ISSN: 2321-7782 (Online)
Volume 2, Issue 3, March 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Paper / Case Study
Available online at: www.ijarcsms.com

Data prediction: A useful technique

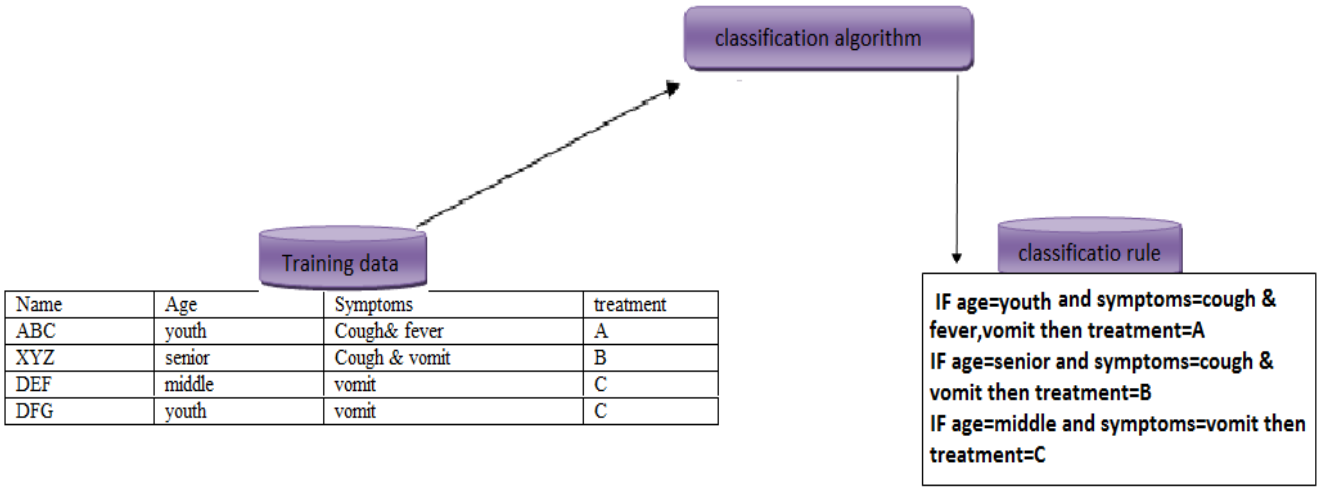
Monika D. Khatri[1]
Dept. of Computer Engineering
Sipna College of Engg. and Technology
Amravati, Maharashtra – India

S. Dhande [2]
Assistant Professor
Dept. of Computer Engineering
Sipna College of Engg. and Technology
Amravati, Maharashtra – India

Abstract: ***Data is a collected part about a particular topic. As a database contains large amount of data that is all data about related topic as well as nearby data about a topic. Example, Google search engine has its own database which stores large amount of data. Whenever we enter a query to search all the data related to that topic appears and then we select data according to our need. That's why it is true that databases are nowadays becoming data rich but information poor [1]. So, to improve the quality of our database and extract information from it a technique called as data mining is used. There are various data mining techniques such as classification, prediction, clustering, association, etc. This paper gives detail review of prediction technique.***

