A manual for Contrast Targeted Positive and Negative Association Rule Mining (CTR)

Contents

1   Introduction

This package provides an implementation of the *Contrast Targeted Positive and Negative Association Rule Mining* (CTR) algorithm proposed in [1, 2]. The CTR algorithm generates targeted positive/negative rules and contrast between multiple datasets. The package is implemented under Matlab environment. In the following sections we illustrate the usage of this package on a sample dataset.

2   Backgrounds

In this demo, we aim to discover associations between hotel customers demographic information and their opinions on hotel *value* and *room quality*. For example, if "$female, age2 \rightarrow room0$" is a strong rule, then it indicates that young female customers are unlikely to satisfied with *room quality*. Further more we want to contrast the stregnth of the rules across the years to see whether there exists differences or trends among the years.

3   Dataset

A sample dataset is provided in the code package (sample.xlsx). The dataset has a total of 7 attributes as shown in Table 1. The first attribute *year* is the contrast attribute which contains values to splite the dataset into sub-datasets. The *value* and *room* are the target attributes. Remaining attributes are the demographic attributes.

4   Code Execution

Open Matlab and change the current directory to the code package. The sample dataset "sample.xlsx" has already been put under the same directory of the code. Execute the following Matlab command:

```
poNeContrast('sample.xlsx', 'Sheet1', 2, 0.05, 0.005, 0.01, 3);
```

there are 7 parameters for the function, where

parameter 1 - name of the excel data file.

parameter 2 - name of the worksheet inside data file.

parameter 3 - number of the target attributes. the target attributes must start from the second column. In this demo the target attributes are "value" and "room" therefore we set 2 to indicates there are 2 target attributes.

parameter 4 - minimum support. 0.05, 0.1, 0.2 are some typical values, refer to the paper for more details.

parameter 5 - minimum interest. Refer to the paper for more details.

Table 1: Attributes in the dataset

Attribute	Description	Possible values	Label
YEAR	this attribute splits the whole data set into several sub-datasets based on different year values	2008 2009 2010 2011	
VALUE	Value of the hotel	0 (Not satisfied) 1 (Satisfied)	$value0$ $value1$
ROOM	Room quality	0 (Not satisfied) 1 (Satisfied)	$room0$ $room1$
AGE	Age	1 (Under 25) 2 ($26 - 35$) 3 ($36 - 45$) 4 ($46 - 55$) 5 (56 years old or above)	$age1$ $age2$ $age3$ $age4$ $age5$
GENDER	Gender	1 (Male) 2 (Female)	$male$ $female$
MOTIVATION	Motivation of trip	1 (For work) 2 (For fun) 3 (For work and fun)	$work$ fun $workfun$
CITIES	Number of cities visited	1 (1-10) 2 (11-20) 3 (21-30) 4 (31 or more)	$cities1$ $cities2$ $cities3$ $cities4$

parameter 6 - surprising threshold. Refer to the paper for more details.

parameter 7 - max level of itemsets, i.e. the max length of the association rule. For example, length of the rule "$female, age2 \rightarrow room0$" is 3.

You may need to change the values for your own project. The above Matlab command will execute for a while and 3 output files will be generated, which contain the extracted rules.

5 Results Interpolation

In this demo, we are contrasting the rules across 4 years (see the first column of the data file). The Matlab command will generate three excel files. Open the one named "PoDeviation.xls" and look at the worksheet "Sheet1". Worksheet "Sheet1" contains length 2 rules. If you run the Matlab command correctly, the first rule should be: "female value0 20.000 32.258 50.000 56.667" where $female \rightarrow value0$ is the rule and the following four values are the support values of this rule on 2008, 2009, 2010, and 2011 datasets. Refer to the papers for more details on rules interpolation.

3

References

[1] R. Law, J. Rong, H.Q. Vu, G. Li, and H.A. Lee. Identifying changes and trends in hong kong outbound tourism. *Tourism Management*, 32(5):1106–1114, 2011.

[2] Jia Rong, Huy Quan Vu, Rob Law, and Gang Li. A behavioral analysis of web sharers and browsers in hong kong using targeted association rule mining. *Tourism Management*, 33(4):731 – 740, 2012.