Implementation of Normalization using Min_Max Method

Nang Wai Wai Phyo, Chit Su Hlaing
Computer University (Monywa) Myanmar
Nangcherry74@gmail.com

Abstract

This paper presents a method to normalize logical propositions in numerical data. Type of min_max normalization is emphasized in this study. Data mining is concerned with analyzing large volume of data to automatically discover interesting regularities or relationships. A function obtained by multiple regression analysis in which data are normalized to [0,1]. Therefore, the function represents a non-classical logical proposition and it can be approximated by a Boolean function representing a classical logical proposition This paper presents a method of data transformation. Aim of this paper is to normalized numerical data and reduce the occupied memory space.

1. Introduction

Data mining is the process of extracting hidden patterns from data. As more data is gathered the amount of data doubling every three years data mining becomes an increasingly important tool to transform this data into knowledge. It is commonly used in a wide range of applications such as, marketing, fraud detection and scientific discovery. Data mining can be applied to data sets of any size, and while it can discover hidden patterns, it can not discover patterns which are not already present in the data set.

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

Data mining can do primarily used today by companies with a strong consumer focus retail, financial, communication, and marketing organizations. It enables these companies to determine relationships among "internal" factors such as price, product positioning, or staff skills, and "external" factors such as economic indicators, competition, and customer demographics. And, it enables them to determine the impact on sales, customer satisfaction, and corporate profits. Finally, it enables them to "drill down" into summary information to view detail transactional data. With data mining, a retailer could use point-of-sale records of customer purchases to send targeted promotions based on an individual's purchase history. By mining demographic data from comment or warranty cards, the retailer could develop products and promotions to appeal to specific customer segments. Data mining work are while large-scale information technology has been evolving separate transaction and analytical systems, data mining provides the link between the two. Data mining software analyzes relationships and patterns in stored transaction data based on open-ended user queries.

Data mining consists of five major elements:

1.Extract, transform, and load transaction data onto the data warehouse system.

2. Store and manage the data in a multidimensional database system.

3. Provide data access to business analysts and information technology professionals. Step in the Process of Data Mining are data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation and knowledge presentation.

2. Related work

Data transformation involves smoothing, generalization of the data, attribute construction and normalization. Data mining seeks to discover unrecognized associations between data items in an existing database and Step in the Process of Data Mining.[8]

Min_Max normalization description is a function to apply min-max normalization to a

matrix or data frame. Mmnorm (data, minval=0,maxval=1).Data the dataset to be normalized, including classes: Minval and Maxvalue{the maximum value of the transformed range}.[5]

Min-max normalization subtracts the minimum value of an attribute from each value of the attribute and then divides the difference by the range of the attribute.[2]

The process of normalization is a formal method that identifies relations based on their primary or candidate keys and the functional dependencies among their attributes. [10]

Normalization methods that improve the multimodal system performance. They present three well-known normalization methods, and a new method, which we call adaptive normalization. They denote a raw matching score as s from the set S of all scores for that matcher, and the corresponding normalized score as n. Min-Max (MM): This method maps the raw scores to the [0, 1] range. [7]

3. Theory background

Data transformations are commonly used tools that can serve many functions in quantitative analysis of data. Data transformations are the application of a mathematical modification to the values of a variable. There are a great variety of possible data transformations, from adding constants to multiplying, squaring or rising to a power, converting to logarithmic scales, inverting and reflecting, taking the square root of the values, and even applying trigonometric transformations such as sine wave transformations. Data transformation where data are transformed or consolidated into form appropriate operation.

Data transformation involves smoothing, generalization of the data, attribute construction and normalization. Data mining seeks to discover unrecognized associations between data items in an existing database. It is the process of extracting valid, previously unseen or unknown, comprehensible information from large databases. The growth of the size of data and number of existing databases exceeds the ability of humans to analyze this data, which creates both a need and an opportunity to extract knowledge from databases.

Firstly, performing any transform by hand is prone to error. Even if both source and target languages are well understood by the person doing the transform, any transform of significant size poses a great number of opportunities for error. Conversely computer programs can easily be made to run in a deterministic manner. Therefore, their output is predictable. Another reason for using transformational software is that a type of transformation can be proven to be correct (using formal proofs) for every instance of its application.

