

Cluster based rule discovery model for enhancement of Government's tobacco control strategy

Shamsul Huda¹, John Yearwood¹, Ron Borland²

1. Centre for Informatics and Applied Optimization (CIAO), University of Ballarat, Australia

2. VicHealth Centre for Tobacco Control, The Cancer Council Victoria, Melbourne, Australia

E-mails: 1. s.huda@ballarat.edu.au, j.yearwood@ballarat.edu.au 2. Ron.Borland@cancervic.org.au

Abstract—Discovery of interesting rules describing the behavioural patterns of smokers' quitting intentions is an important task in the determination of an effective tobacco control strategy. In this paper, we investigate a compact and simplified rule discovery process for predicting smokers' quitting behaviour that can provide feedback to build an scientific evidence-based adaptive tobacco control policy. Standard decision tree (SDT) based rule discovery depends on decision boundaries in the feature space which are orthogonal to the axis of the feature of a particular decision node. This may limit the ability of SDT to learn intermediate concepts for high dimensional large datasets such as tobacco control. In this paper, we propose a cluster based rule discovery model (CRDM) for generation of more compact and simplified rules for the enhancement of tobacco control policy. The cluster-based approach builds conceptual groups from which a set of decision trees (a decision forest) are constructed. Experimental results on the tobacco control data set show that decision rules from the decision forest constructed by CRDM are simpler and can predict smokers' quitting intention more accurately than a single decision tree.

Keywords: Tobacco control, Cluster analysis, Decision rule, Univariate Decision Tree, Multivariate decision tree, rule discovery.

I. INTRODUCTION:

Tobacco smoking has a large influence on health and is a significant cause of death. It is one of the main causes of death (ITCEP, WHO) [1] and currently 5.4 million people die every year due to tobacco smoking (ITCEP, WHO) [1] in the world. Smoking is the top major cause of death (DHS, Melbourne) [2] and every year 4000 people die due to smoking in Victoria (VTCS) [4] with an additional cost over \$5 billion each year for Victorians (VTCS) [4]. There were approximately 2.9 million people aged 14 years or older who smoked tobacco daily (NDSH)[3] in Australia, in 2007. In Victoria, 17.3% of adults are regular smokers (CCV) [5],[6]. Tobacco smoking causes death for more than 15,000 Australians every year [7], [8]. Therefore controlling tobacco smoking has become a social demand. Governments of many countries have included it as a major state regulation and started implementing different policies to encourage tobacco smokers to quit tobacco smoking. The Australian government has also adopted the "Australian National

Tobacco Strategy" and has applied different types of regulation and policies to tobacco. However controlling tobacco smoking and determining corresponding policies is a difficult task since it is related to human habit, behaviour and activities as well as relationships with tobacco industries. Therefore policy maker need feedback from research to adopt more fruitful policies. This feedback is usually obtained from ground level surveys. However survey data does not directly help much and can not explore the overall picture of the effect of the policies. In this case, decision support systems (DSS) can be very useful.

DSS often engages in predictive data mining tasks and discovers interesting rules which can give important guidance to policy makers. Growing a decision tree from available data is one of the most popular approaches of predictive data mining and widely used in the machine learning and data mining community. Each path from the root of a decision tree to one of its leaves can be easily transformed to a decision rule by co-joining the intermediate nodes' test-conditions. This forms the antecedent part of the rule and the class value of the leaf forms the consequent part of the rule. In this way a decision tree is also widely used as an interesting rule generator.

Researchers in tobacco control have also used decision trees in determining effective policy. Previous attempt on tobacco control using decision tree based approaches were based on pre-defined concept oriented datasets [9], [10] such as demographic, psychological or particular age group (more on these concepts has been described in section-2). However, the separation of training data based on predefined concepts may bias the generated rule set which may fail to reflect the actual effect of a control policy. Another approach considered a single standard decision tree (SDT) [11] to generalize the input and target attributes relationship [9], [10].

SDTs show very good expressive power, however, SDTs are univariate [12]. This may limit the capacity of an SDT to learn the intermediate concepts for high dimensional large datasets such as tobacco control. Researchers have tried to use multivariate splitting criteria [12] in decision trees. However finding the best multivariate criteria for a high dimensional data set is also complicated and computationally expensive. Therefore, in this paper, we propose a cluster-based rule discovery model (CRDM) for the generation of more compact and simplified rules for enhancement of tobacco control policy. The cluster-based approach builds conceptual groups from which a set of decision trees (a decision forest) is constructed. Then interesting rule sets can

be extracted from the decision forest. Extracted rules from the decision forest are simple and show more correct prediction capability.

The rest of the paper is organized as follows. The next section describes the tobacco control data sets. Section 3 gives a detailed description of the proposed CRDM. Experimental analysis of CRDM and results are described in section 4. The conclusion of this study is given in the last section.

