IDENTIFYING DIFFERENCES IN SAFE ROADS AND CRASH PRONE ROADS USING CLUSTERING DATA MINING

Emerson D [a], Nayak R [a], Weligamage J [b].

a. Computer Science Discipline, Faculty of Science and Technology, Queensland University of Technology, Brisbane, Queensland, Australia.

b. Road Asset Management Branch, Queensland Government Department of Main Roads, Brisbane, Queensland, Australia.

ABSTRACT

Road asset managers are overwhelmed with a high volume of raw data which they need to process and utilise in supporting their decision making. This paper presents a method that processes road-crash data of a whole road network and exposes hidden value inherent in the data by deploying the clustering data mining method. The goal of the method is to partition the road network into a set of groups (classes) based on common data and characterise the class crash types to produce a crash profiles for each cluster. By comparing similar road classes with differing crash types and rates, insight can be gained into these differences that are caused by the particular characteristics of their roads. These differences can be used as evidence in knowledge development and decision support.

1 INTRODUCTION

Road safety is a major concern worldwide with road crashes costing countries between one and three percent of annual Gross Domestic Product. WHO predicts road traffic crashes emerging as the 3rd leading cause of disease or injury burden [1]. The annual economic cost of road crashes in Australia conservatively estimated at \$18 billion per annum, with crashes having the potential for devastating social impact [2]. World-wide, road authorities follow strict design codes applying known engineering principles within the contexts of safety, cost, driver expectation, and economic and environmental parameters [3]. Research is continually contributing to development of these principles.

Historically, statistically studies have been used to analyse crashes from homogeneous road subsets [4], however there is a demand for whole of network analysis. An Australian road research organization, the ARRB Group contends that use of safety risk assessment at the network level (rather than concentrating on single sites) can apply consistency in the approach to road safety across the entire network [5]. Depair el el (2006) provide the theoretical framework and a case study for use of probability-based clustering data mining for analysing crashes from heterogeneous road sets [6], but the case study is limited to a set of urban roads.

This paper evaluates an applied data mining approach to using clustering data mining using the probabilistic expectation maximization method [6] on a large, diverse rural-urban network, and shows clustering's value by presenting the practical decision support outcomes of the case study. The method uses relationships between road-related attributes and crash types to build clusters and document a profile of individual cluster crash characteristics, which subsequently can be used to define the risks faced by motorists as they travel a particular group of road segments. This paper describes the clustering data mining approach used to develop road clusters by deploying the following road and crash characteristics as input variables: surface measures, surface type, age, wear, damage, road geometry, traffic and crash type. Clustering, using the selected attributes, develops natural groups in an unsupervised fashion, then assigns instances to those clusters, resulting in grouping of instances according to their attributes. The dimensionality of the attribute value range active in each cluster is reduced, and the clusters make new datasets suitable for statistical analysis [6].

Data mining methods are increasingly in demand in business situations [7] and are being applied in many innovative ways in road crash studies [8-9]. Anderson [10] recently used clustering in a road crash study to analyse crash hotspots using spatial density. Our study's whole of network approach classified four year of crashes, their 1 km road segments, and the road segment crash counts [11]. These crash, road, traffic and conditions characteristics provided the input variables for the clustering models. Studies such as Quin [12] demonstrate that road factors influence crashes, with significant factors including speed limit, traffic rates, time of day, volume/capacity ratio, percent of passing lanes, shoulder width, number of intersections and driveways. These characteristics coincide with many of the variables in our dataset.

The incentive to investigate data mining clustering as a method of road & crash benchmarking evolved from our preliminary clustering of road characteristics during an investigation into road segment crash proneness [13]. The study showed that road segments clusters developed from road attributes had a well-differentiated crash counts range, characterising each cluster's into a category of very low, moderate or high crash.

The contribution of this paper is to demonstrate that clustering methods can be successfully applied to whole-of-network road-crash data for the development of useful road cluster crash profiles for use in decision support, plus provide guidelines for method deployment. The paper describes development and outcomes of the study in the following sections. Section 2 describes the data organization and pre-processing, while section 3 describes the data mining methodology and method assessment. Section 4 discusses the methods used in the case study to analyse, organize and present information, and demonstrates use of the road segment crash profiles in decision making. The conclusion provides the summary of the outcomes and proposes future directions.

2 DATA SET DEVELOPMENT

The dataset is four year list of 2004 -2007 crash instances populated with the road data, provided by the Queensland Department of Transport and Main Roads (QDTMR). Of the 42,388 crashes reported, sufficient data was available for inclusion of one third of the crashes for analysis. The attributes provided the necessary road and traffic information to develop a series of road cluster types with sufficient differentiation to allow comparison between high risk and low risk road clusters and their crash types. Road attributes are modelled in the following logical groups:

- Road surface, design and wear. (19 attributes) included *roadway surface with* road surface friction*,* texture depth, seal age and seal type; *road surface wear and damage with* rutting and roughness; *roadway design:* with speed limit, traffic, divided road, dual or single carriageway, road type such as highway, urban arterial etc, carriageway type and the lane count; *roadway features* such as roundabouts, bridges, intersections; traffic control with tights, signs etc; *traffic rates*; *geometry with* horizontal and vertical alignment and terrain
- Roadway and cracking. (24 attributes) consisting of the attributes above and five cracking attributes
- Roadway and crash details (32 attributes): addition of *crash details* such as crash type, road surface wetness, monthly rainfall, day, time, and lighting.

