Nayak, Richi and Buys, Laurie and Lovie-Kitchin, Jan E. (2006) Influencing Factors in Achieving Active Ageing. In *Proceedings 6th IEEE International Conference on Data Mining - Workshops (ICDMW)*, pages pp. 858-862, Hong Kong.

Influencing Factors in Achieving Active Ageing

Richi Nayak
School of Information Systems
Queensland University of Technology
Brisbane, Australia
r.nayak@qut.edu.au

Laurie Buys
Centre for Social Change Research
Queensland University of Technology
Brisbane, Australia

Jan Lovie-Kitchin
Faculty of Health
Queensland University of Technology

Abstract

The Triple A project at QUT performed a national wide postal survey of older australian on a wide range of 'work', 'learning', 'social', 'spiritual', 'emotional', 'health and home', 'life events' and 'demographics' variables. With the use of predictive modelling techniques, this paper examines the impact of these interrelated variables on older people. The predictive modeling results successfully highlight interesting trends in the data and describe the significance of different factors in achieving active ageing.

1. Introduction

With continuous increase in the numbers and percentages of the older population in most countries over the next three or five decade, ageing is now viewed as a priority issue (Council in the Ageing, 2001; Kinsella and Phillips, 2005; WHO, 2005). The interest in developing a concept 'Active Ageing' reflects changes to prevailing theories of social and psychological aspects of ageing indicative of a focus on people's strengths as opposed to their deficits or pathology (Beardon, 1996).

The Australian Active Ageing (Triple A) project at Queensland University of Technology (QUT) conducted a national-wide postal survey to collect responses of older people on a wide range of questions related to 'work', 'learning', 'social', 'spiritual', 'emotional', 'health and home', 'life events' and 'demographics'. This dataset is the first of its kind reflecting a wide variety of aspects of older adults life. Previously, studies such as the Australian Longitudinal Study of Ageing (ALSA) (Andrews et al. 2002), and the Dubbo Study of Ageing (Simons et al. 1990) focused on social aspects as well as the more usual psychological and behavioral (ALSA), and bio-medical issues (Dubbo Study) only.

This paper attempts to conceptualize active ageing in terms of complex issues that intertwine and converge with the ageing experience, rather than in singular health or social dimensions. We apply the predictive data mining technique to examine the large number of interrelated variables and the complex relationships between aspects of older people's experience. The results successfully highlight interesting trends in the data and describe the significance of different factors in respect to quality of life and engagement in life. The trends and patterns of this data set will assist to develop the concept of active ageing and contribute significantly to future policy directions.

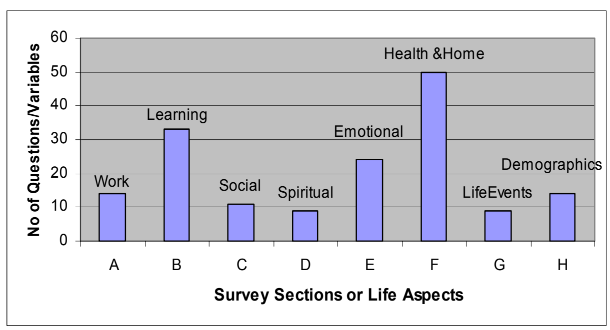

Figure 1: Distribution of Survey Questions/ Variables

2. Data Pre-processing

The Triple A project compiled a survey that includes responses of members of a large Australia wide seniors organization on a wide range of questions related to their life. Figure 1 shows the areas of variables covering various aspects of older adults' life, and their overall distribution in the data set. Survey

Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06)
0-7695-2702-7/06 $20.00

questions on 'learning', 'social', 'emotional', 'health', 'home' and 'life events' aspects were adapted from existing studies. The team developed survey questions on 'work', 'spiritual' and 'demographics'.

A total of 2655 surveys were returned at 46% response rate. Of these 32 incomplete surveys were excluded due to missing age, postcode and/or non-response to survey questions; others were excluded due to respondents being younger than 50 years old. The sample included an expected balance of gender (57% female) and state representation (NSW: 32%, Vic: 24% and Qld: 20%). Most of the respondents expected to or were living on a pension, but in general their financial status was some what higher than the national average. The majority were living with a companion, in their own home in a metropolitan area and were well educated. The distribution analysis showed us that the data set is suitable for predictive data mining techniques to be applied.

A major problem in preparing data for mining was that some questions requested the respondents to skip questions depending on their response to certain questions. These variables were subsequently modified to limit the effect of the large percentage of respondents who didn't answer them. For example, variable A1 requested respondents to skip to A4 if they answered no, which led to a large percentage of respondents leaving variables A2 and A3 blank. Therefore A2 and A3 were modified to include a response of N/A for those respondents who were not required to do so.