II. DATA SETS

International Tobacco Control Policy Evaluation Project (ITC Project) completed a four country survey (known as ITC-4 data) [13], [14] and [15] with a target of estimating the impact of psychological and behavioural impact of the key policies of Framework Convention on Tobacco Control (FCTC) [1], [13], [14] and [15] organized by the World Health Organization (WHO). The Four-Country Survey was made among randomly selected smokers in four English-speaking countries: Canada, the United States, the United Kingdom, and Australia. ITC-4 participant smokers are adult who have smoked more than 100 cigarettes in their lifetimes and have smoked at least once in the past 30 days. The survey consists of four waves. First survey (Wave-1) was completed during October-December 2002. Following every 8 or 9 months (approximately) a survey was conducted. Wave-2 was conducted during May-August 2003, Wave-3 during June-December 2004, and Wave 4 from September-December 2005. Eighty five or more questions have been considered to evaluate the impact of tobacco control policy measures for different waves. Survey question are mainly based on psychosocial – beliefs about smoking, beliefs about quitting, psychosocial questions such as perceived risk and health worry, smoking behaviour such as total minutes to first cigarette, addictedness to cigarettes), knowledge of health effects/tobacco constituents, socio-demographic questions such income, smokers’ reaction and outcome on cessation advice and services, smokers’ reactions on warning labels, advertising, monitoring of anti-tobacco campaigns, price/taxation and sources of tobacco, smokers’ reactions and effect on smoking restrictions. The main outcome questions is whether the smokers’ have made any attempt to stop smoking since they were interviewed last or they have stopped smoking for about 6 months. Questions of ITC-4 are described in the Appendix section and in [1], [13], [14].

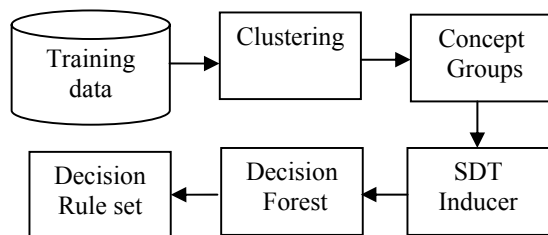


Figure-1: Proposed CRDM for rule discovery in tobacco control

III. METHODOLOGY:

In this paper we propose a cluster-based rule discovery model (CRDM) for determining effective tobacco control policy. In CRDM, conceptual groups are formed from sample data and then a set of SDTs (a decision forest) is constructed. Decision rules are extracted from the decision forest. The next section discusses the detail of CRDM.

A. Standard Decision Tree (SDT)

SDTs [11], [12] are one of the popular approaches in predictive data mining tasks and widely used in decision support systems (DSS). A SDT is a rooted tree with a node called the root that has no incoming edge. Nodes with both incoming and outgoing edges are called internal nodes. Nodes with no outgoing edges are called leaves. In general, SDTs are constructed by following a divide and conquer search strategy that recursively partition the training spaces into subspaces according to the value of a single feature. The selection of an input features in partitioning the sample space is done by some goodness measure. The goodness measure ranks the features and the best feature is chosen. Many goodness measure have been proposed such as impurity based criteria [11], [16] likelihood ratio [17], Gain ratio [11], [16]. Each path from a root of the SDT to one of its leaves can easily be transformed to a decision rule by co-joining the intermediate nodes’ test-conditions that forms the antecedent part of the rule and class value of the leaf forms the consequent part of the rule.

B. Cluster Analysis

Clustering is a process of grouping of a set of samples in a manner that maximizes the intraclass similarity and minimizes the interclass similarity. The process is also known as unsupervised classification where a set of unsupervised data are separated into a discrete set of natural and hidden structures. When sample spaces are clustered the samples within a cluster have high similarity with each other and show high dissimilarity to the samples of other clusters. The clustering process can use various proximity measures (e.g. Minkowski distance [18], Mahalanobis distance [19], Pearson co-relation [20], Global-K-Means [21] and various criterion functions (sum square error [22], Maximum Likelihood [23] in grouping the unsupervised data.

C. Proposed Cluster-based Rule Discovery Model (CRDM)

SDT is univariate where the decision boundaries in the feature space are geometrically orthogonal to the axis of the feature of a particular decision node. This may limit SDT’s ability to learn the intermediate concepts for high dimensional data set. Use of multivariate splitting criteria can overcome the problem up to a certain extent. Multivariate splitting criteria is based on a linear combination of input attributes. However finding best multivariate criteria at each intermediate node of the tree is also complicated and computationally expensive.

We propose a cluster-based rule discovery model (CRDM). In CRDM, a clustering algorithm is employed to find the natural groups from the sample space based on the

hidden pattern of the data. Then SDT is applied on the conceptual groups from which a set of decision trees (a decision forest) is constructed. Since several conceptual groups are built based on the hidden data structure, the decision trees based on the conceptual groups become more simple and compact. The CRDM is presented in the Figure-1.

In the first step of CRDM, a clustering algorithm, Global-K-Means [21] is used which is based-on K-means clustering algorithm. It is an incremental algorithm that dynamically adds one cluster centre at a time and uses each data point as a candidate for the k-th cluster centre [21]. A starting point for the k-th cluster centre in this algorithm is computed by minimizing an auxiliary cluster function [21]. When data are clustered, an efficient decision tree inducer, an efficient version of C4.5 (EC4.5) [24] has been used to generate the decision forest from the conceptual groups. The clustered data is processed using the decision tree. EC4.5 [24] is an update version of C4.5 which improves the information gain computation for continuous attributes and thus is more computationally efficient. In the last step, decision rules are extracted from the decision forest.

The cluster Characteristics has been determined by relabelling the training and test data according to their cluster label and then by generating the rules for the cluster using decision tree from each cluster.