An standardized copy of the data was generated for modelling. The natural range of the AADT attribute, from 61 to 64,452 vehicles per day, is many magnitudes higher than road friction (F60) ranging from 0.08 - 0.66. Standardized ranges were -1.077 to 4.195 vs. -4.462 to 4.185 respectively. Cluster models were developed with both the original non-standardized and processed standardized versions and compared.

3 DATA MINING METHODOLOGY, RESULTS AND DISCUSSION

This section examines how the clustering methods were configured and discusses the results of the trials. The clustering application selected was the probabilistic method called *expectation maximization* (EM) [6]. The probabilistic method allocates instances to clusters on the highest likelihood of an instance being in a given cluster. EM provides a model goodness measure, *log likelihood.* Our study progressed through two stages. In stage one we performed clustering benchmarking to get to know how the data would perform with the EM clustering method, and to assess the usefulness of resulting clusters. Stage two saw benchmarking of the EM configuration parameters in conjunction with the automated cluster count determination function.

In the stage 1 we ran a series of tests with combinations of the road & crash attributes, cluster counts and standardized and non-standardized data. Cluster count can be set in the EM application configuration interface, and clusters were created with the individual cluster count ranging between 2 and 16 clusters. The models were evaluated in three ways: by comparing the method's output statistics, analysing the cluster structure and comparing how well, for a given trial, the clusters differentiated themselves. Extensive visualization was used to compare the clustered attributes values to evaluate usefulness of a given model in its relative contribution to the road domain knowledge.

Results showed that the model quality statistic *log likelihood* was not very useful determining the best models and that producing a standardized dataset had only marginal benefit. Models from standardized and non-standardized datasets in most cases produced identical clusters, while having wildly different log likelihood values; -22.9 vs. -55.5 for standardized vs. non-standardized datasets.

While searching for the optimal seed configuration value, we found that the variation of seed affected the standardized or non-standardized datasets differently, producing clusters with identical instance allocation in some cases and different clusters in others. We favoured seed values that produced identical clusters.

Completion of benchmarking at stage 1 provided a set of known clusters, useful for later comparisons in assessment of procedural accuracy and configuration for model quality and suitability.

In stage 2 we sought the ideal clustering count, however the problem of the cluster count dependency on the seed value needed resolution. We used the traditional data mining *validation method* of developing a model by training on 66% of the instances and validating the model by testing on the remaining data subset. While EM has its own internal cross validation method to progressively develop and test the model, the training / testing method was extremely valuable in selecting the best seed. An optimal cluster count and seed value was selected for each of three standard attribute sets, based on seeds where the cluster numbers and instance allocations from both training and testing were identical.

For the *road only* dataset the seed value of 100 produced a cluster model with 12 groups. The *road and cracking* dataset the seed of 75 produced 15 clusters. With *road, cracking and crash* dataset, the seed of 75 produced 8 clusters. The unexpected reduction in cluster count with the inclusion of crash data is thought to be caused by the similarity in crash outcomes for some clusters causing them to coalescence, but this requires conformation.

Once the prime benchmarked clusters were documented, a range of new clusters were generated and evaluated. Configuring a model for a higher cluster count caused individual clusters to split and produce cluster subgroups. The split was created on a significant attribute and results were found to be useful for examination of topics such as serious crashes and wet crashes.

Using model statistics in comparing the performance of our EM implementation with the Depaire's work [6] was difficult because of different measures in use, and also in our study the log likelihood, while indicating cluster compactness did not relate to the actual cluster formations and instance allocations. However comparing models at the information level showed that our clusters, like Depaire's, were able to strongly differentiate on the available attributes; in our case with road type and function, crash types, road surface wetness, speed limit, lane counts, crash type and so on.

4 CASE STUDY: METHOD, RESULTS AND DISCUSSION

The third stage describes how the clustering models were investigated and presents a subset of results from one of our case studies to illustrate how intelligent partitioning of the data can be used in knowledge management and decision making. As with many road studies, this study is driven by goals of discovering how roads contribute to crashes, and how we can use outcomes to improve road safety. Clustering provides an extremely valuable tool to partition the dataset into instance groups with a strong relationship to the attribute properties, and provided clusters relating to our topics of investigation: wet crash and serious crash. Since this was a preliminary clustering study with this dataset, we were investigating ways to extract the value from the models to contribute to the goals. Our investigations included;

- Visualizing cluster distributions geographically to contextualize them spatially, e.g. plotting the crashes by latitude and longitude to show geographic cluster distribution and allow relation to all other temporal-spatial data.
- Using the characteristics to name the cluster and develop named cluster groups e.g. formation of the *city urban and regional roads cluster group* from a number of similar clusters.
- Plotting clusters by attributes to gather evidence on the big questions of road crash causation, e.g. by plotting cluster vs. seal type and relating the attribute range to the cluster's significant properties such as the relative cluster crash rate.
- Examining the built in cluster/attribute cross tabulation reports which document the relative contribution of attribute values to the individual clusters and using these values in developing cluster profiles. e.g. rural highway clusters were found that have high vehicle collision with stationary objects and high vehicle overturn type of crash. We used these relative crash values to create cluster crash profiles.
- Filtering clusters by attribute values of interest and using cluster attribute distributions to describe the entity characteristics, e.g. obtaining the crash instance cluster types for a particular road or road type and using the cluster crash profiles to understand the crash characteristics of the road.