Data transformation such as normalization may improve the accuracy and efficiency of mining algorithms involving neural networks, nearest neighbor and clustering classifiers. Such methods provide better results if the data to be analyzed have been normalized, that is, scaled to specific ranges such as [0.0, 1.0]. An attribute is normalized by scaling its values so that they fall within a small-specified range, such as 0.0 to 1.0. Normalization is particularly useful for classification algorithms involving neural networks, or distance measurements such as nearest neighbor classification and clustering. If using the neural network back propagation algorithm for classification mining, normalizing the input values for each attribute measured in the training samples will help speed up the learning phase. For distanced-based methods, normalization helps prevent attributes with initially large ranges from outweighing attributes with initially smaller ranges.

Normalization is a process in database design which groups data into various tables which are then cross linked by a particular field. In a properly normalized database, there will be only one field in one table for a particular piece of data, eliminating possibility for conflicts. Breaking the database up into numerous smaller tables, and eliminating redundancies, eases management and enhances efficiency.

A major goal of normalization is to eliminate redundancy by having a data element represented in only one place. Where the attribute data are scale so as to fill within a small specified range, such as 0.0, 1.0.There are many methods for data normalization includes min-max normalization, z-score normalization and normalization by decimal scaling. So this system is intended to implement the min-max normalization.

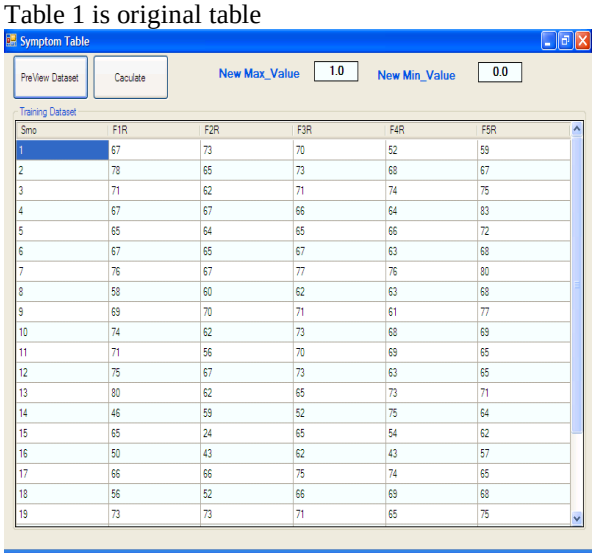
3.1 The purpose of normalization

Normalization is a technique for producing a set of relations with desirable properties, given the data requirements of an enterprise. The process of normalization is a formal method that identifies relations based on their primary or candidate keys and the functional dependencies among their attributes.

3.2Normalization Result

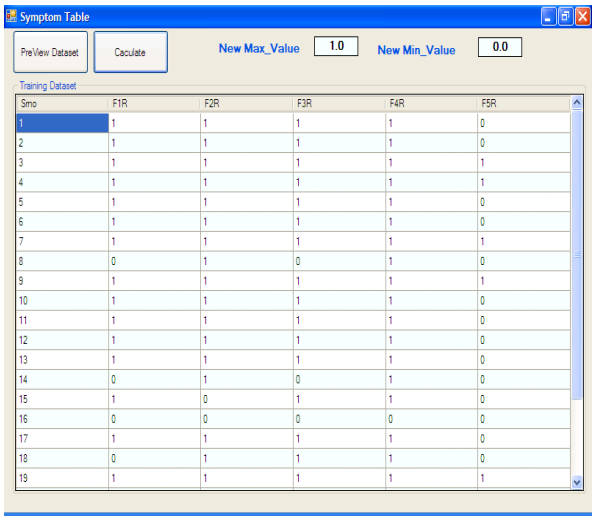
Table 1 is normalized to table 2

Table 1 is original table



Smo	F1R	F2R	F3R	F4R	F5R
1	67	73	70	52	59
2	78	65	73	68	67
3	71	62	71	74	75
4	67	67	66	64	83
5	65	64	65	66	72
6	67	65	67	63	68
7	76	67	77	76	80
8	58	60	62	63	68
9	69	70	71	61	77
10	74	62	73	68	69
11	71	56	70	69	65
12	75	67	73	63	65
13	80	62	65	73	71
14	46	59	52	75	64
15	65	24	65	54	62
16	50	43	62	43	57
17	66	66	75	74	65
18	56	52	66	69	68
19	73	73	71	65	75

