Filomat 34:1 (2020), 175–185
https://doi.org/10.2298/FIL2001175M

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

Missing Value Prediction for Qualitative Information Systems

T. Medhat[a], Manal Elsayed[b]

[a]*Electrical Engineering Department, Faculty of Engineering, Kafrelsheikh University, Kafrelsheikh, Egypt*
[b]*Physics and Engineering Mathematics Department, Faculty of Engineering, Kafrelsheikh University, Kafrelsheikh, Egypt*

Abstract. Most information systems usually have some missing values due to unavailable data. Missing values have a negative impact on the quality of classification rules generated by data mining systems. They make it difficult to obtain useful information from the data set. Solving the missing data problem is a high priority in the fields of knowledge discovery and data mining. The main goal of this paper is to suggest a method for converting a qualitative information system into a binary system, by using a distance function between condition attributes, we can detect the missing values for decision attribute according to the smallest distance. Most common values can be used to solve the problem of repeated small distance for some cases. This method will be discussed in detail through a case study.

1. Introduction

Missing data is the situation where some values of some cases are missing. Dealing with missing data [9, 10, 14, 16, 17] is time-consuming. In our experience, fixing up problems caused by missing data sometimes takes longer than the analysis itself. Missing data or more generally incomplete data [1] (e.g. censored data which are partially missing because we know which intervals they fall into) occur frequently in medical studies, in the forms of nonresponse in patient surveys, noncompliance in clinical trials, nonreporting or delayed reporting to health surveillance systems, just to list a few. Three major difficulties with such incomplete-data problems are: (I) loss of information, efficiency or power due to loss of data; (II) complication in data handling, computation and analysis due to irregularities in the data patterns and nonapplicability of standard software and (III) potentially very serious bias due to systematic differences between the observed data and the unobserved data.

The best way of dealing with these problems, of course, is to avoid them in the first place. Unfortunately, in most real-life studies, be they medical or otherwise, the problem of incomplete data is unavoidable even if we have made the greatest possible efforts [1]. There are three different methods for imputing missing data as described in [3]. Similarity measure is fundamental to many machine learning and data mining; similarity is learned from the data, so it will be more unbiased than that measured by traditional similarity metrics [6]. In the real-world applications, heterogeneous interdependent attributes that consist of both discrete and numerical variables can be observed ubiquitously [15].

2010 *Mathematics Subject Classification*. Primary O159
Keywords. Information System, Binary System, Reduction, Missing Values, Distance Function, Most Common Values
Received: 01 February 2019; Revised: 10 April 2019; Accepted: 18 June 2019
Communicated by Biljana Popović
Corresponding author: T. Medhat
Email addresses: tmedhatm@eng.kfs.edu.eg (T. Medhat), melsayed@eng.kfs.edu.eg (Manal Elsayed)

In many predictive modeling applications, useful attribute values ("features") may be missing. For example, patient data often have missing diagnostic tests that would be helpful for estimating the likelihood of diagnoses or for predicting treatment effectiveness. Consumer data often do not include values for all attributes useful for predicting buying preferences. It is important to distinguish two contexts: features may be missing at induction time, in the historical "training" data, or at prediction time, in to-be-predicted "test" cases. This paper introduces techniques for handling missing values at prediction time. Research on missing data in machine learning and statistics has been concerned primarily with induction time. Much less attention has been devoted to the development and to the evaluation of policies for dealing with missing attribute values at prediction time. Importantly for anyone wishing to apply models such as classification trees, there are almost no comparisons of existing approaches nor analyses or discussions of the conditions under which the different approaches perform well or poorly [13]. Rough Set Theory, proposed in 1982 by Zdzislaw Pawlak, is in a state of constant development. Its methodology is concerned with the classification and analysis of imprecise, uncertain or incomplete information and knowledge, and it is considered one of the first non-statistical approaches in data analysis [11, 12]. The fundamental concept behind Rough Set Theory is the approximation of lower and upper spaces of a set, the approximation of spaces being the formal classification of knowledge regarding the interesting domain [8]. The subset generated by lower approximations is characterized by objects that will definitely form part of an interesting subset, whereas the upper approximation is characterized by objects that will possibly form part of an interesting subset. Every subset defined through upper and lower approximation is known as Rough Set.