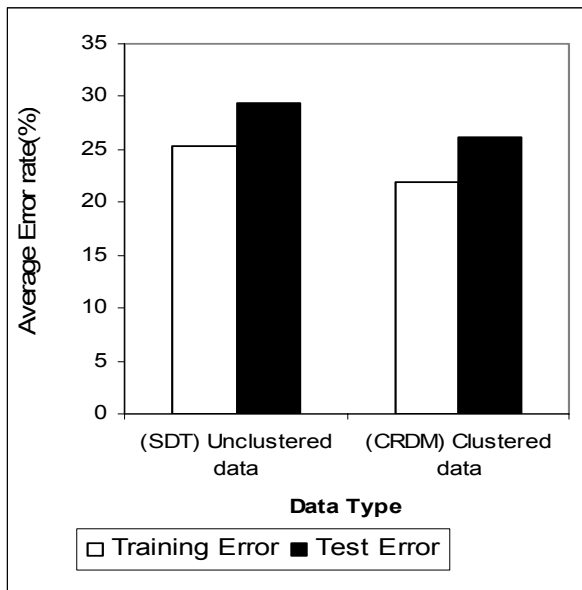


Figure-2: Comparison of average error rate (%) of CRDM (Clustered data) and SDT (Unclustered data)

IV. EXPERIMENTAL ANALYSIS

A. Settings

ITC-4 (Wave-1) [13], [14] AND [15] data set has been used for rule discovery in tobacco control and to test the efficiency of the rules. We applied Global-K-Means [21] to cluster the data sets. Then EC4.5 [24] was applied on each cluster to build the decision trees.

Training Data and Test Data:

For each cluster, we randomly divide the data into training and test sets. Two-third of data of each cluster has been taken for training and the remaining one-third has been taken as test data. The random division of data into training and test has been done for five trials. Therefore, five different tests from the random divisions of each cluster data have been performed for each cluster. The test with lowest error has been adopted for the decision tree. The decision tree constructed from each cluster is applied on the corresponding test set of each cluster to verify the prediction ability of the decision rules. In a set of separate experiments (total five), the whole data set has been divided into training and test sets. EC4.5 [24] has been applied on the whole training set and the corresponding SDT is applied on the test set. Finally, the average error rate from all clusters (by CRDM) has been compared with the average error rate of a complete test (a single decision tree without clustering).

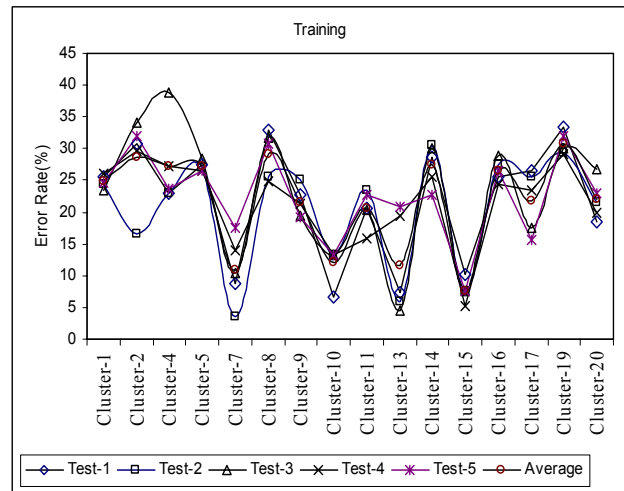


Figure-3: Error rate (%) of Five tests and average error rate (%) of CRDM in cluster 1 to 20 in training data.

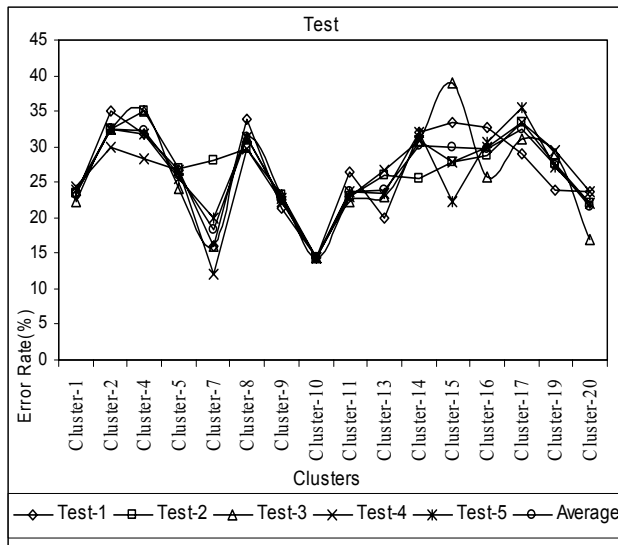


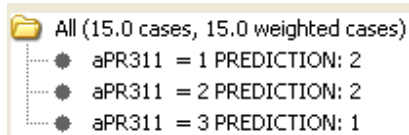
Figure-4: Error rate (%) of five tests and average error rate (%) of CRDM in cluster 1 to 20 of test data.

