# A Study on Comparison between Normalization Techniques and Best Normalization Technique

Dhruvi Sonani [#1] , Prof. Binita B. Acharya [#2]
BE Computer Engineering (Pursuing) ,
[#] Shree Swami Atmanand Saraswati Institute of technology, Surat – 395006, Gujarat, India.
[1]dhruvisonani7878@gmail.com
Assistant Professor Computer Engineering Dept.
[#] Shree Swami Atmanand Saraswati Institute of technology, Surat – 395006, Gujarat, India.
[2]chahwala.binita09@gmail.com

*Abstract*—As we know Normalization is a pre-processing stage of any problem statement. There is quite a few number of techniques to normalize the data namely Min-Max normalization, Z-score normalization and Decimal scaling normalization. By referring these techniques we are going to differentiate each from another and will decide the best method to normalize dataset.

*Index Terms*—Normalization, Comparison of Normalization technique, transformation

## I. INTRODUCTION

In recent trend to pre-process data, data normalization is mandatory to represent data in some range form. All three methods works efficiently but they all are different from each other in their mathematical area, conceptual difference and representation. This difference is made to ensure which method to use for better performance and boost up the pre-processing. These removes redundancy which is mainly used for application developers as it is difficult to store object in a relation database to maintain same information on more places.

Data normalization in machine learning is known as feature scalling[4]. Main advantage of normalization is to increase consistency, reducing in-consistency of data and easier object-to-data mapping, highly-normalized data scheme closer to conceptually to object-object schemas to prompt high cohesion and loose coupling between class result in similar solution minimum a data point of view.

An attribute is normalized using various techniques such that they fall within a small-specified range. Particularly useful for classification algorithm involving clustering, neural network.

Min-max normalization performs linear transformation on original data. In Z-core normalization, the values for and attribute are normalized based on mean and standard deviation. By decimal Scaling, normalization is performed by moving decimal point moved depends on maximum absolute Value[5].

## II. PROCEDURE FOR PAPER SUBMISSION

### A. Figures

Normalization of ratings means adjusting values measured on different scales to declared prior scale. Normally scale ranges [0,1]. Using data normalization technuques data can be sorted according to relation to eliminate redaudancy as well as establish relationship between entities.[2]
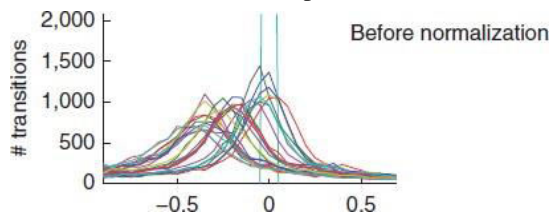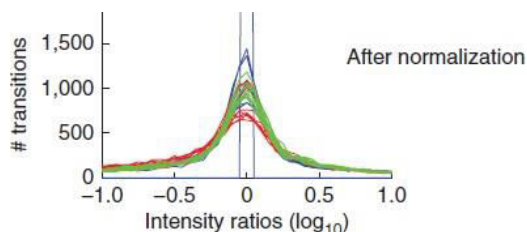

Fig. 1. Dataset before applying Normalization Technique


Fig. 2. Dataset after applying Normalization Technique

## III. UNITS

As we have studied multiple sources to find unit and we found, The goal of normalization is to change the value of numeric columns in the dataset to a common scale, without distorting differences in the ranges of values i.e. normalization of rating means adjusting values measured on different scales to a nationally common scale, here no unit is defined to represent data set value.

## IV. HELPFUL HINTS

### A. References

Reference to the claims are denoted in square brackets with reference number inside. Example in Figures [2] is reference to claimed sentence. Multiple reference can be given as [2]-[3], represents reference 2 and 3.

### B. Abbreviations and Acronyms

SD – standard Deviation
min – minimum
max - maximum
i.e – that is

### C. Equations

Equations play a major role to pre-process data using Normalization. Individual method has separate equation to perform calculation. Diffrentiating each method, Start with **Decimal Scalling** for normalization , it normalizes by moving the decimal point of value of data. In this we divide each value of data by max absolute value of data.

$$Vi' = Vi / 10\,\hat{}\,j$$

Where, Vi – Data value
          Vi'- Normalized data of Vi
          j – Smallest integer such that $max(|Vi'|) < 1$

Next is **min-max normalization**, in this technique of data normalization linear transformations performed on the original data.

$$v' = ((v - min(A)) / max(A) - min(A)) *$$
$$((new\_max(A)-new\_min(A))+new\_min(A))$$

Where, A – attribute data,
      min(A), max(A) - the miimum and maximum
        absolute value of A resp.
      v' – new value of each entry in data.
      v – old value of each entry in data.
      new_max(A) , new_min(A) – max and min value of
        the range resp.

Last is **Z-score normalization ,** values are normalized based on mean and standard deviation of data A.

$$V' = (V - mean\ of\ A) / SD$$

Where, V' – new entry in data
      V – old entry in data

## V. CONCLUSION

As there are some major difference between the methods referring them accuracy and efficiency can be increased and the concluded Decimal Scalling normalization is best to use as it does not use complex mathematical equation and confusing substitution, no higher computing knowledge is required as it has a simplest method among all. The best fit is depend on the nature and aim of study. Best normalization method is to divide physical maximum to compare results in relative frequency term.

## CONFLICT OF INTEREST

We Authors declare no conflict of interest. make the research worthful, Her guidance took me the level to make knowledgeable paper.

## AUTHOR CONTRIBUTIONS

I, Dhruvi Sonani and Guide has given our best efforts to make this paper valuable, effective and efficient. Her constant guidance motivated me to write this paper with complete concentration and faith.

## REFERENCES

[1] S.Gopal Krishna Patro and Kishore Kumar sahu, Normalization: A Preprocessing Stage
[2] Mahadev Asaram Jadhav, Deogiri college, when and why do we need data normalization in data mining algorithm. (**Google reference**)
[3] Alessandro Giuliani, best normalization method.
[4] Håkon Hapnes Strand, Data Scientist, M.Tech Data Science, Indraprastha Institute of Information Technology, Delhi (2017) and Abhinav Sharma, Product Designer at Quora, former ML Engineer at Facebook
[5] Luai Al Shalabi, Zyad Shaaban and Basel Kasasbeh Applied Science

## BIOGRAPHY

**Dhruvi S. Sonani**
Student of BE in Computer Engineering
SSASIT, SURAT – 395006, Gujarat, India.
Interested in Artificial intelligence

**Binita B. Acharya**
Assistant Professor Computer Engineering Dept,
SSASIT, SURAT – 395006, Gujarat, India.
Interested in Operating System, Web technology, Unix Programming, Computer Network