Over the years Rough Set Theory has become a valuable tool in the resolution of various problems, such as: representation of uncertain or imprecise knowledge; knowledge analysis; evaluation of quality and availability of information with respect to consistency; identification and evaluation of data dependency; reasoning based on uncertain and reduct of information data. The extent of rough set applications used today is much wider than in the past, principally in the areas of medicine, analysis of database attributes and process control. In this paper, we use the concepts of Rough Sets for attribute reduction [2, 4]. Many methods are sensitive to the used distance metric [5, 7]. The proposed method heavily depends on distance, and that is not sensitive to use the distance metric, because we calculate the distance after converting the system into binary system. Therefore, the result of distances belongs to the limited set of values: $\{0, 1, \sqrt{2}, \sqrt{3}, \sqrt{4}, \dots, \sqrt{N}\}$ where N is the number of condition attributes.

In this paper, we begin with section 2 as an introduction to the basic concepts. Section 3 introduces the method of converting qualitative information system into the binary information system. Section 4 introduces the method of missing values prediction, and at the last section, we conclude this paper.

2. Basic Concepts

2.1. information system:

Let $IS = (U, A \bigcup \{d\})$ be an information system, where U is the universe with a non-empty set of finite objects. A is a nonempty finite set of condition attributes, and d is the decision attribute (such a table is also called decision table). $\forall a \in A$ there is a corresponding function $f_a : U \to V_a$, where V_a is the set of values of a. If $P \subseteq A$, there is an associated equivalence relation [8, 11, 12]: as shown in Table 1.

Table 1: Example of information system table

U/A	Condition attributes			Decision attribute
	a	b	c	D
O1	1	1	5	accept
O2	2	0	3	reject
O3	1	0	3	reject
O4	2	0	4	accept

2.2. Qualitative information system:

If some values of condition attributes are non-numerical values, then that information system is called qualitative information system as shown in Table 2.

Table 2: Example of qualitative information system table

U/A	Condition attributes			Decision attribute
	a	b	c	D
x1	1	high	yes	yes
x2	2	low	no	no
x3	1	low	no	no
x4	2	medium	yes	yes

2.3. Binary information system:

If all values of condition attributes are binary values (0 or 1), then that information system is called binary information system as shown in Table 3.

Table 3: Example of binary information system table

U/A	Condition attributes			Decision attribute
	a	b	c	D
1	1	1	0	yes
2	0	1	1	yes
3	0	0	0	no
4	1	0	1	yes

2.4. Reduction of Condition Attributes Relative to Decision Attribute

Definition 2.1.
If C and D are the condition and decision attributes respectively. Then relative discernibility matrix [11] is:

$$M_C^D(x, y) = \{a \in C : a(x) \neq a(y), D(x) \neq D(y)\}$$

Definition 2.2.
If C and D are the condition and decision attributes respectively. Then relative discernibility function [11] is:

$$f_C^D = \wedge\{\vee a : a \in M_C^D \neq \emptyset\}$$

*which is used for **reduction** of condition attributes relative to decision attribute.*

2.5. Missing values:

If some values of decision attribute have quotation mark "?" as a value, then this value is called missing value, and we need to detect the values of missing values according to the given information system.

2.6. *Distance function:*

The distance between objects according to the values of condition attributes, can be calculated by the following function:

$$dis(O_i, O_j) = \sqrt{\sum_{k=1}^{N} [c_k(O_i) - c_k(O_j)]^2}$$

$$where\ \ O_i,\ O_j\ \in U,\ \ c_k \in C,\ N = \|C\| = number\ of\ condition\ attributes$$

2.7. *The most common value:*

The value of decision attribute which is repeated more than others, which is congruent to the statistical concept **mode**.

3. Converting into Binary System

In the following examples, **by converting each column of condition attribute into a number of columns equal to the number of its different values**. If its value exists "it takes 1" else "it takes 0". If the value of an attribute is only two values, then we put it as a single column "one of its value takes 1 and other takes 0" instead of increasing the columns without any usefulness. If the value of an attribute is only one value, we delete its column. See the following figure:

3.1. *Conversion rules:*

The conversion rules according to the types of information system are shown in Figure 1 as follows:

1. If an attribute has one value → We delete this attribute.
2. If an attribute has two different values → We convert its values one of them takes "0" and the other takes "1".
3. If an attribute has three different values → We convert it into three attributes (columns), takes values "0" or "1", according to the position of values.

