International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 04 | Apr -2017 www.irjet.net p-ISSN: 2395-0072

# Analysis of road accidents using data mining techniques

**Poojitha Shetty[1], Sachin P C[2], Supreeth V Kashyap[3], Venkatesh Madi[4]**

[1,2,3,4] *Dept. of Computer science and engineering,National Institute of Engineering, Karnataka, India.*

---

***Abstract:****Globalization has affected many countries. There has been a drastic increase in the economic activities and consumption level, leading to expansion of travel and transportation. The increase in the vehicles, traffic lead to road accidents. Considering the importance of the road safety, government is trying to identify the causes of road accidents to reduce the accidents level. The exponential increase in the accidents data is making it difficult to analyse the constraints causing the road accidents. The paper describes how to mine frequent patterns causing road accidents from collected data set. We find associations among road accidents and predict the type of accidents for existing as wel as for new roads. We make use of association and classification rules to discover the patterns between road accidents and as well as predict road accidents for new roads.*

***Key Words***:Data mining, Association rule, Classification rule, Apriori algorithm, Naïve Bayes algorithm.

## 1.INTRODUCTION

India has secondlargest road network in the world. Road accidents happen quite frequently and they claim too many lives every year. It is necessary to find the root cause for road accidents in order to avoid them. Suitable data mining approach has to be applied on collected datasets representing occurred road accidents to identify possible hidden relationships and connections between various factors affecting road accidents with fatal consequences. The results obtained from data mining approach can help understand the most significant factors or often repeating patterns. The generated pattern identifies the most dangerous roads in terms of road accidents and necessary measures can be taken to avoid accidents in those roads.

## 2. METHODOLOGY

Descriptive or predictive mining applied on previous road accidents data in combination with other important information as weather, speed limit or road conditions creates an interesting alternative with potentially useful and helpful outcome for all involved stakeholders.

Association rule mining is used to analyse the previous data and obtain the patterns between road accidents. The two criterion used for association ule mining are support and confidence. Apriori algorithm is one of the techniques to implement association rule mining. In the proposed system, we use apriori algorithm to predict the patterns of road accidents by analyzing previous road accidents data.

The steps for the apriori algorithm:

- Scan the data set and find the support(s) of each item.
- Generate L1 (Frequent one item set).Use Lk-1, join Lk-1 to generate the set of candidate k - item set.
- Scan the candidate k item set and generate the support of each candidate k – item set.
- Add to frequent item set, until C=Null Set.
- For each item in the frequent item set generate all non empty subsets.
- For each non empty subset determine the confidence. If confidence is greater than or equal to this specified confidence .Then add to Strong Association Rule.

  INPUT DATASET ( A,B,C,D and E are accident types):

| TID | Items |
|---|---|
| 1 | A,C,D |
| 2 | A,C,E |
| 3 | A,B,C,E |
| 4 | B,E |

**Minimum Support = 50%**

**Minimum Confidence = 80%**

Item set: A, B, C, D, and E

**STRONG ASSOCIATION RULE:**

This is the result obtained.

1. {B}->{E}
2. {CE}-> {A}
3. {AE}->{C}
4. {A}-> {C}
5. {C}->{A}

Classification rule is used to predict road accidents for new roads. In the poposed system, Naive Bayes algorithm is used to implement classification rule.

Naïve Bayes algorithm steps:

**Step 1:** Scan the dataset (storage servers)

**Step 2:** Calculate the probability of each attribute value. [n, n_c, m, p]

**Step 3:** Apply the formulae

P(attributevalue(ai)/subjectvaluevj)=(n_c + mp)/(n+m)

*Where:*

n = the number of training examples for which v = vj

- nc = number of examples for which v = vj and a = ai
- p = a priori estimate for P(aijvj)
- m = the equivalent sample size

**Step 4:** Multiply the probabilities by p

**Step 5:** Compare the values and classify the attribute values to one of the predefined set of class.

**Sample Example:**

Attributes(Constraints) – SpeedLimit, Wheather, PedistrianDistance [m=3]

Subject (Accident Type) – A1, A2 [p=1/2=0.5]

**Training Dataset**

| Road | SpeedLimit(X,Y,Z) | Wheather(A,B,C) | Pedistrian Distance(P,Q,R) | Accident Type |
|---|---|---|---|---|
| Road1 | X | A | P | A1 |
| Road2 | X | B | Q | A1 |
| Road3 | Y | B | P | A2 |
| Road4 | Z | A | R | A1 |
| Road5 | Z | C | R | A2 |

**New Road6 Features – SpeedLimit - X,Wheather – A,PedistrianDistance - R Which Accident Type – A1/A2?**

P=[n_c + (m*p)]/(n+m)

| **A1** | **A2** |
|---|---|
| X<br>P=[n_c + (m*p)]/(n+m)<br>n=2, n_c=2,m=3,p=0.5<br>p=[2+(3*0.5)]/(2+3)<br>p=0.7 | X<br>P=[n_c + (m*p)]/(n+m)<br>n=2, n_c=0,m=3,p=0.5<br>p=[0+(3*0.5)]/(2+3)<br>p=0.3 |
| A<br>P=[n_c + (m*p)]/(n+m)<br>n=2, n_c=2,m=3,p=0.5<br>p=[2+(3*0.5)]/(2+3)<br>p=0.7 | A<br>P=[n_c + (m*p)]/(n+m)<br>n=2, n_c=2,m=3,p=0.5<br>p=[2+(3*0.5)]/(2+3)<br>p=0.3 |
| R<br>P=[n_c + (m*p)]/(n+m)<br>n=2, n_c=1,m=3,p=0.5<br>p=[1+(3*0.5)]/(2+3)<br>p=0.5 | R<br>P=[n_c + (m*p)]/(n+m)<br>n=2, n_c=1,m=3,p=0.5<br>p=[1+(3*0.5)]/(2+3)<br>p=0.5 |

**A1** = 0.7 * 0.7 * 0.5 * 0.5 (p)

=0.1225

**A2** = 0.3 * 0.3 * 0.5 * 0.5 (p)

=0.0225

Since **A1**>**A2**

So this new road6 is classified to **A1**

## 3. CONCLUSIONS

Current system is manual where government sector make use of ledger data and analyze the data manually, based on the analysis they will take the precautionary measures to reduce the number of accidents.Proposed system uses road accidents data to mine frequent patterns and important factors causing different types of accidents.It discovers the associations among road accidents using apriori algorithm .It also predicts the common accidents that may cause for new roads with the help of Naïve Bayes algorithm.

## REFERENCES

[1] R. Agrawal, T. ImieliĔski, A. Swami, “Mining Association Rules Between Sets of Items in Large Databases”, Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, ACM, New York, NY, USA, pp. 207–216, 1993.

[2] R. Agrawal, R. Srikant, “Fast Algorithms for Mining Association Rules in Large Data-bases”, Proceedings of the 20th International Conference on Very Large Data Bases, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp 487-499, 1994.

[3] A Araar et al., “Mining road traffic accident data to improve safety in Dubai”, Journal of Theoretical and Applied Information Technology, 47(3), pp. 911-927, 2013.

[4] L. Breiman, “Random Forests”, Machine Learning, Vol. 45, pp. 532, 2001. [