There were some variables that were transformed to fulfil the goal of the mining process. For example, the variable H2 asked respondents to supply their age. Since the Triple A project is only interested in finding for age groups, therefore their responses were modified to reflect the group that they belong to (55-64; 65-74; 75 years and over). Similarly, the variable H3 asked respondents to supply their postcode. We found that the postcode had little effect on the data, and therefore used it to calculate the state that the respondent lived in, whether they lived in a metropolitan or regional area, and to which socio-economic status they belong.

3. Conducting Predictive Data Mining

The processed data consists of 2,623 respondents and 165 variables/attributes. Majority of attributes are categorical. The experiments are conducted with 'SAS 9.1 Enterprise Miner'.

We applied predictive (classification) data mining to understand the deep relationships between various sections and various variables. We had many choices for predictive modeling. Since, our main goal is to understand how the variables from different life aspects are related to each other, good comprehensibility and high accuracy were the main requirements in a method. *Decision tree* is chosen due to their ability to obtain reasonable accuracy, good comprehensibility, efficiency, robustness and scalability.

We wanted to examine how a variable in one section (or life aspect) is related to other variables. The challenge was how to choose the target variables. We have chosen four of the target variables in each life aspect based on those who have equal distribution among their representative class values.

The Figure 2 shows the details of the classifiers for each target with the average performance on 10-fold CV experiments. We report the number of variables contained in the rules, the number of classification rules, and the misclassification rate for the classifiers. The relatively small misclassification rates (26 out of 32 have < 30% misclassification) show that classifiers can be found that can accurately describe some of the data. This indicates that there are strong patterns within the data.

A large number of variables appearances in rule sets show the higher dependency of variables in the data set. This in some extent proves that active ageing does not just depend on health or social issues, but, it is a complex process and depends on all aspects of life.

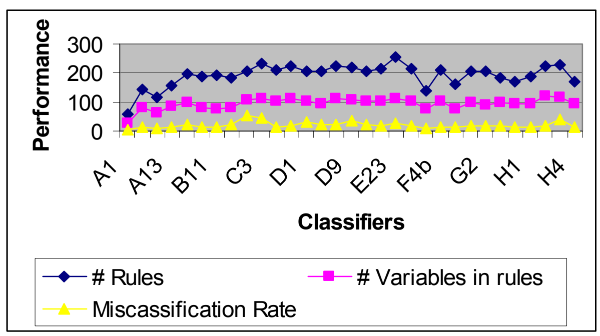

Figure 2: Classifiers Details

The results in Table 1 show the percentage of variables from each aspect that forms the classifier for the given target. The average of the targets for each aspect is calculated by dividing the number of variables for the given aspect by the number of variables required to classify the target. The "Variables" row shows the percentage of variables that exist in each section (or life aspect) in the survey. The "Mean" row shows the mean percentage of dependence for each aspect based on experiments.

The mean row shows that the Learning, Emotional, and Health and Home aspects (B, E, F) are the most significant, as combined they contribute 64.49% of all classification variables. This indicates that physical and emotional health combined with the desire to learn are the most significant factors when considering active ageing. However, these aspects also contribute 65.24% of the total variables in the data set so this could be expected. It is therefore necessary to conduct further analysis to measure the significance of each life aspect.

Target	Section Breakdown							
Section Average	**A Work**	**B Learning**	**C Social**	**D Spiritual**	**E Emotional**	**F Health & Home**	**G Life Events**	**H Demographics**
Variables	8.54%	20.12%	6.71%	5.49%	14.63%	30.49%	5.49%	8.54%
A	12.48%	20.76%	9.30%	5.19%	14.07%	18.46%	7.03%	12.71%
B	10.99%	27.36%	6.20%	5.58%	19.78%	19.15%	5.38%	5.56%
C	8.91%	23.83%	7.99%	5.08%	18.33%	25.86%	4.08%	5.93%
D	8.84%	22.93%	8.02%	6.28%	17.58%	24.17%	4.74%	7.44%
E	6.93%	23.11%	7.42%	6.54%	19.48%	24.23%	4.66%	7.64%
F	8.53%	20.87%	7.32%	5.94%	15.06%	29.88%	4.75%	7.65%
G	9.23%	23.67%	8.47%	4.71%	15.11%	26.29%	6.01%	6.50%
H	9.13%	22.25%	6.65%	5.96%	19.55%	24.10%	4.53%	7.84%
Mean	9.38%	23.10%	7.67%	5.66%	17.37%	24.02%	5.15%	7.66%