U/A	a1	a2	a3
c1	H	L	5
c2	H	L	1
c3	H	M	1
c4	H	M	3
c5	H	L	1
c6	H	M	3
c7	H	L	5
c8	H	M	1

U/A	a1	a2	a3(1)	a3(3)	a3(5)
c1		0	0	0	1
c2		0	1	0	0
c3		1	1	0	0
c4		1	0	1	0
c5		0	1	0	0
c6		1	0	1	0
c7		0	0	0	1
c8		1	1	0	0

Figure 1: Example of converting into Binary system

4. Prediction of missing values

We will introduce a method (depending on the reduction of attributes, distance function, and most common values) to predict the decision for missing values. This will be done by:

1. Convert qualitative information system into a binary system.
2. Divide the binary information table into two tables, one of them is complete and the other is incomplete.
3. If needed, make a reduction of attributes for the complete decision table.
4. Compute the distance metric between objects of complete table and incomplete table.
5. The smallest distance means that the decision for missing value of the incomplete decision table equals the decision value of the complete decision table of the complete object which makes that distance.
6. If the small distance is repeated with more than one object, then we use the method of most common values, where we select the decision which has the largest number of repetition with the complete decision table.

See the following example.

Example:
The optometrist's data collection concerns the optician's decision as to whether or not the patient is suitable for contact lenses. The set of all possible decisions is listed in Table 4, which has 6 missing values "?" of the decision.
Where U={P1,P2,. . . . . . ,P24}, A={age, Spectacle, Astigmatic, Tear production rate}, and D={ Optician's decision}

EXPERIMENTAL RESULTS:

1) Converting the qualitative information system into a binary system, we get to Table 5:
2) Dividing the binary system into complete decision table and incomplete decision table as shown in Table 6 and Table 7.

3) Making a reduction of condition attributes relative to decision attribute of the complete decision table, which give us the following reducts:
Reduct={{b, c, d, g, h},{b, c, d, g, i},{a, c, e, f, i},{b, c, d, f, i},{a, b, e, f, i},{a, b, e, g, i},{a, b, d, g, i},{a, b, e, f, h},{b, c, e, f, h},{a, b, d, g, h},**{a, c, d, f, h}**,{a, c, d, g, h},{a, c, d, f, i},{a, b, e, g, h},{b, c, e, f, i},{b, c, e, g, h},{a, c, e, f, h},{a, c, e, g, h},{b, c, e, g, i},{a, c, d, g, i},{a, b, d, f, h},{b, c, d, f, h},{a, c, e, g, i},{a, b, d, f, i}}
We select one of them, which gives Table 8 and Table 9.

4) Calculating the distance function between the objects of incomplete decision table and complete decision table after reduction of condition attributes:

From Table 10, we find that:
The decision of object P13 ⇒ will be "no contact lenses".
The decision of object P14 ⇒ will be "soft contact lenses".
The decision of object P19 ⇒ will be "no contact lenses".
And the decision of object P20 ⇒ will be "hard contact lenses".
But we can't predict the missing values of objects P1, and P7, according to the repetition of small distance with more than one object which has different decisions. So we need to calculate the most common values of decision values.

5) Calculating the most common values for all values of decision attribute, and Appling this method to detect the missing values for repeated small distance, as shown in Table 11.