Clusters provided a range of new ways to characterize road types and provided new data and evidence for development and evaluation of the inferences that help make our roads safer.

The following section presents some of the outcomes of our preliminary case study. This clustering method was configured to produce an model extended from the 8 main classes identified as the prime clusters, and we configured this model with 14 clusters, with the purpose of splitting the default clusters in the search of knowledge relating topics highlighted above: wet and serious crash.

The model was developed from 43 road-crash attributes, using non-standardized data with natural values to produced a comprehensive cluster model with human decipherable reports. The model was generated by the EM algorithm as a testing set, using 5,695 randomly selected instances of the total 16,750 crashes. The model develops cluster distributions of crashes for the whole main roads network and nicely resolved cluster groups with differing crash profiles. Results of the case study are shown in the appendix in Table 2

The geographic distribution of all clusters plotted on latitude and longitude is shown in Figure 1. Brisbane to Cairns defines the eastern seaboard freeways and Mackay/Townsville to Mt Isa shows the central inland road systems. The emergent patterns differentiate rural highways from the more complex city and regional centres.

Tabulation and correlation of the common characteristics of the of each cluster lead to grouping of common clusters (sample groups in Table 1). Analysis showed that clusters were strongly influenced by the road seal type, speed limit, design features such as the number of lanes, single or dual carriageway and the roadway features such as intersections, roundabouts and so on. Each cluster group maintained a characteristic road segment *average* crash rate ranging from high (8 crashes per year) to low (less than 1 crash per year). Table 1 presents two of the five cluster groups devised to demonstrate the interaction between crash types, crash rates and road attributes.

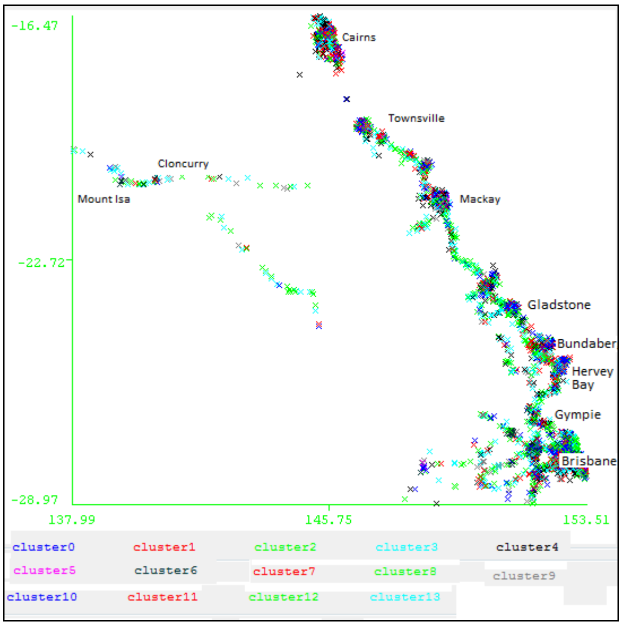


Figure 1 Geographic distribution of main road crash clusters

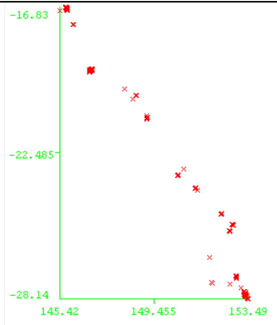
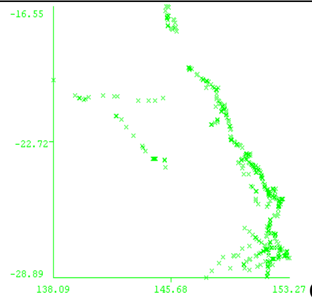
Sample cluster geographic distribution	I km road segment cluster profiles
 Cluster Group A	**Cluster Group A**: Main road / highway, proximity to city. **Clusters 1,5,6** **Road profile:** Seal type of **dense graded asphalt**, with daily traffic rates of between 13k and 17k vehicles per day. Predominantly single carriageway, high percentage of divided road of both 2 and 4 lane @ **60-80km/h.** **Annual average crash counts: are the highest** ranging between 5 and 8 crashes per road segment with a maximum of 35. Serious crash averaged: average between 17% and 22% **Crash Profiles: high levels of crossings and t-junctions involved** in crashes, with angle and rear collisions being the predominant crash types. Wet crash: low with av. values 3% to 12%. Wet crash were a low percentage between 0-9%
 Cluster Group D	**Cluster Group D** :Freeways and Main roads **Clusters 2,13** **Road profile:** Seal type of **spray seal**, with low daily traffic rates of between 3.2k and 3.5k vehicles per day. Predominantly single carriageway, very high percentage of divided road and exclusively 2 lane @ **80-110km/h** **Annual average crash counts: are the lowest** ranging between 0.5 and 0.8 crashes per road segment with a maximum of 32. **Serious crash rate** is the highest with cluster averages between 33% and 100% **Crash Profiles:** highest percentage of **road features not involved** in crash (97%)with relative high levels of high levels of hit stationary object, overturned, head-on being the predominant crash types.