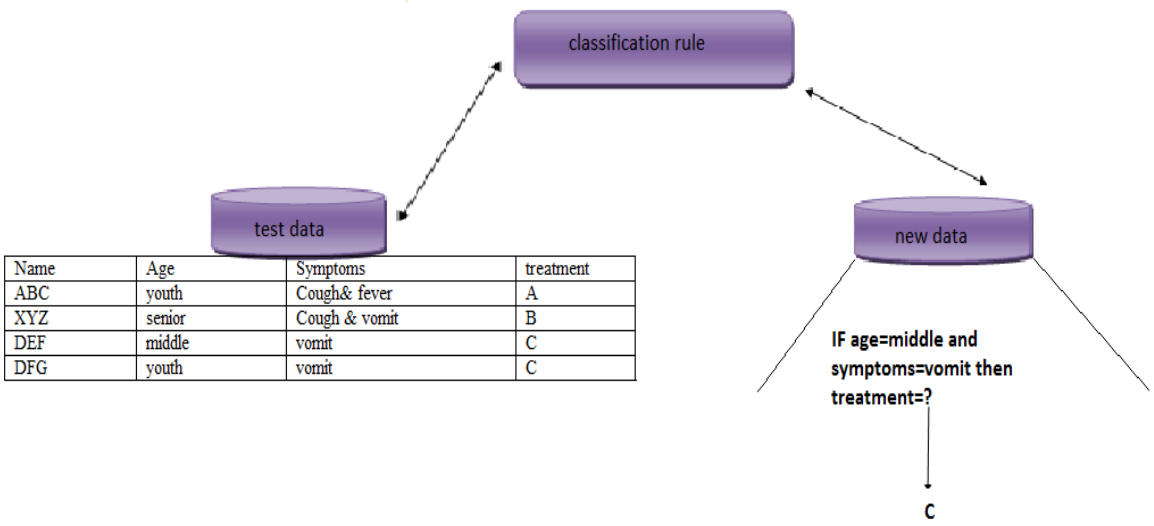
I. INTRODUCTION

Basically, data mining can be performed on two tasks: predictive and descriptive. Predictive data are done on current data to predict future conditions whereas descriptive task gives the general properties of the data. There are two ways by which data analysis can be done to predict the future and they are: Classification and Prediction. Classification is used for categorical data and prediction for numeric values [1].Prediction con also be done on text data. The most famous social media are Twitter and Facebook. They are in top 10 sites in the world according to Alexa ranking [2].This prediction is done by machine because they have lower cost as compared to human prediction [3]. Example, Suppose a Doctor wants to decide the treatment for a patient depending on its current symptoms then the decision will be either treatment 'A' or 'B' or 'C' where the data analysis is categorical data. If the treatment is decided and then according to the patient state of health doses of medicine is decided that is tablet 1=2,tablet 2=1,tablet 3=3where data analysis is numeric data.



Name	Age	Symptoms	treatment
ABC	youth	Cough& fever	A
XYZ	senior	Cough & vomit	B
DEF	middle	vomit	C
DFG	youth	vomit	C

Fig.1. classification



Name	Age	Symptoms	treatment
ABC	youth	Cough& fever	A
XYZ	senior	Cough & vomit	B
DEF	middle	vomit	C
DFG	youth	vomit	C

Fig.2 prediction

II. METHOD OF PREDICTION

A. Steps to build model

Generally predictive models are developed by statistical specialist and contain following steps

1) Selecting input variables: Choice of input variables depends on:
   - Data available
   - Factors relevant to task
2) Defining output classes: The model that is being developed should be given by giving a precise definition to each output class.
3) Preparing data: Transformation of input variables into a form that is appropriate for the prediction task is called as preparing data.
4) Building models: By using set of training data for which output class is known we can develop various models that can classify set of examples.
5) Performing model selection: From the set of the model we can select any model based on the accuracy access.
6) Evaluating the selected model on access: Evaluation is must to estimate the models expected accuracy [2].

B. Types of models:

1) Linear regression:

Regression is a procedure for predicting value of a dependant variable from an independent variable. It is the simplest form of regression. People generally use linear regression than nonlinear regression, exponential, logarithmic, polynomial, etc. Ex, data on digg.com the association is upon logarithmic converted data [4]. y as a linear function of x. The equation of linear regression will be:

$Y= b+wx$

Here, y as a linear function of x. It can also be written as:

$Y=w_{0+}w_1x$

These can solve by method of least squares. The formula is:

$W_1=\frac{\sum_{i=1}^{|D|}(xi-x)(yi-y)}{\sum_{i=1}^{|D|}(xi-x)^2}$ .....Equation (1)

Where, D=training set which contains data points of the form (x_1,y_1) to $(x_{|D|}, y_{|D|})$.