Table 4: The optician's decisions data set

U/A	Condition attributes				Decision attribute (Optician's decision)
	Age	**Spectacle**	**Astigmatic**	**Tear production rate**	
P1	Young	Hypermetropia	No	Reduced	?
P2	Young	Hypermetropia	No	Normal	soft contact lenses
P3	Pre-presbyopic	Hypermetropia	No	Reduced	no contact lenses
P4	Pre-presbyopic	Hypermetropia	No	Normal	soft contact lenses
P5	Presbyopic	Hypermetropia	No	Reduced	no contact lenses
P6	Presbyopic	Hypermetropia	No	Normal	soft contact lenses
P7	Young	Hypermetropia	Yes	Reduced	?
P8	Young	Hypermetropia	Yes	Normal	hard contact lenses
P9	Pre-presbyopic	Hypermetropia	Yes	Reduced	no contact lenses
P10	Pre-presbyopic	Hypermetrope	Yes	Normal	no contact lenses
P11	Presbyopic	Hypermetropia	Yes	Reduced	no contact lenses
P12	Presbyopic	Hypermetrope	Yes	Normal	no contact lenses
P13	Young	Myope	No	Reduced	?
P14	Young	Myope	No	Normal	?
P15	Pre-presbyopic	Myope	No	Reduced	no contact lenses
P16	Pre-presbyopic	Myope	No	Normal	soft contact lenses
P17	Presbyopic	Myope	No	Reduced	no contact lenses
P18	Presbyopic	Myope	No	Normal	no contact lenses
P19	Young	Myope	Yes	Reduced	?
P20	Young	Myope	Yes	Normal	?
P21	Pre-presbyopic	Myope	Yes	Reduced	no contact lenses
P22	Pre-presbyopic	Myope	Yes	Normal	hard contact lenses
P23	Presbyopic	Myope	Yes	Reduced	no contact lenses
P24	Presbyopic	Myope	Yes	Normal	hard contact lenses

Table 12 shows the prediction of missing values of objects P1 and P7 after computing the most common values of decision attribute.

Table 5: The optician's decisions data set after converting to a binary system

U/A	**Age**			**Spectacle**		**Astigmatic**		**Tear production rate**		**Decision**
	Young	**Pre-presbyopic**	**Presbyopic**	**Hypermetropia**	**Myope**	**No**	**Yes**	**Reduced**	**Normal**	
P1	1	0	0	1	0	1	0	1	0	?
P2	1	0	0	1	0	1	0	0	1	soft contact lenses
P3	0	1	0	1	0	1	0	1	0	no contact lenses
P4	0	1	0	1	0	1	0	0	1	soft contact lenses
P5	0	0	1	1	0	1	0	1	0	no contact lenses
P6	0	0	1	1	0	1	0	0	1	soft contact lenses
P7	1	0	0	1	0	0	1	1	0	?
P8	1	0	0	1	0	0	1	0	1	hard contact lenses
P9	0	1	0	1	0	0	1	1	0	no contact lenses
P10	0	1	0	1	0	0	1	0	1	no contact lenses
P11	0	0	1	1	0	0	1	1	0	no contact lenses
P12	0	0	1	1	0	0	1	0	1	no contact lenses
P13	1	0	0	0	1	1	0	1	0	?
P14	1	0	0	0	1	1	0	0	1	?
P15	0	1	0	0	1	1	0	1	0	no contact lenses
P16	0	1	0	0	1	1	0	0	1	soft contact lenses
P17	0	0	1	0	1	1	0	1	0	no contact lenses
P18	0	0	1	0	1	1	0	0	1	no contact lenses
P19	1	0	0	0	1	0	1	1	0	?
P20	1	0	0	0	1	0	1	0	1	?
P21	0	1	0	0	1	0	1	1	0	no contact lenses
P22	0	1	0	0	1	0	1	0	1	hard contact lenses
P23	0	0	1	0	1	0	1	1	0	no contact lenses
P24	0	0	1	0	1	0	1	0	1	hard contact lenses