Table 1: Average Section Dependencies

	Misclassification Rate for Classifier								
Target	**Base**	**A Work**	**B Learning**	**C Social**	**D Spiritual**	**E Emotional**	**F Health & Home**	**G Life Events**	**H Demographics**
A1	3.33%	4.97%	30.34%	36.76%	39.90%	32.13%	30.45%	42.58%	17.98%
B11	14.66%	39.81%	**16.89%**	41.69%	42.99%	34.20%	35.70%	45.80%	41.69%
C5	14.78%	31.04%	25.04%	21.43%	28.99%	23.63%	23.18%	32.10%	31.53%
D3	20.15%	41.22%	32.40%	37.00%	28.78%	27.22%	30.27%	48.56%	43.04%
E12	18.70%	32.61%	26.63%	30.90%	30.44%	**21.75%**	26.29%	34.74%	33.64%
F3g	10.09%	24.51%	18.58%	24.59%	25.77%	20.63%	**12.11%**	24.74%	23.56%
G6	12.67%	**15.66%**	**14.82%**	**15.93%**	16.51%	**14.86%**	**14.36%**	**16.28%**	**15.97%**
H4	12.97%	19.84%	**15.63%**	18.89%	19.95%	17.15%	17.30%	21.40%	19.12%

Table 2: Classifier versus Section

The results in Table 1 that are shaded dark grey highlight results that have a 20% or more difference from the "Variables" row. These indicate that the targets from the given aspect rely on the prescribed aspect more than is statistically expected. Therefore the variables from those aspects are quite important in determining those targets. Targets from aspect A are statistically more reliant on variables in aspects A, C, G, and H and less reliant on aspect F. Targets from aspect B are statistically more reliant on variables in aspect A, B, and E and less reliant on aspects F and H. Targets from aspect C are statistically more reliant on variables in aspect E, and less reliant on aspects G and H. Targets from aspects D, E, and H are statistically more reliant on variables in aspect E, and statistically less reliant on aspect F. Targets from aspect G are statistically more reliant of variables in aspect H. Therefore it can be concluded that variables in E (Emotional) aspect are statistically more significant, and variables in F (Health & Home) are statistically less significant than any other aspects (according to the difference in values of "Variables" and "Mean" row).

Despite this it is still evident that aspects 'Learning', 'Emotional', and 'Health and Home' (B, E, and F) are the most important in this survey.

Further tests were performed on the most accurate classifier (selected from Figure 2) built for each aspect to investigate the dependence on variables from other aspects in classifying the same target. The results in Table 2 compare the "Base" misclassification rate versus classifiers reliant on only variables from the specified aspect. The "Base" misclassification rate is obtained from the Figure 2 for the classifier when it includes variables from all sections. The results highlighted in grey indicate the classifiers that are within 20% of the most accurate aspect dependent classifier for the target. These results show that every target except G6 and H4 produce significantly more

Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06)
0-7695-2702-7/06 $20.00 © 2006 IEEE

accurate classifiers when only relying on variables from their respective aspect. The G6 and H4 targets can be accurately predicted (<= 21.4% misclassification) using variables from any aspect. A total of 81.81% of respondents answered "No" to G6: "Were they involved in any childcare activities in the past year?" and 76.07% of respondents answered "Australia" to H4: "Country of Birth". Therefore when such a high percentage of respondents answered the same way it is highly likely that strong relationships exist between many variables that can explain the data.

The results in bold indicate the classifiers that are within 30% of the accuracy of the base classifier. The G6 target is within 30% for every aspect except section D, which reiterates the previous conclusion that G6 can be explained by several variables. However, excluding G6, only B11, E12, and F3g have classifiers that rely on their own aspect so strongly. Therefore it is reasonable to assume that strong correlations exist within an aspect. Again 'Learning', 'Emotional', and 'Health and Home' (B, E, and F) appear to be relied upon the most heavily.

We analyzed several classifiers to actually see the impact of various variables on positive responses. We only report some of the discoveries. The rules infer that the working status of people highly depends upon how they feel about their life. If they generally feel confident and positive about themselves, it is reflected in their working life. Their learning desires such as if they have a personal goal to achieve and want to keep up-to-date with technology, make them to enjoy the work. Data mining techniques however did not draw a direct link between enjoying the work and health limitations.

People regarded self motivation, obstacles in having transport, general health and physical health limitations as main factors to keep them away from learning new things. In addition, people over 70 years old considered mainly their memory and transport obstacles are barriers to learning new things.

Attributes like 'engagement in social voluntary work', 'feeling family member and friends understand them or not', 'age group' and 'whether they feel contented in their present life are' playing a very important role in determining the current situation of older people with their social network support.

Attributes like 'age group', 'whether they feel control of their life', 'relationship with others', 'interests on learning' and 'accepting new things' are important in determining autonomy, personal growth and purpose of life.