$W_0=y-w_1x$ ……Equation(2)

Where, x is mean of $x_1,x_2...x_{|D|}$ and y is mean of $y_1,y_2....y_{|D|}$.

Example:

TABLE I

SALARY DATA

X	Y
3	30
8	57
9	64
13	72
3	36
6	43
11	59
21	90
1	20
16	83

We will calculate the mean of x and y which would be 9.1 and 55.4.Putting this values in equation number 1,we get

$W_1=\frac{(3-9.1)(30-55.4)+(8-9.1)(57-55.4)+\cdots.+(16-9.1)(83-55.4)}{(3-9.1)^2+(8-9.1)^2+\cdots+(16-9.1)^2}=3.5$

Now putting this value in equation 2

$W_0=55.4-(3.5)*(9.1)=23.6$

The equation will be y=23.6+3.5x

From the above equation we can predict that a person having 10 years of experience will have salary of $58,600 [1].

2) Non linear regression:

Regression model is used when the relationship between X and Y is straight line but when the relationship is curved we need to use nonlinear regression [1]. This model creates new variables from the data. If these nonlinear variables are predicted properly than we can convert this nonlinear regression to linear regression.Consider the following equations:

$Y=Ae^{bx}u$.....Equation(3)

Here, y is growing at a constant rate of b.

To convert this into linear form, we need to take logarithmic on both sides of equation(3),we will get

$In(y)=In(Ae^{bx}u)$

By algorithmic rule we will get,

In(y)=In(A)+bX+In(u)….Equation(4)

Then, put In(y)=y,In(A)=a and In(u)=v in equation(4),we will get

Y=a+bX+v.

This is a linear equation.

Second equation is

$Y= Ax^{b}u$…..Equation(5)

Here, the elasticity of y with respect to x is constant.

To convert this into linear form we will take logarithmic to both sides of equation(5),we will get

$In(y)=In(Ax^{b}u)$

Now by algorithmic rule we get

In(y)=In(A)+bX+In(u)…Equation(6)

To make this linear take two variables such as y=In(y) and x=In(x),let a=In(A) and v=In(u) in equation(6)

Y=a+bX+v

Here v is 0to behave as an error because it is acting as an error in linear equation. To do this we have assume the value of u=1.

That is In(u)=In(1)=0…logarithmic rule.

U is never negative or 0.v can be positive or negative because if u<1 than v=In(u)<0 [5].

III. Application of Prediction

A. Web page prediction

When we search something on web we get instant result without any wastage of time .This is because of perfected web pages. If perfecting is not done properly then problems such as web pages are not visited by users, delay in receiving requested web pages in available bandwidth. Therefore, efficient model for perfecting should be used such as Markov model. Web log files contain all the users' activities which can be used by the model to predict the next web page visit [6].

B. Crime prediction

Crime is "an act committed or omitted in violation of a law forbidding or commanding it and for which punishment is imposed upon conviction". Generally, whenever a crime is committed a report of it is taken either on telephone or police themselves visit the crime site which is called as crime report. It contains following things:

- Time, day and date of crime
- Location
- Offence type
- Victim information
- How crime is committed.

This information in criminology to predict reoffending. This was started by "Ernest W. Burgess" in the 1920s .Future events are based on past events that helps to predict the crime[7].