Table 1 Samples cluster groups devised from the fourteen clusters in the model.

Cluster distributions were closely related to the distribution of the seal types, roadway features, roles of the roads and settlement patterns.

The chart showing relative cluster crash characteristics (Figure 2) clearly demonstrates that each road-crash cluster has its own characteristic set of crashes, and our development of individual cluster crash profiles was based on this results.

This outcome demonstrates that clustering method meets the objective of this paper, to develop a crash profile for each of the cluster groups for use in decision support. The following examples illustrate some differences in the clusters. Freeway and main roads clusters, *cluster 2 and 13*, have high levels of head-on crashes as well as having the highest level of vehicles having overturned and crashes involving animals. These road segments, while having relatively low 1 km road segment crash counts, have the most severe with *cluster 13* having all of its crashes involving hospitalization or fatality. *Clusters 1, 5 and 6* have the greatest relative frequency of angle crashes, where road crossings were involved and are among road segment with the highest crash counts.

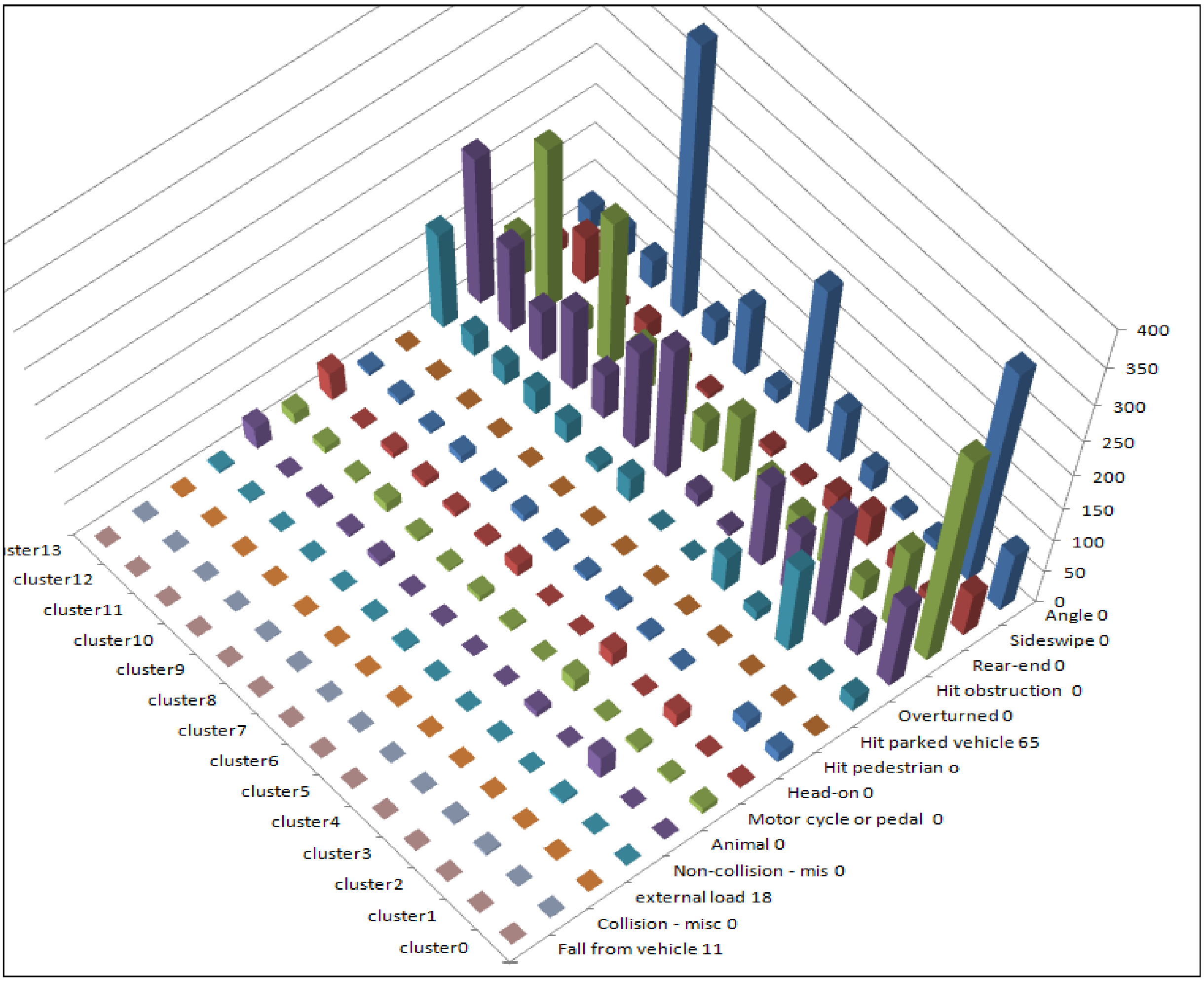


Figure 2 Crash profile of each cluster.