4. Discussion and Conclusion

With the use of predictive data mining techniques, this paper examined the inter-relationships of a wide range of 'work', 'learning', 'social', 'spiritual', 'emotional', 'health and home', 'life events' and 'demographics' variables in order to identify those that contributed most strongly to positive responses and could therefore be indicators of "Active Ageing".

A number of classifiers or predictors were built to explore the relationships between various variables targeted to certain variables. The main goal of this application was to understand the survey data and learn if there are patterns and generalized rules reflecting the population. This emphasizes the need for good comprehensibility of output results, which demonstrates the reason that decision trees were chosen for classification analysis.

Data mining analysis successfully highlighted complex issues that intertwine and converge with the ageing experience, rather than in singular health or economic dimensions. Our results present a portrait of Australian older adults that is distinctly different from the stereotype generated by models which negatively portray ageing as a process of decline.

We further tested to check which sections in the survey (or domains of life for older people) are more strongly associated. Results show that the 'Learning', 'Emotional', and 'Health and Home' are more significant, as combined they contribute 64.49% of all classification variables. This reinforces that physical and emotional health combined with the desire to learn are significant factors when considering the positive relationship to well being.

Further analysis shows that 'Emotional' feelings are statistically more significant, and 'Health and Home' is statistically less significant than any other aspect. Factors related to 'Learning' are also significant. A majority of rules have dominance of emotional feelings, and learning needs and interests in explaining active ageing concepts. For example, the results strongly suggest that people feel contented in their current life when they have more enjoyment engaging in social activities and higher intention of learning new things.

The identified relationships assist us to better understand the impact of ageing upon older Australians' engagement in society and their general sense of well being. This study viewed the ageing from a developmental perspective so that the identified rules and trends can be used to contribute to the conceptualizing and understanding of 'active ageing' through the significant relationships required for quality of life. The identified patterns will help policy

Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06)
0-7695-2702-7/06 $20.00 © 2006 IEEE

makers and service providers to provide services that meet the needs of older people in the future.

Our survey data includes a homogenous population of respondents in terms of education and finance status. For more insight, we need to mix various populations such as those with diverse living status, lower education, lower financial status etc. Regardless, the decision trees were able to capture the essence of the data set to reflect general positive engagement in life.

The aim of this paper was to highlight interesting trends in the data, though future research will more comprehensively describe the significance of different factors in respect to conceptualizing active ageing. These trends in the data will form the basis to do more in-depth qualitative study.

5. References

[1] Agrawal, R., & Srikant, R. (1994). *Fast Algorithms for Mining Association Rules*. IBM Research Report RJ9839, IBM Almaden Research Center.

[2] Andrews G, Clark M, Luszcz M. (2002) Successful aging in the Australian Longitudinal Study of Aging: Applying the MacArthur model cross-nationally. Journal of Social Issues; 58(4): 749-765.

[3] Beardon L. (1996). Successful aging: What does the "good life" look like? NC State University. *The Forum For Family & Consumer Issues* 1: 3. [Accessed 17/2/2003] Available online at: http://www.ces.nesu.edu/depts/fcs/pub/aging.htm

[4] Council on the Ageing (2001). *Strategic Ageing: Australian issues in Ageing.* Occasional papers. [Accessed 21/2/2003] Available online at: http://www.cota.org.au/stratage.htm

[5] Han J. (2001) *Data Mining: Concepts and Techniques* Morgan Kaufmann Publishers, San Francisco

[6] Kinsella K. Phillips, DR (2005). Global aging: The challenge of success. Population Bulletin 60(1): 3-40.

[7] Lim, T. S. & Loh, W. Y. (2000). A comparison of prediction accuracy, complexity and training time of thirty three old and new classification algorithms. *Machine Learning, 40*(3), sep. 203-228.

[8] Simons LA, McCallum J, Simons J, Powell I, Ruys J, Heller RLC (1990). The Dubbo study: An Australian prospective community study of the health of elderly. Australian & New Zealand Journal of Medicine; 20: 783-789.

[9] World Health Organisation [WHO]. Active Ageing: A Policy Framework. [Online] [Accessed 2005 April]. Available from www.who.int/hpr/ageing/ActiveAgeingPolicyFrame.pdf

[10] Anderson, R.E. Social impacts of computing: Codes of professional ethics. Social Science Computing Review, 2 (Winter 1992), 453-469.

6. Acknowledgement

This work was supported by the QUT ECR and the QUT Strategic Collaborative Grant Scheme. We would like to thank all the team members of the Triple A project. We would also like to thank Calum Robertson, Lei Zhang, Xin Li and Ying Luo for their assistance in conducting experiments.