B. Results:

Global K-means finds total 20 clusters. The results have been presented in the Figure-2, 3, 4, 5. Figure-3, 4 describes the error rate of all five tests for cluster-1 to cluster-20 and their average error rate. It is seen in figures 2, 3 and 4 that the average error rate (26.21%) by the decision forest of the clustered data obtained by CRDM is less than the error rate (29.35%) by the single SDT of un-clustered data. This proves the effectiveness of CRDM based rule discovery approach. The extracted rules from the decision forest constructed by CRDM have been presented in the Table-1. The attributes' description of the rule is given in the Appendix section. It is seen in Table-1 that rules extracted by CRDM are also very simple. Some of the rules even have only one decision node in CRDM. Clusters-3, 6, 12, 18 have only one member. Therefore these have not been included in the results and less significant. Decision trees constructed by CRDM have small size and the tree-depth is very low. In contrast, the SDT from un-clustered data has 389 nodes and maximum depth 10. This gives very complex rules in which are not suitable for policy makers.

C. Interpretation of extracted rules

Rules can be easily extracted from the decision forest constructed by CRDM. For a particular decision tree from a cluster given below, the rule is extracted for cluster-10 as follows:

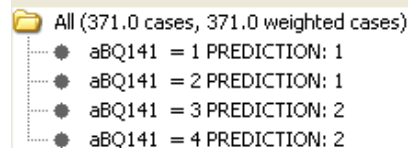


In cluster-10, the characteristics of the cluster is that smokers have (aSB012v>780) (aSB012v = total minutes to first cigarette).

The rule: [IF (aPR311 = 3) THEN Smokers made a quit attempt]. Otherwise they did not make a quit attempt

Where aPR311 means the question to the smokers as: **To what extent, if at all, has smoking damaged your health?**
Choice of answer: 01 – Not at all, 02 – Just a little, 03 – A fair amount, 04 – A great deal.

Another example of a decision tree for cluster-19 is as follows:



In Cluster-19, the smokers have general characteristics such that (aDE212v = 1 AND aSB012v <= 35 AND apu555 <= 31.78). This means that smokers have least income (less than 10,000\$) and total minutes to first cigarette is less than 35 and price/unit cigarette is greater than 33.

The rule: [IF (aBQ141 =1 OR aBQ141 = 2) THEN Smokers made a quit attempt].

Where aBQ141 means the question to the smokers as: **Are you planning to quit smoking?**

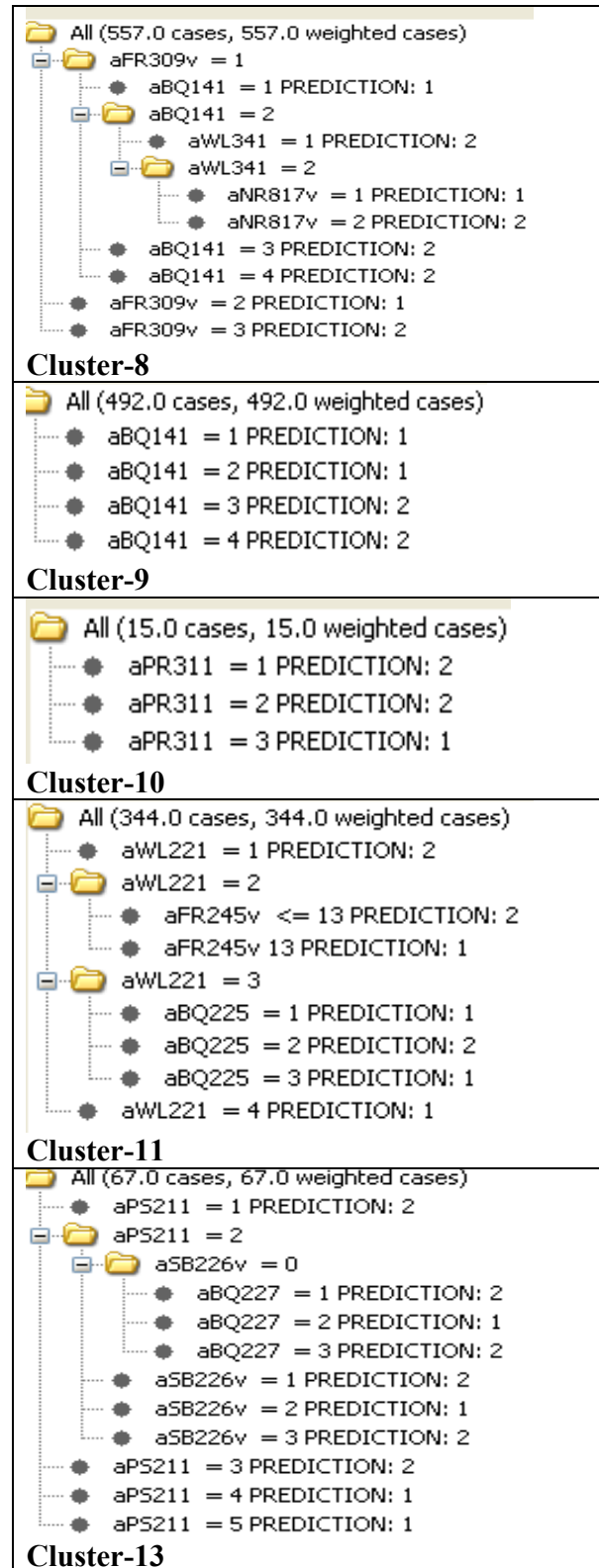
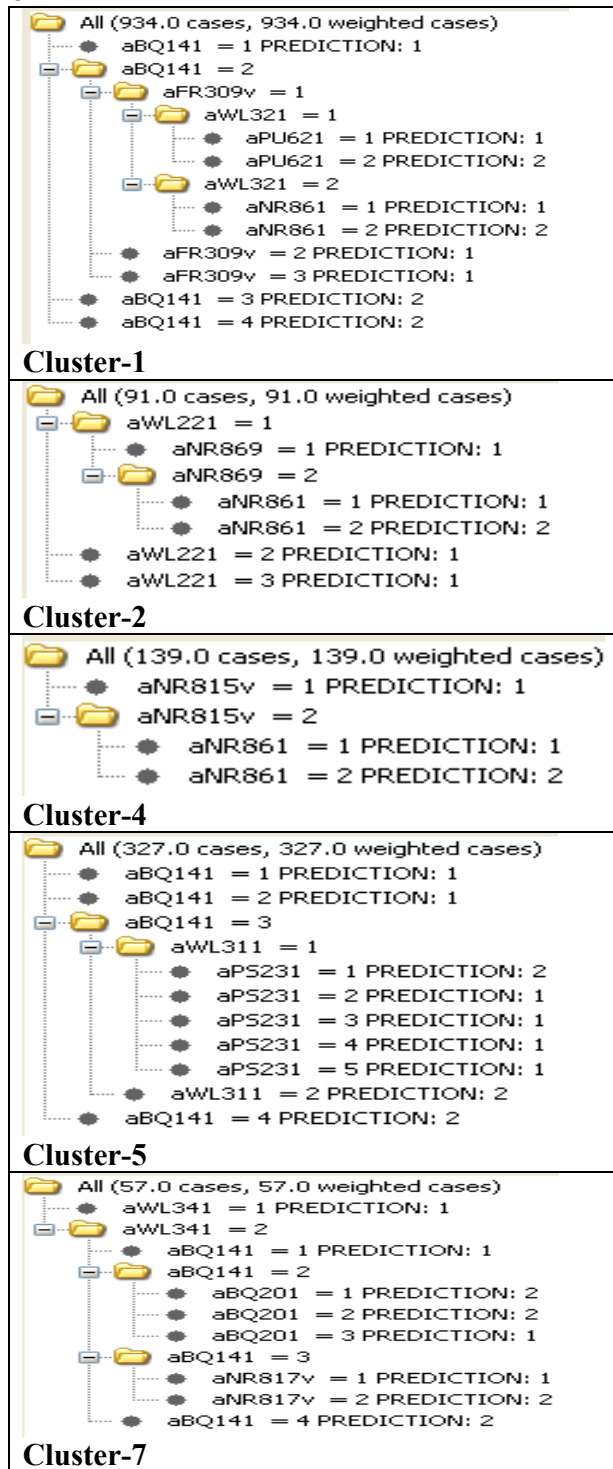
Choice of answer: 01 – Within the next month?, 02 – Within the next 6 months?, 03 – Sometime in the future, beyond 6 months,04 – Not planning to quit.