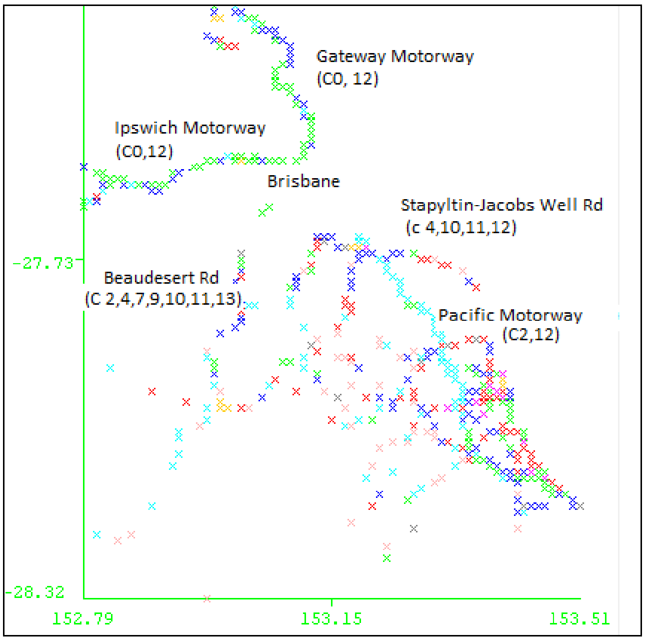
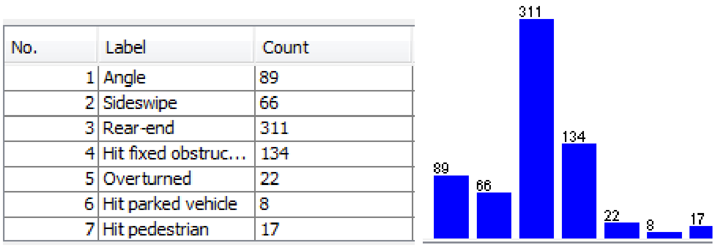


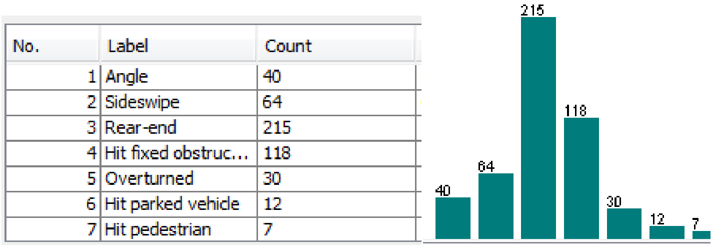
Figure 3 Cluster components of Brisbane motorways and main roads.

Individual cluster profiles were used to describe and explain the effective crash characteristics of a roadway; in this study comparing three Brisbane motorways (Figure 3). *Clusters 0, 2 and 12* are predominant clusters of the Ipswich, Gateway and Pacific motorways, and their respective crash types profiles help explain their crash behaviour. The seal type at the crash locations for Ipswich and Pacific motorways is predominantly dense graded asphalt, whereas the Pacific highway crash locations are predominantly cement concrete along with other seals. The highway crash profiles appear to have a strong relationship with seal type. The Ipswich and Gateway Motorways of dense graded asphalt have predominantly crash *cluster 0 and 12,* whereas the Pacific Motorway is mostly *clusters 3 and 12.*



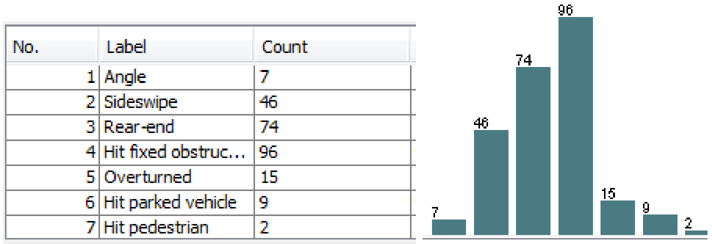
No.	Label	Count
1	Angle	89
2	Sideswipe	66
3	Rear-end	311
4	Hit fixed obstruc…	134
5	Overturned	22
6	Hit parked vehicle	8
7	Hit pedestrian	17

Figure 4 Crash profile of cluster 0 for dense graded asphalt motorway.



No.	Label	Count
1	Angle	40
2	Sideswipe	64
3	Rear-end	215
4	Hit fixed obstruc…	118
5	Overturned	30
6	Hit parked vehicle	12
7	Hit pedestrian	7

Figure 5 Crash profile of cluster 12 for spray seal motorway.



No.	Label	Count
1	Angle	7
2	Sideswipe	46
3	Rear-end	74
4	Hit fixed obstruc…	96
5	Overturned	15
6	Hit parked vehicle	9
7	Hit pedestrian	2

Figure 6 Crash profile for cluster 3.