Table 6: Complete decision table

#	U/A	Age			Spectacle		Astigmatic		Tear production rate		D
		a	b	c	d	e	f	g	h	i	
		Young	Pre-presbyopic	Presbyopic	Hypermetropia	Myope	No	Yes	Reduced	Normal	Decision
1	**P2**	1	0	0	1	0	1	0	0	1	soft contact lenses
2	**P3**	0	1	0	1	0	1	0	1	0	no contact lenses
3	**P4**	0	1	0	1	0	1	0	0	1	soft contact lenses
4	**P5**	0	0	1	1	0	1	0	1	0	no contact lenses
5	**P6**	0	0	1	1	0	1	0	0	1	soft contact lenses
6	**P8**	1	0	0	1	0	0	1	0	1	hard contact lenses
7	**P9**	0	1	0	1	0	0	1	1	0	no contact lenses
8	**P10**	0	1	0	1	0	0	1	0	1	no contact lenses
9	**P11**	0	0	1	1	0	0	1	1	0	no contact lenses
10	**P12**	0	0	1	1	0	0	1	0	1	no contact lenses
11	**P15**	0	1	0	0	1	1	0	1	0	no contact lenses
12	**P16**	0	1	0	0	1	1	0	0	1	soft contact lenses
13	**P17**	0	0	1	0	1	1	0	1	0	no contact lenses
14	**P18**	0	0	1	0	1	1	0	0	1	no contact lenses
15	**P21**	0	1	0	0	1	0	1	1	0	no contact lenses
16	**P22**	0	1	0	0	1	0	1	0	1	hard contact lenses
17	**P23**	0	0	1	0	1	0	1	1	0	no contact lenses
18	**P24**	0	0	1	0	1	0	1	0	1	hard contact lenses

Table 7: Incomplete decision table

#	U/A	Age			Spectacle		Astigmatic		Tear production rate		D
		a	b	c	d	e	f	g	h	i	
		Young	Pre-presbyopic	Presbyopic	Hypermetropia	Myope	No	Yes	Reduced	Normal	Decision
1	**P1**	1	0	0	1	0	1	0	1	0	?
2	**P7**	1	0	0	1	0	0	1	1	0	?
3	**P13**	1	0	0	0	1	1	0	1	0	?
4	**P14**	1	0	0	0	1	1	0	0	1	?
5	**P19**	1	0	0	0	1	0	1	1	0	?
6	**P20**	1	0	0	0	1	0	1	0	1	?

Table 8: Complete decision table after reduction

#	U/A	Age (Young)	Age (Presbyopic)	Spectacle (Hypermetropia)	Astigmatic (No)	Tear production rate (Reduced)	Decision
		a	**c**	**d**	**f**	**h**	**D**
1	**P2**	1	0	1	1	0	soft contact lenses
2	**P3**	0	0	1	1	1	no contact lenses
3	**P4**	0	0	1	1	0	soft contact lenses
4	**P5**	0	1	1	1	1	no contact lenses
5	**P6**	0	1	1	1	0	soft contact lenses
6	**P8**	1	0	1	0	0	hard contact lenses
7	**P9**	0	0	1	0	1	no contact lenses
8	**P10**	0	0	1	0	0	no contact lenses
9	**P11**	0	1	1	0	1	no contact lenses
10	**P12**	0	1	1	0	0	no contact lenses
11	**P15**	0	0	0	1	1	no contact lenses
12	**P16**	0	0	0	1	0	soft contact lenses
13	**P17**	0	1	0	1	1	no contact lenses
14	**P18**	0	1	0	1	0	no contact lenses
15	**P21**	0	0	0	0	1	no contact lenses
16	**P22**	0	0	0	0	0	hard contact lenses
17	**P23**	0	1	0	0	1	no contact lenses
18	**P24**	0	1	0	0	0	hard contact lenses

Table 9: Incomplete decision table after reduction

#	U/A	Age (Young)	Age (Presbyopic)	Spectacle (Hypermetropia)	Astigmatic (No)	Tear production rate (Reduced)	Decision
		a	**c**	**d**	**f**	**h**	**D**
1	**P1**	1	0	1	1	1	?
2	**P7**	1	0	1	0	1	?
3	**P13**	1	0	0	1	1	?
4	**P14**	1	0	0	1	0	?
5	**P19**	1	0	0	0	1	?
6	**P20**	1	0	0	0	0	?

Table 10: Decision Table of Missing Values of Some Objects

Objects with missing decision	Objects with decision	Small distance	Decision of complete objects	Decision of incomplete objects
P1	P2	1	soft contact lenses	?
	P3	1	no contact lenses	?
P7	P8	1	hard contact lenses	?
	P9	1	no contact lenses	?
P13	P15	1	no contact lenses	no contact lenses
P14	P2	1	soft contact lenses	soft contact lenses
	P16	1	soft contact lenses	
P19	P21	1	no contact lenses	no contact lenses
P20	P8	1	hard contact lenses	hard contact lenses
	P22	1	hard contact lenses	