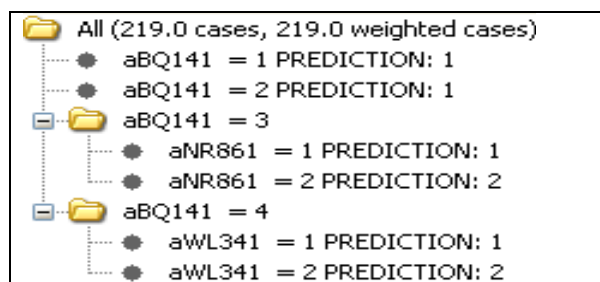
All decision trees have been given in Table-1 from which rules can be extracted and attributes have been described in the Appendix section.

V. CONCLUSIONS:

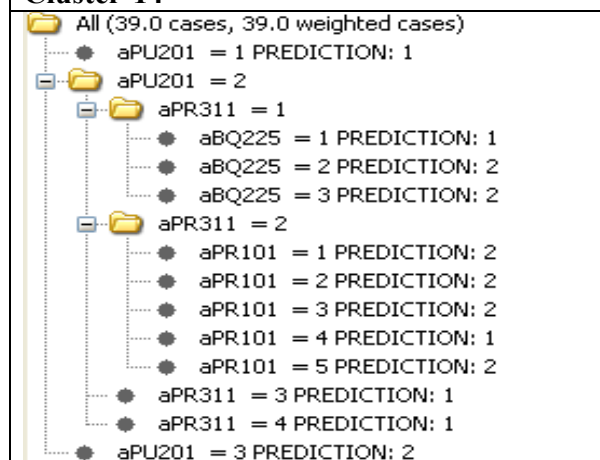
A cluster-based rule discovery model (CRDM) has been proposed for predicting smokers' quitting intentions that helps tobacco control policy makers in determining scientific-evidence based, cost-effective and adaptive policy. The cluster-based approach in CRDM is able to overcome the univariate problem of SDT for high dimensional data (such as tobacco control data). Experimental analysis on the real tobacco control data set shows that CRDM generates more simplified and compact decision rules than a single SDT for tobacco control system. In a single SDT, the tree size becomes large and the depth of the tree is high which means that very complex rules are generated. Moreover, the average prediction error rate of CRDM is less than a single SDT. In future we will apply CRDM on the other waves of ITC-4 survey and compare the result using some other clustering algorithms.