Table 2 normalized table



Smo	F1R	F2R	F3R	F4R	F5R
1	1	1	1	1	0
2	1	1	1	1	0
3	1	1	1	1	1
4	1	1	1	1	1
5	1	1	1	1	0
6	1	1	1	1	0
7	1	1	1	1	1
8	0	1	0	1	0
9	1	1	1	1	1
10	1	1	1	1	0
11	1	1	1	1	0
12	1	1	1	1	0
13	1	1	1	1	0
14	0	1	0	1	0
15	1	0	1	1	0
16	0	0	0	0	0
17	1	1	1	1	0
18	0	1	1	1	0
19	1	1	1	1	1

Table 1 is original manual simple data. Table 2 is the result of normalization of table 1.The use of normalization, in order to provide data in the simplest and significant way. By using normalization, some risks can be easily transparent by the user immediately. Normalization are also help with the reduction of the processing time and memory space. Normalization could help numerical data in generation better accuracy. So this paper is mainly done about normalization.

3.3 Min_Max normalization description

This is a function to apply min-max normalization to a matrix or data frame.

3.4 Min_Max normalization Usage

Mmnorm (data, minval=0,maxval=1)

3.5 Min_Max normalization Arguments

Data the dataset to be normalized, including classes:

Minval

Maxvalue

{the maximum value of the transformed range}

3.6 System Flow Diagram

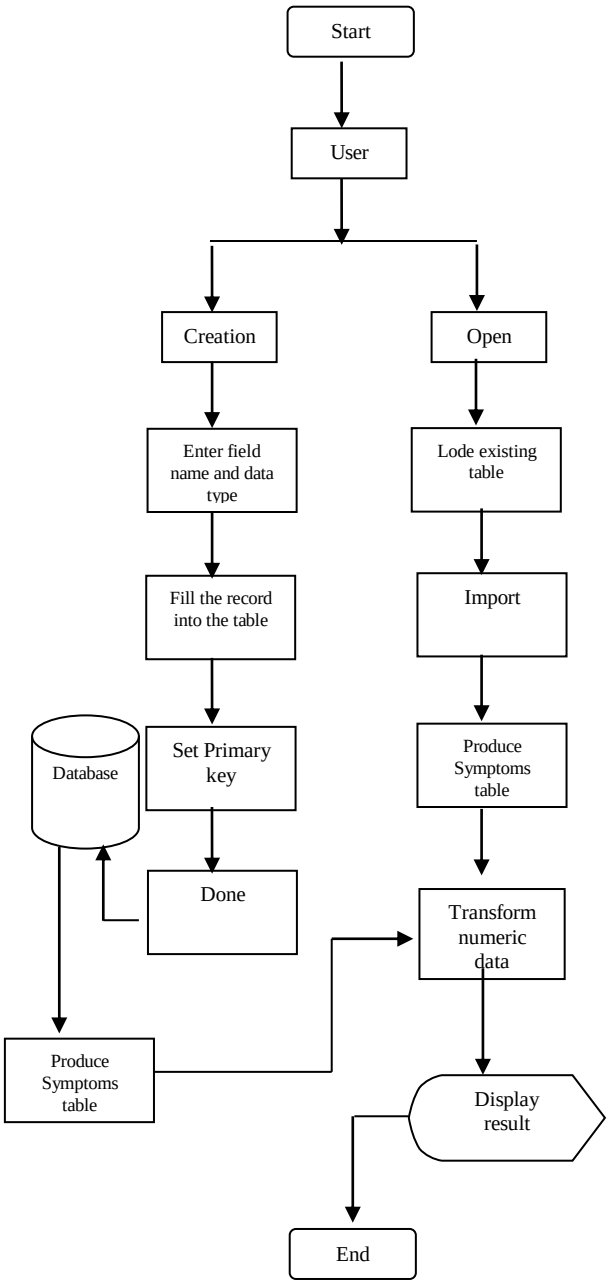


Figure1 : System Flow Diagram

This system flow diagram, the user can choose two parts which normalized the creative data and existing data. In data creation, the user has to type the field name data type in table and normalize the numeric data. When a user normalize the existing data, the user has to take the saved table in anywhere into the system after that normalize the numeric data A major goal of normalization is to eliminate redundancy by having a data element represented in only one place. Normalization is a "scaling down" transformation of the features. Within a feature there is often a large difference between the maximum and minimum values.

3.7 Min_Max normalization method

Min-max normalization subtracts the minimum value of an attribute from each value of the attribute and then divides the difference by the range of the attribute. These new values are multiplied by the new range of the attribute and finally added to the new minimum value of the attribute. These operations transform the data into a new range, generally [0, 1]. Min-max normalization performs a linear transformation on the original data. Suppose that mina and maxa are the minimum and the maximum values for attribute A. Min-max normalization maps a value v of A to v' in the range new-mina, new-maxa by computing:

v'= ((v-mina) / (maxa – mina)) * (new-maxa – newmina) + new-mina eq (1)

For Example Suppose that minimum and maximum value are $ 12,000 and $98,000.we would like the range [0.0, 1.0]. By min_max normalization a value is 73,600.

