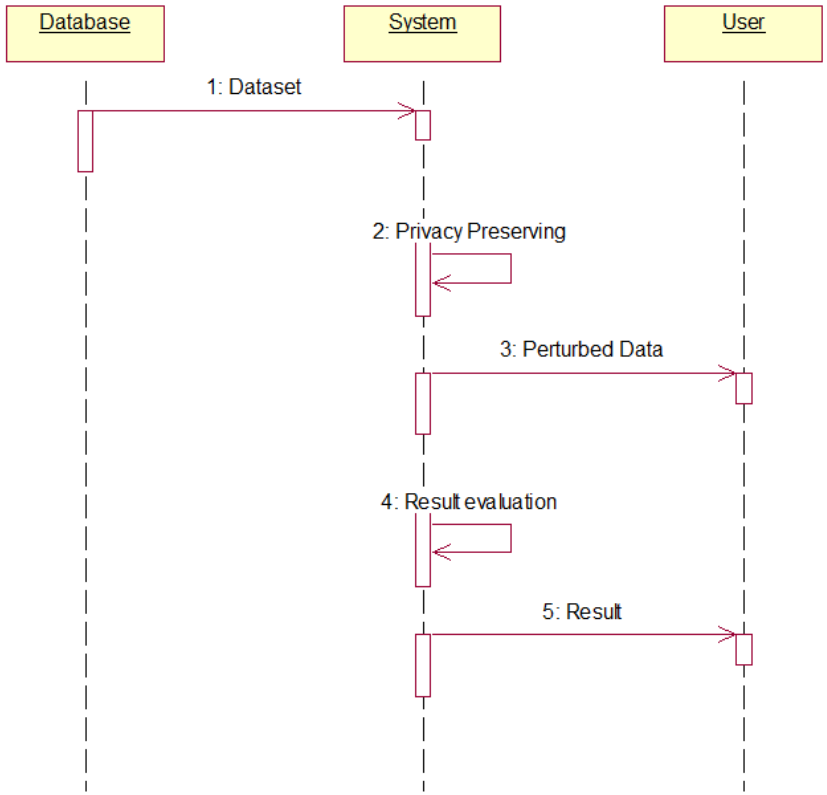
System Design

Outline
Introduction
Literature survey
Proposed System
Architecture
System Design
Algorithms
Advantages



System Design

Outline
Introduction
Literature survey
Proposed System
Architecture
System Design
Algorithms
Advantages



Sequence diagram

Outline

1 Introduction
2 Literature survey
3 Proposed System
4 Architecture
5 System Design
6 Algorithms
7 Advantages
8 Disadvantages
9 Conclusion
10 Reference

1.Algorithm

Initialize variable i,j,mean,count=0,n,value
For value of i=1 to 10
Normalized Value=mean-i/i
i=i+Normalized value
end for
count++
Check if (count==1) Then
average=average(i)
For value of j=1 to n
value=i*j
perturbed Data=Value+Sensitive attribute
end for
Check if(count==1) Then
reset i=1
end

Outline

1 Introduction
2 Literature survey
3 Proposed System
4 Architecture
5 System Design
6 Algorithms
7 Advantages
8 Disadvantages
9 Conclusion
10 Reference

Advantages

- Provide summary statistical information without disclosing individuals confidential data.
- Data loss will be minimized.
- The proposed mechanism will give high performances and low error rate compared with existing methods.

Outline

1 Introduction
2 Literature survey
3 Proposed System
4 Architecture
5 System Design
6 Algorithms
7 Advantages
8 Disadvantages
9 Conclusion
10 Reference

Disadvantages

- Time consuming.
- Complex to understand.

Outline
Introduction
Literature survey
Proposed System
Architecture
System Design
Algorithms
Advantages

Outline

1 Introduction
2 Literature survey
3 Proposed System
4 Architecture
5 System Design
6 Algorithms
7 Advantages
8 Disadvantages
9 Conclusion
10 Reference

Conclusion

- Proposed approach focused on data perturbation by randomization noise addition to preserve privacy of sensitive attributes. Proposed approach tried to keep statistical relationship among the tuple attributes intact. We can predict that Results show fairly good level of privacy with reasonable accuracy in almost all test cases. Privacy of original data after applying perturbation has been quantified using misclassification error.

Outline

Outline
Introduction
Literature survey
Proposed System
Architecture
System Design
Algorithms
Advantages

1 Introduction
2 Literature survey
3 Proposed System
4 Architecture
5 System Design
6 Algorithms
7 Advantages
8 Disadvantages
9 Conclusion
10 Reference

References

- Lambodar Jena, Ramkrushna Swain, IEEE,"Comparative study on Privacy Pre- serving Association Rule Mining Algorithm,"International Journal of Internet Computing, Vol.1, 2011.
- Agrawal R and Srikant R, (2000) Privacy-preserving data mining, In Proceeding of the ACM SIGMOD Conference on Management of Data. ACM Press, pp 439-45 International Journal of Data Mining and Knowledge Management Process (IJDKP) Vol.3, No.3, May 2013.
- International Journal on Recent and Innovation Trends in Computing and Communication Volume: 2 Issue: 9 May 2015.

Outline
Introduction
Literature survey
Proposed System
Architecture
System Design
Algorithms
Advantages

Thank You...