Table-1: Extracted rules from decision forest obtained by CRDM

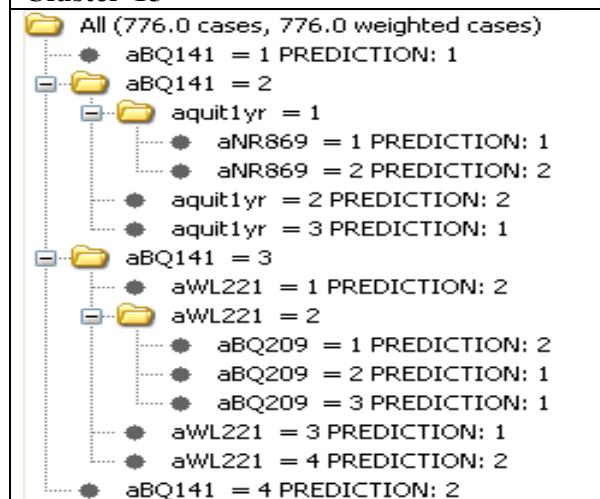




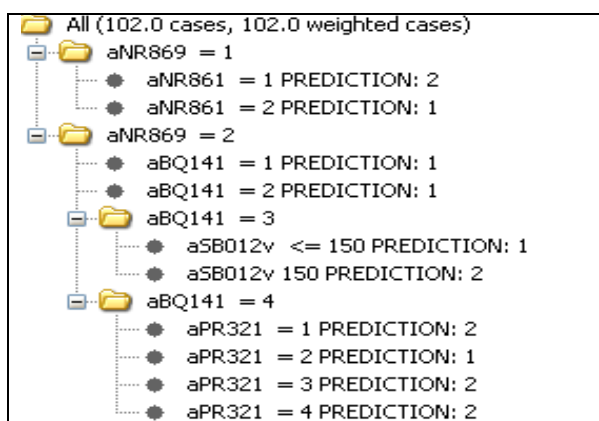
Cluster-14



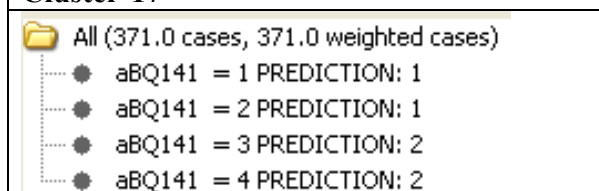
Cluster-15



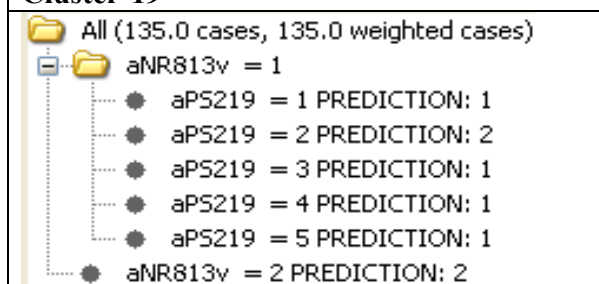
Cluster-16



Cluster-17



Cluster-19



Cluster-20

Table-1: Extracted rules from decision forest obtained by CRDM

REFERENCES:

[1] ITCEP, International Tobacco Control policy Evaluation Project (ITCEP), <http://www.itceproject.org/>.

[2] WHO, 2008, WHO REPORT on the global TOBA CCO epidemic, 2008, The MPOWER package http://www.who.int/tobacco/mpower/gtcr_download/en/index.html.

[3] NDSH, 2008, 2007 National Drug Strategy Household Survey, First results, April 2008, Australian Institute of Health and Welfare, Canberra, Cat. no. PHE 98, <http://www.aihw.gov.au/publications/phe/ndshs07-fr/ndshs07-fr-no-questionnaire.pdf>

[4] VTCS, 2008, Victorian Tobacco Control Strategy, 2008–2013, Report produced by Published by the Victorian Government Department of Human Services Melbourne, Victoria, Copyright State of Victoria 2008, <http://www.health.vic.gov.au/tobaccoreforms/downloads/vtcs0813.pdf>

[5] CCV, 2004, The Cancer Council Victoria 2004, Centre for Behavioural Research in Cancer, <http://www.cancervic.org.au..>