v'= ((v-mina) / (maxa – mina)) * (new-maxa – newmina) + new-mina

=73,000-12,000/98,000-12,000(1.0 - 0.0) + 0

=61000/86000*1 =0.716

4. Conclusion

This paper uses normalization method to be the best comperss for data value. If the user wants to compress the data value, the user can use normalization method. This paper may compress data by min_max normalization method. Normalization helps with the reduction of processing time and memory space in the dataset . Creating primary key constraints will reduce the number of empty or null values in columns and reduces the overall size of the database This paper examines the performance of data mining using transformation theory with min_max normalization method. When applying data mining to the real world learning from data that fall within a large specific range is an evitable situation. Trying to normalize data is an obvious solution where data are scaled so as to fall within a small specific range. The paper presented an overview and brief insight into the research in the area of data mining. The transformation programme will deliver a truly data focused service that is efficient, cost effective and relevant to the needs of local economies . Normalization involves changing the form of data in a table, so that the normalize table takes up less space than the original table. By using normalization, some risks can be easily transparent by the user immediately. Normalization can also help with the reduction of the risk. Normalization could help numerical data in generation better accuracy. By normalization of data, this system can reduce memory space about 1/4 percent.

5. References

[1] C.J.DATE, "AN INTRODUCTION TO DATABASE SYSTEMS",SIXTH EDITION, 1995.

[2] J.Han, M. Kamber, "Data Mining Concepts and Techniques", Morgan Kaufman Publishers, 2001.

[3] S. B. Kotsiantis, D. Kanellopoulos and P. E. Pintelas, " Data Preprocessing for Supervised Leaning", 2006.

[4] E.Rahm H. H.Do, "Problems and Current Approaches",University of Leipzig, Germany.

[5] C.Rodriguez and E.Acuna,"The discovery of logical prepositions in numerical data", Japan.

[6] L. A.Shalabi, Z.Shaaban and B.Kasasbeh, "A Preprocessing Engine**",** Applied Science University, Amman, Jordan.

[7] R.Snelick,U.Uludag, A.Mink, M.Indovina and A.Jain normalization and fusion methods that improve the multimodal system performance.

[8] http://www.cse.cuhk.hk/~csc3170/Index.html.

[9]http://en.wikipedia.org/wiki/Knowledge_representation.

[10]http://ori.dhhs.gov/education/products/n_illinois_u/datamanagement/dstopic.html.

[11]www.cs.tut.fi/~anix/databases/Normalization.ppt.