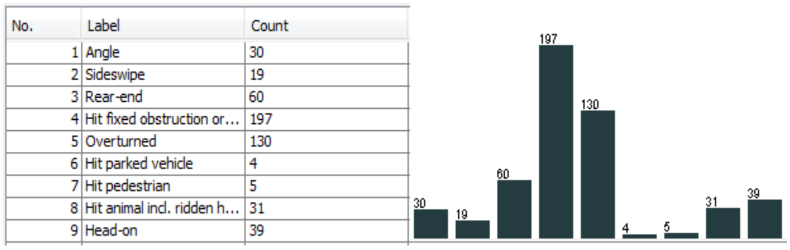
These clusters link the road and its attributes with crash characteristics. The clusters and their relationships with the motorways are described below (Figures 4-7).

Cluster 0, a motorway/highway type, found on the Ipswich and Gateway motorways, has the following characteristics; seal of dense graded asphalt, with 92% being undivided two lane, dual carriageway with speed limit of 100km/h and average traffic rate of 20k vehicles per day. Count of individual crash instances decisions made by on-site crash assessors indicate that 80% of crashes are not related to roadway features, with the remainder associated with T junctions, roundabouts and crossings. Road segment crash count is relatively high, averaging at 4 crashes per year of which a quarter being serious. The crash profile (Figure 4) of this cluster shows that rear-end crashes are the most common by a magnitude of 2, followed by hit fixed obstruction, sideswipe, angle and sideswipe. The cluster profile could be used in decision making to inform signage to alert drivers to the predominance of rear end crashes.

Cluster 12 (Figure 5), a component of all three motorways, shares all attributes in common with cluster 0 (Figure 4), except that the seal type is spray seal. Almost all crashes are assessed as not being related to roadway features. Once again we could apply the crash characteristics provided by the cluster to inform motorists of the impending risks by using appropriate signage.

Cluster 3 (Figure 6), making up a high proportion of the Pacific Motorway, has seal of predominantly cement concrete with sections of dense graded and open graded asphalt. The road is 99% undivided, dual carriageway with over half being 4 lane with the speed limit of 100-110 km/h and average traffic rate of 20k vehicles per day. 95% of crashes are not associated with road features. While the traffic rate at 45k vehicles per day is double that of *cluster 0*, the road segment crash count is substantially lower at 2 crashes per year, of which a quarter are serious. The cluster crash profile (Figure 6) shows that a dominant crash type of hit fixed obstruction, followed closely by rear-end, then by sideswipe. Similar to cases above, the crash profile may be of help in the development of signage messages to inform motorists.

Main roads, presenting a more complex set of conditions to motorists, present a more diverse set of crashes than motorways. Figure 3 shows that the sections making up the Brisbane to Beaudesert Road include clusters 2,4,7,9,10,11 and cluster13. Motorists are confronted with more frequently changing conditions and risks than with motorway driving. While the road has lower traffic volume at 3.5k vehicles per day, some road segments are prone to a high proportion of serious crashes. The presence of Cluster 13 (Figure 7) is of significance, with all crashes of the cluster being serious, and including a relatively high percentage potentially fatal crashes, such as *vehicle overturned* and *head-on crash*.



No.	Label	Count
1	Angle	30
2	Sideswipe	19
3	Rear-end	60
4	Hit fixed obstruction or…	197
5	Overturned	130
6	Hit parked vehicle	4
7	Hit pedestrian	5
8	Hit animal incl. ridden h…	31
9	Head-on	39

Figure 7 Crash profile for Cluster 13

Each of the clusters represent a microcosm that the motorist will encounter during their journey. With information from the clusters, action can be taken based on characteristics of each microcosm to provide appropriate modifications to driver instructions relating to the changing traffic control and risk; actions which have the potential to lead to an a safer journey for the motorist and potentially contribute to an improvement in road safety.

Incidentally, since the data survey, substantial road works have been carried to modernise this road section as part of QDTMR's normal development of state roads.

In addition to providing views of road sets, clustering provide new data view for statistical analysis. The clustering algorithm annotates each instance in the dataset with the cluster number, thus giving the analyst groups of homogeneous roads selected from across the whole network. These new groups of instances can be analysed using statistical tools in the knowledge extraction process [4].

5 CONCLUSION

The paper provides an applied data mining methodology to develop a system of benchmarked clusters for a whole road network and demonstrates the value of the outcome with a case study. This paper demonstrated the benefit of applying clustering within the road management and road safety domains by developing and applying cluster-based crash profiles to describe the crash characteristics of road sections associated by their cluster. This information can be applied in new ways to make decisions about reducing road crashes.

Resulting clusters were differentiated by value ranges within the road attributes, the crash types and the annual crash rates, and this information was used to develop crash profile for each group of roads. The study demonstrates that the crash profiles provide meaningful information, useful in decision making, shown in an example by identifying and applying the predominant crash type for a section of homogeneous road segments selected from the whole road network. Information including risk and driving strategies could be generated and be presented to the motorist using roadway signage. These driver messages could be refined for specific to time of day and time of year and the prevailing weather conditions.

Further, the study describes significant differentiation of attribute values between clusters, allowing analysis of the road characteristics and the individual crash-road instance within and between cluster allocations. In future work this differentiation could be used to support higher-order knowledge development. Standard statistical methods can be used in analysis to support hypothesis development, which in turn could be used to support road design and maintenance procedures focused on road safety.