- [6] Germain, D., Wakefield, M. & Durkin, S. (2008) 2008 Smoking prevalence and consumption in Victoria: Key findings from 1998–2007 population surveys, Centre for Behavioural Research in Cancer, The Cancer Council of Victoria.
- [7] Begg S, Vos T, Barker B, Stevenson C, Stanley L, Lopez A. The burden of disease and injury in Australia 2003. PHE 82. Canberra: Australian Institute for Health and Welfare-2007. <http://www.aihw.gov.au/publications/index.cfm/title/10317>
- [8] Collins D, Lapsley H. The costs of tobacco, alcohol and illicit drug abuse to Australian society in 2004–05. P3 2625. Canberra: Department of Health and Ageing; 2008. [http://www.nationaldrugstrategy.gov.au/internet/drugstrategy/publishing.nsf/Content/mono64/\\$File/mono64.pdf](http://www.nationaldrugstrategy.gov.au/internet/drugstrategy/publishing.nsf/Content/mono64/$File/mono64.pdf)
- [9] X., J., Ding, S., Bedingfield et al, “A Decision Tree Approach for Predicting Smokers’ Quit Intentions”, Journal of Electronic Science and Technology of China, vol-6(3), pp 220-224.
- [10] X., J., Ding, S., Bedingfield et al, “A Rule Based Approach to Modeling the Effect of Tobacco Control Policies” 4th International conference on Information and Automation for sustainability(ICIAES), 2008, pp 496-501.
- [11] J. R. Quinlan, “C4.5 programs for machine learning”, Morgan Kaufmann, 1987
- [12] L. Rokach, O., Maimon, “Top-Down induction of decision trees classifier- A survey”, IEEE transaction on system, man and cybernetics Part-C, Vol-35(4), 2005, pp 478-487.
- [13] G. T. Fong, K. M. Cummings, R. Borland, G. B. Hastings, P. Hyland, G. A. Giovino, D. Hammond, and M. E. Thompson, “The conceptual framework of the International Tobacco Control (ITC), Policy Evaluation Project,” Tobacco Control, vol. 15, Suppl. 3, pp. 3-11, 2005.
- [14] M. E. Thompson, G. T. Fong, D. Hammond, C. Boudreau, P. Driezen, P. Hyland, R. Borland, K. M. Cummings, G. B. Hastings, M. Siahpush, A. M. Machintosh, and F. L. Laux, “Methods of the International Tobacco Control (ITC) Four Country Survey”, Tobacco Control, vol. 15, Suppl. 3, pp. 12-18, 2006.
- [15] A. Hyland, R. Borland, Q. Li, H-H. Yong, A. McNeill, G. T. Fong, R. J. O’Connor, and K. M. Cummings, “Individual- level predictors of cessation behaviours among participants in the International Tobacco Control (ITC) Four Country Survey”, Tobacco Control, vol. 15, Suppl. 3, pp. 83-94, 2006.
- [16] J. R. Quinlan, “Simplifying decision trees,” Int. J. Man-Mach. Studies, vol. 27, pp. 221–234, 1987.
- [17] F. Attneave, Applications of Information Theory to Psychology. New York: Holt, Rinehart and Winston, 1959.
- [18] R. Hathaway, J. Bezdek, and Y. Hu, “Generalized fuzzy c-means clustering strategies using L norm distances,” IEEE Trans. Fuzzy Syst., vol. 8, no. 5, pp. 576–582, Oct. 2000.
- [19] J. Mao and A. Jain, “A self-organizing network for hyperellipsoidal clustering (HEC),” IEEE Trans. Neural Netw., vol. 7, no. 1, pp. 16–29, Jan. 1996.
- [20] M. Eisen, P. Spellman, P. Brown, and D. Botstein, “Cluster analysis and display of genome-wide expression patterns,” in Proc. Nat. Acad. Sci.USA, vol. 95, 1998, pp. 14 863–14 868.
- [21] A.M. Bagirov, Modified global k-means algorithm for minimum sum-of-squares clustering problems, Pattern Recognition, Volume 41, 2008, pp 3192-3199.
- [22] A. Likas, N. Vlassis, and J. Verbeek, “The global K-means clustering algorithm,” Pattern Recognition, vol. 36, no. 2, pp. 451–461, 2003.
- [23] G. McLachlan and T. Krishnan, The EM Algorithm and Extensions. New York: Wiley, 1997.
- [24] S. Ruggieri, “Efficient C4.5”, IEEE transaction on knowledge and data engineering, vol-14 (2), 2002, pp 438-444.

Appendix: Attributes

Prediction=1 (Quit Attempt)

Prediction=2 (No Quit Attempt)

aBQ141

a. Are you planning to quit smoking:

01 – Within the next month?

02 – Within the next 6 months?

03 – Sometime in the future, beyond 6 months

04 – Not planning to quit

aFR250V

cigarettes per day

0= 1-10cigs, 1= 11-20cigs, 2= 21-30cigs, 3= 31+ cigs

aPU621

In the last 6 months, since, have you spent money on cigarettes that you knew would be better spent on household essentials like food.

01 – YES, 02 – NO

aNR861V

Since [LSD], have you received advice or information about quitting smoking from

Telephone or quit line services?

1-Yes, 2-No, 7- NA, 8-Refused, 9- Don't know

aNR815V

quitting RX from doctor, overall (incl those who did not visit the doctor)

AWL221

In the last month, have the warning labels stopped you from having a cigarette when you were about to smoke one? Would you say:

01 – Never, 02 – Once, 03 – A few times, 04 – Many times

aBQ201

In the past 6 months, have each of the following things led you to think about quitting, not at all, somewhat, or very much:

Concern for your personal health?

01 – Not at all, 02 – Somewhat, 03 – Very much

aFR309V

Smoking status: 1=daily, 2=weekly, 3=monthly, 4=quit<1mth, 5=quit 1-6m, 6=quit>6m

aNR817

pamphlet on quitting, from doctor, overall (incl those who did not visit the doctor)

aNR869

Since [LSD], have you received advice or information about quitting smoking from. Local stop-smoking services (such as clinics or specialists)?

aBQ225

In the past 6 months, have each of the following things led you to think about quitting, not at all, somewhat, or very much:

Advertisements or information about the health risks of smoking?