Table 11: The most common values of decision values

Decision values	**Number of repetition**
no contact lenses	11
soft contact lenses	4
hard contact lenses	3

Table 12: Decision Table of Missing Values of P1 and P7

Objects with missing decision	**Objects with decision**	**Small distance**	**Decision of complete objects**	**Decision of incomplete objects**
P1	P2	1	soft contact lenses	no contact lenses
	P3	1	no contact lenses	
P7	P8	1	hard contact lenses	no contact lenses
	P9	1	no contact lenses	

5. Conclusion

By converting the qualitative information system tables into binary tables, we can make a reduction of condition attributes and can predict the missing values of decision attribute according to the distance function and the most common values. This method provides us a new technique, which predicts the missing values according to the real data instead of making a mapping from the qualitative values to numbers.

Acknowledgement

The authors would like to thank Prof. Dr. A. M. Kozae, Prof. Dr. Dominik Selzak, and Prof. Dr. Abdalsalam Alsabbgh for their encouragement and support, and sincerely thank the anonymous reviewers whose comments have greatly helped clarify and improve this paper.

References

[1] J. Barnard, X. Meng, Applications of multiple imputations in medical studies: from AIDS to NHANES, Statistical Methods in Medical Research 8(1999) 17-36.
[2] Z. Geng, Z. Qunxiong, A New Rough Set-Based Heuristic Algorithm for Attribute Reduct, Proceedings of the 6th World Congress on Intelligent Control and Automation, ISBN 1-4244-0332-4, Dalian-China, Jun. 21-23, 2006, Wuhan University of Technology Press, Wuhan-China (2006) 3085-3089.
[3] T. Hastie, R. Tibshirani, G. Sherlock, Imputing missing data for gene expression arrays, Technical Report, Division of Biostatistics, Stanford University (1999) 1–9.
[4] X. Hu, N. Cercone, J. Han,W. Ziarko, GRS: A Generalized Rough Sets Model, in Data Mining, Data Mining, Rough Sets and Granular Computing, T.Y. Lin, Y.Y.Yao and L. Zadeh (eds), Physica-Verlag (2002) 447-460.
[5] Z. Kang, H. Pan, S. C. H. Hoi, Z. Xu, Robust Graph Learning From Noisy Data, IEEE Transactions on Cybernetics (2018) 1-11.
[6] Z. Kang, C. Peng, Q. Cheng, Kernel-driven Similarity Learning, Neurocomputing 267 (2017) 210-219.
[7] Z. Kang, L. Wen, W. Chen, Z. Xu, Low-rank Kernel Learning for Graph-based Clustering, Knowledge-Based Systems 163 (2019) 510-517.
[8] E. F. Lashin, A. M. Kozae, A. A. Abo Khadra, T. Medhat, Rough set theory for topological spaces, International Journal of Approximate Reasoning 40(1-2) (2005) 35-43.
[9] T. Medhat, Missing Values Via Covering Rough Sets, IJMIA: International Journal on Data Mining and Intelligent Information Technology Applications, 2(1) (2012) 10-17.
[10] T. Medhat, Prediction of Missing Values for Decision Attribute, I.J. Information Technology and Computer Science 4(11) (2012) 58–66.
[11] Z. Pawlak, Rough sets: Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Dordrecht, Boston, London, 1991.
[12] Z. Pawlak, Rough sets, International Journal of Computer and Information Sciences 11 (1982) 341?356.
[13] M. Saar-Tsechansky, F. Provost, Handling Missing Values when Applying Classification Models, Journal of Machine Learning Research 8 (2007) 1217-1250.
[14] K. L. Sainani, Dealing With Missing Data, PM and R 7(9) (2015) 990-994.
[15] C. Wang , C. Chi, W. Zhou, R. Wong, Coupled Interdependent Attribute Analysis on Mixed Data, In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI–15). (2015) 1861–1867.
[16] H. Wang, S. Wang, Towards optimal use of incomplete classification data, Comput. Oper. Res. 36 (2009) 1221-1230.
[17] Y. Zha, A. Song, C. Xu, H. Yang, Dealing with missing data based on data envelopment analysis and halo effect, Applied Mathematical Modelling 37(9) (2013) 6135-6145.