C. Manufacturing

Predictive data are used by industries for decision making. Firstly to predict the future of a product in accompany we must have knowledge about the trends, taste of customers, market, etc. If we talk about car manufacturing company then the company will not directly interact with the customers. They will gather data about the car from vendors, previous year data and the manufacturing capacity of their company and predict the next year manufacturing rate [8]. The linear regression is best way to predict this

D. Movie box-office review using social media

Social media like Facebook has more than 1.11 billion active users[9].Along with the other applications mentioned in this paper social media can also be useful to predict movie box-office[10].Movies being part of entertainment for people, so large data is available of it. IMDB is internet Movie Database on which there is more than 200 films which are USA box-office movies that are released per year. There are more than 100,000 talks for movies [10].It is possible to calculate the opening weekend income and gross by simply IMDB. They both add 25% of total sales [11].Some movies are having low gross predicted and some have high. Sometimes it happens that the prediction appeared to be perfect.[12].Now the people who are interested in movie will definitely post on the site and this indicates that they are going to watch it. Basically there release first week is most important for the movie gross than any another [12].Some researchers also tried to predict Oscar winners [3][13][14].Thus social media can help for movie box-office.

IV. CONCLUSION

In our normal life also we do general predictions like if its cloudy weather then it's going to rain, if I don't do studies than I am going to fail. This is normal prediction but for real world prediction we require some techniques that will help us to predict correct data. Based on this prediction we can accomplish important task such as detection of a disease to cure it as early as possible, to detect the stock prices of company, production of a product, etc. This major decisions need to be accurate otherwise company may fail to progress, etc. So this paper provides the detail information about prediction method.

References

1. Data mining: concepts and techniques second edition,Jiawei Hn,University of lions at Urbana Champaign,Micheline Kamber.
2. Alexa Internet Inc, "Alexa Top 500 Global Sites". http://www.alexa.com/topsites. [Accessed march 2014]
3. E. Bothos, D. Apostolou, and G. Mentzas, "Using Social Media to Predict Future Events with Agent-Based Markets," IEEE Intelligent Systems, vol. 25, no. 6, pp. 50-58, Nov. 2010.
4. G. Szabo and B. a. Huberman, "Predicting the popularity of online content," Communications of the ACM, vol. 53, no. 8, p. 80, Aug. 2010.
5. Non linear regression, ©2006-08 Samuel L. Baker.
6. Web page prediction techniques: A review, International Journal of computer trends and Technology(IJCTT),volume 4 Issue&-July2013, by Sunil Kumar, assistant professor, computer science dept MIT, Moradabad and Ms. Mala Karla, Assistant professor in computer science and engineering dept, NITTTR, Chandigarh
7. Review of current crime prediction technique by visas Grover, Richard Adele and Max Bramer, University of Portsmouth, UK AE solutions.
8. Predicting the future of car manufacturing industry using mining techniques by dr. m. hanuman thappa, department of computer science and application, dayananda sagar college, banglore, india and dept of computer science and application ,banglore university, banglore, india, ACEEE int. j. on Information technology, vol.01,no.02,sep 2011.
9. Facebook, "Statistics". http://www.statisticbrain.com/facebook-statistics/[accessed 2014]
10. S. Asur and B. A. Huberman, "Predicting the future with social media," in Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on, 2010, vol. 1, no. 6, pp. 492–499.
11. J. S. Simonoff and I. R. Sparrow, "Predicting movie grosses: Winners and losers, blockbusters and sleepers," Chance, vol. 13, no. 3, pp. 15–24, May 2000.
12. W. Zhang and S. Skiena, "Improving Movie Gross Prediction through News Analysis," in 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, 2009, vol. 30, no. 2, pp. 301-304.
13. L. Liviu, "Predicting Product Performance with Social Media," Informatics in education, vol. 15, no. 2, pp. 46-56, 2011.
14. Google, " Search trends: a clue to 2011 Oscar winners?" http://googleblog.blogspot.com/2011/02/search-trends-clue-to-2011-oscar.html.

AUTHOR(S) PROFILE

Monika Khatri received B.Tech. degree in Information Technology in 2013 from Government College of Engineering Amravati(An Autonomous Institute of Government of Maharashtra) and now pursuing M.E. from Sipna College of Engineering and Technology, Amravati (under Sant Gadge Baba Amravati University).