01 – Not at all, 02 – Somewhat, 03 – Very much)

aPS215

Please tell me whether you strongly agree, agree, neither agree nor disagree, disagree, or strongly disagree with each of the following statements. If you had to do it over again, you would not have started smoking.

01 – Strongly agree, 02 – Agree, 03 – Neither agree nor disagree, 04 – Disagree, 05 – Strongly disagree

01 – Strongly agree, 02 – Agree, 03 – Neither agree nor disagree, 04 – Disagree, 05 – Strongly disagree

aNR813v

referral from doctor to help stay quit, overall (incl those who did not visit the doctor)

aBQ203

In the past 6 months, have each of the following things led you to think about quitting, not at all, somewhat, or very much: Concern about the effect of your cigarette smoke on non-smokers? 01 – Not at all, 02 – Somewhat, 03 – Very much

aBQ223

In the past 6 months, have each of the following things led you to think about quitting, not at all, somewhat, or very much: Availability of telephone

helpline/Quitline/information line?

01 – Not at all, 02 – Somewhat, 03 – Very much

aFR260v

Derived variable: Heaviness of smoking index

aPS211

Please tell me whether you strongly agree, agree, neither agree nor disagree, disagree, or strongly disagree with each of the following statements. You enjoy smoking too much to give it up. 01 – Strongly agree, 02 – Agree

03 – Neither agree nor disagree, 04 – Disagree

05 – Strongly disagree,

aET221

Which of the following best describes smoking in your home? (read)

01 – Smoking is allowed anywhere in your home, 02 – Smoking is never allowed anywhere in your home, 03 – Something in between

aSB031

Do you consider yourself addicted to cigarettes? 01 – Not at all, 02 – Yes–somewhat addicted, 03 – Yes–very addicted

aFR245

On average, how many cigarettes do you smoke each day/week, including both factory-made and roll-your own cigarettes?

aBQ209

In the past 6 months, have each of the following things led you to think about quitting, not at all, somewhat, or very much:

The price of cigarettes? 01 – Not at all

02 – Somewhat, 03 – Very much

aPU555

Calculated variable: price per unit, regardless of packaging.

Aquit1yr

Tried to quit in the last year: 1=never, 2=tried>1yr ago, 3=tried within last year

aSB221

In the last month—since [1M Anchor], have you

[AUS/UK=stubbed]

[CAN/US/=butted] out a cigarette before you finished it because you thought about the harm of smoking?

01 – YES, 02 – NO

aSB012v

total minutes to first cigarette (continuous)

aDE312v

Education 3-High; 2-Medium; 1-Low

aBQ221

In the past 6 months, have each of the following things led you to think about quitting, not at all, somewhat, or very much:

Free or lower-cost stop-smoking medication?

01 – Not at all, 02 – Somewhat, 03 – Very much

In the past 6 months, have each of the following things led you to think about quitting, not at all, somewhat, or very much:

aBQ229

Setting an example for children?

aBQ201

Concern for your personal health?

01 – Not at all, 02 – Somewhat, 03 – Very much

aSB226

a. In the last month—since, have you

[AUS/UK=stubbed]

[CAN/US/=butted] out a cigarette before you finished it because you thought about the harm of smoking?

b. Was that once, a few times, or lots of times?

01 – Once, 02 – A few times, 03 – Lots of times

aPS223

Smoking is an important part of your life.

Please tell me whether you strongly agree, agree, neither agree nor disagree, disagree, or strongly disagree with each of the following statements. 01 – Strongly agree 02 – Agree 03 – Neither agree nor disagree 04 – Disagree 05 – Strongly disagree

05 – Strongly disagree

ABQ121

How easy or hard would it be for you to completely quit smoking if you wanted to?

01 – Very easy, 02 – Somewhat easy, 03 – Neither easy nor hard, 04 – Somewhat hard 05 – Very hard

APS229

Please tell me whether you strongly agree, agree, neither agree nor disagree, disagree, or strongly disagree with each of the following statements. People who are important to you believe that you should not smoke. 01 – Strongly agree, 02 – Agree, 03 – Neither agree nor disagree, 04 – Disagree, 05 – Strongly disagree

aWL341

In the last month, have you made any effort to avoid looking at or thinking about the warning labels: by not buying packs with particular labels?

01 – YES, 02 – NO

aWL211

In the last month, how often, if at all, have you read or looked closely at the warning labels on cigarette packages?

01 – Never, 02 – Rarely, 03 – Sometimes

04 – Often, 05 – Very often

aPS227

Please tell me whether you strongly agree, agree, neither agree nor disagree, disagree, or strongly disagree with each of the following statements. You have strong mixed emotions both for and against smoking, all at the same time.

01 – Strongly agree, 02 – Agree, 03 – Neither agree nor disagree, 04 – Disagree, 05 – Strongly disagree

aSB013v

Minutes to first cigarette

aSB205

The following questions ask you about how often you've had certain thoughts in the last month, that is, . For each question, please answer using. Think about the harm your smoking might be doing to you?

01 – Never, 02 – Rarely, 03 – Sometimes

04 – Often, 05 – Very Often

aPS229

People who are important to you believe that you should not smoke.