References

1 World Health Organization. (2002). Global Health: today's challenges. *The World Health Organization*, Retrieved from http://www.who.int/whr/2003/en/Chapter1.pdf. Retrieved February 11, 2011.

2 Queensland-Fire-and-Rescue-Service. (2002). Firefighters called to record number of road crashes,. *Queensland Fire and Rescue Service, Queensland Government, Brisbane,* Last update: November 18, 2002), Retrieved from http://www.fire.qld.gov.au/news/view.asp?id=207 Retrieved October, 2008.

3 Queensland Department of Transport Main Roads. (2011). Road planning and design manual, Design Philosophy. *Queensland Government, Transport and Main Roads. Brisbane,* Queensland. Last Update: February 11, 2011 Accessed from http://www.tmr.qld.gov.au/Business-and-industry/Technical-standards-and-publications/Road-planning-and-design-manual.aspx, Retrieved August 02, 2010.

4 Huang, H., Chin, H. and Haque, M. (2009). Empirical Evaluation of Alternative Approaches in Identifying Crash Hot Spots. *Transportation Research Record*, 21032009), 32-41.

5 ARRB Group. (2011). Tools for improving road safety. *Safe Systems*, ARRB Group,2011, http://www.arrb.com.au/Safe-Systems/Tools-for-improving-road-safety.aspx. Retrieved September 1, 2011.

6 Depaire, B., Wets, G. and Vanhoof, K. (2008). Traffic accident segmentation by means of latent class clustering. Accident Analysis & Prevention, 40, 4 2008, 1257-1266.

7 Alzghoul, A. and Löfstrand, M. (2011). Increasing availability of industrial systems through data stream mining. *Computers & Industrial Engineering,* 60 (2), pp.195-205.

8 Pande, A. and Abdel-Aty, M. (2009). Market basket analysis of crash data from large jurisdictions and its potential as a decision support tool. *Safety Science,* 47 (1), pp.145-154.

9 Zhao, Z., Jin, X., Cao, Y. and Wang, J. (2010). Data mining application on crash simulation data of occupant restraint system. *Expert Systems with Applications,* 37 (8), pp.5788-5794.

10 Anderson, T. K. (2009). Kernel density estimation and K-means clustering to profile road accident hotspots. *Accident Analysis & Prevention,* 41 (3), 2009, pp. 359-364.

11 Emerson, D., Nayak, R., Weligamage, J. (2011). Using data mining to predict road crash count with at focus on skid resistance values. Proceedings of the 3rd International Surface Friction Conference, SFC 2011, Gold Coast, Australia.

12 Qin, X., Ivan, J. N. and Ravishanker, N. (2004). Selecting exposure measures in crash rate prediction for two-lane highway segments. Accident Analysis & Prevention, 36 (2), pp. 183-191.

13 Nayak, R., Emerson, D., Weligamage, J. and Piyatrapoomi. (2011). Road Crash Proneness Prediction using Data Mining. Proceedings of the Extending Database Technology, EDBT, Uppsala, Sweden.

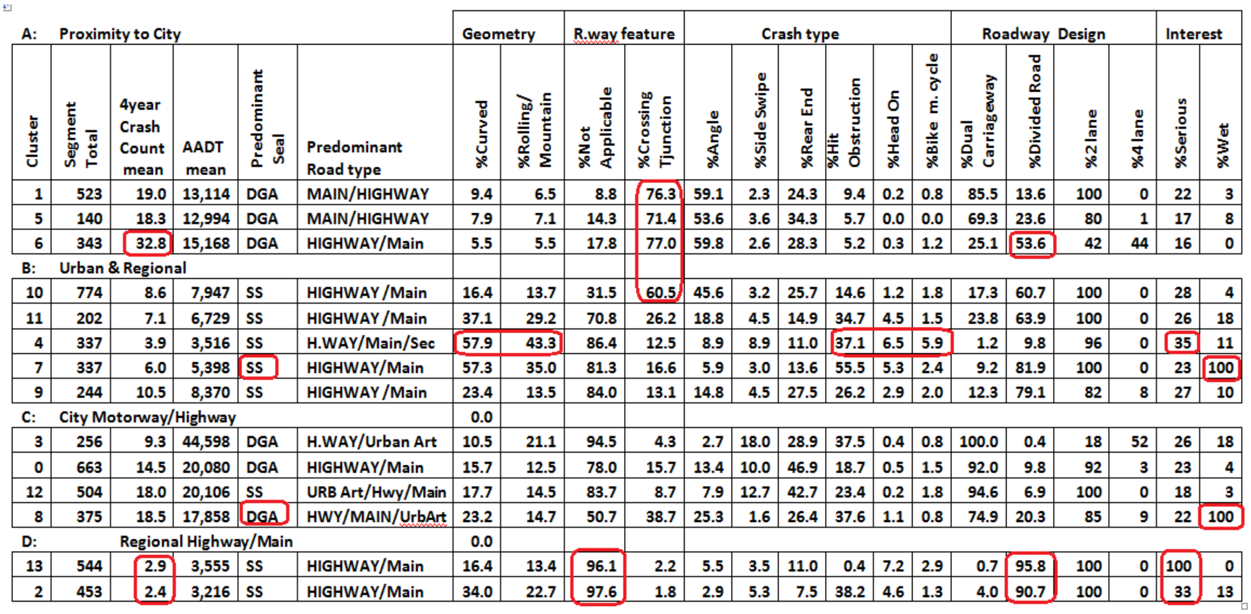
14 Nayak, R., Emerson, D., Weligamage, J. and Piyatrapoomi, N. (2010). Using Data Mining on Road Asset Management Data in Analysing Road Crashes. In Proceedings of the 16th Annual TMR Engineering & Technology Forum, Brisbane, Australia.

15 Emerson, D., Nayak, R., Weligamage, J. and Piyatrapoomi, N. (2010). Identifying differences in wet and dry road crashes using data mining. Proceedings of the Fifth World Congress on Engineering Asset Management, WCEAM 2010, Brisbane, Australia.

Acknowledgments

This paper was developed within the CRC for Infrastructure and Engineering Asset Management (CIEAM), established and supported under the Australian Government's Cooperative Research Centres Programme. The authors gratefully acknowledge the financial support provided by CIEAM. This CIEAM study is a component of a project involving QUT and QDTMR in the development of a skid resistance decision support system for road asset management [11, 13-15]. Queensland Department of Transport and Main Roads provided the road and crash datasets. Data mining operations and presentations were performed in the WEKA and SAS platforms, and substantial computer processing was performed on the QUT High Performance Computing facility (HPC). The views presented in this paper are of the authors and not necessarily the views of the organisations.

Appendix 1:



A:	Proximity to City					Geometry		R.way feature		Crash type						Roadway Design				Interest	
Cluster	Segment Total	4year Crash Count mean	AADT mean	Predominant Seal	Predominant Road type	%Curved	%Rolling/ Mountain	%Not Applicable	%Crossing Tjunction	%Angle	%Side Swipe	%Rear End	%Hit Obstruction	%Head On	%Bike m. cycle	%Dual Carriageway	%Divided Road	%2 lane	%4 lane	%Serious	%Wet
1	523	19.0	13,114	DGA	MAIN/HIGHWAY	9.4	6.5	8.8	76.3	59.1	2.3	24.3	9.4	0.2	0.8	85.5	13.6	100	0	22	3
5	140	18.3	12,994	DGA	MAIN/HIGHWAY	7.9	7.1	14.3	71.4	53.6	3.6	34.3	5.7	0.0	0.0	69.3	23.6	80	1	17	8
6	343	32.8	15,168	DGA	HIGHWAY/Main	5.5	5.5	17.8	77.0	59.8	2.6	28.3	5.2	0.3	1.2	25.1	53.6	42	44	16	0
B:	Urban & Regional																				
10	774	8.6	7,947	SS	HIGHWAY /Main	16.4	13.7	31.5	60.5	45.6	3.2	25.7	14.6	1.2	1.8	17.3	60.7	100	0	28	4
11	202	7.1	6,729	SS	HIGHWAY /Main	37.1	29.2	70.8	26.2	18.8	4.5	14.9	34.7	4.5	1.5	23.8	63.9	100	0	26	18
4	337	3.9	3,516	SS	H.WAY/Main/Sec	57.9	43.3	86.4	12.5	8.9	8.9	11.0	37.1	6.5	5.9	1.2	9.8	96	0	35	11
7	337	6.0	5,398	SS	HIGHWAY/Main	57.3	35.0	81.3	16.6	5.9	3.0	13.6	55.5	5.3	2.4	9.2	81.9	100	0	23	100
9	244	10.5	8,370	SS	HIGHWAY /Main	23.4	13.5	84.0	13.1	14.8	4.5	27.5	26.2	2.9	2.0	12.3	79.1	82	8	27	10
C:	City Motorway/Highway					0.0															
3	256	9.3	44,598	DGA	H.WAY/Urban Art	10.5	21.1	94.5	4.3	2.7	18.0	28.9	37.5	0.4	0.8	100.0	0.4	18	52	26	18
0	663	14.5	20,080	DGA	HIGHWAY/Main	15.7	12.5	78.0	15.7	13.4	10.0	46.9	18.7	0.5	1.5	92.0	9.8	92	3	23	4
12	504	18.0	20,106	SS	URB Art/Hwy/Main	17.7	14.5	83.7	8.7	7.9	12.7	42.7	23.4	0.2	1.8	94.6	6.9	100	0	18	3
8	375	18.5	17,858	DGA	HWY/MAIN/UrbArt	23.2	14.7	50.7	38.7	25.3	1.6	26.4	37.6	1.1	0.8	74.9	20.3	85	9	22	100
D:		Regional Highway/Main				0.0															
13	544	2.9	3,555	SS	HIGHWAY/Main	16.4	13.4	96.1	2.2	5.5	3.5	11.0	0.4	7.2	2.9	0.7	95.8	100	0	100	0
2	453	2.4	3,216	SS	HIGHWAY/Main	34.0	22.7	97.6	1.8	2.9	5.3	7.5	38.2	4.6	1.3	4.0	90.7	100	0	33	13

Table 2 Case study cluster details