5 Transportation Networks, Case-Based Reasoning and Traffic Collision Analysis: A Methodology for the 21st Century

Kaidong Li and Nigel M. Waters

University of Calgary, Calgary, Canada

5.1 Introduction

An average of 400 people are killed and 23,000 injured each year as a result of motor vehicle collisions in the western Canadian province of Alberta (population of 3,146,066; Statistics Canada estimate for 2003). Motor vehicle collisions are the leading cause of death for Albertans under the age of 30 (AMA 2003). These statistics, on a *per capita* basis, are similar to those that might be obtained for many developed states, provinces and nations and their reduction through the application of computer technology is the goal of much recent research in traffic safety analysis (Arthur 2002).

Traffic safety is of paramount concern to not only police and traffic safety engineers, but also to the insurance industry, health care organizations, social workers and the public at large. There is a need for efficient and *intelligent* facilities to study, and eventually enhance, safety on road and other transportation networks. The provision of such a facility is an integral component of work currently being conducted in Calgary's intelligent transportation systems (ITS) community (City of Calgary ITS Workshop No.3, 2003).

Geographic information systems (GIS) have recently received substantial attention in the ITS and traffic safety literature (Arthur 2002; Panchanathan and Faghri 1995; Smith 2000; Spring and Hummer 1995; Waters 2002). The benefits of using a GIS include spatial query, multi-source data overlay and network analysis, among others.

Rule-based expert systems, that embody specialist knowledge in a computer based decision support system (Chris Naylor Research Limited 2003, p. 1), have been used as decision support systems in transportation and road safety research in the past decade (Herland et al. 2000; Panchanathan and Faghri 1995; Paniati and Hughes 2000; Suttayamully et al. 1995). Case-based reasoning (CBR), a parallel methodology to rule-based reasoning, is of increasing interest to the ITS community (Boury-Brisset and Tourigny 2000; Capus and Tourigny 1998; Waters 2002). In CBR, prior experiences are reused and adapted for solving new problems (Leake et al. 1996; Richter et al. 1998). Prior experiences, whether successful or failed, carry valuable information. Similar to human learning and reasoning activities, the system can become more efficient over time as a result of collecting and indexing additional experiences (Leake 1996). CBR approaches are more appealing than the rule-based ones in areas where situations are complex or rules are difficult to generalize (Capus and Tourigny 1998; Khattak and Renski 1999). Traffic safety is a complex function of a wide range of factors, such as road network and environmental conditions, driver behaviour and vehicle condition. Prior experiences rather than general rules might provide more site-specific or case-specific solutions to a given safety problem.

5.2 Applications of GIS in Traffic Safety Analysis

GIS have been widely used in transportation research and management since the late 1980s (ESRI 2003; Thill 2000). Over the past two decades, GIS for Transportation (GIS-T) has evolved and matured to a sophisticated technology that is now applied to a wide range of transportation research (Miller and Shaw 2001; Nyerges 2005; Waters 1999). Traffic safety, an important aspect of transportation research, also shares the benefits of this technology. Early efforts to provide workstation capacities for GIS-based safety analysis were developed in the 1990s for the Federal Highway Administration (FHWA) in the United States. These GIS-based software packages were capable of performing functions that ranged from identifying hazardous intersections to analyzing corridors.

One of the main advantages of a GIS is its ability to access promptly and analyze accurately data distributed across a transportation network. This benefit of GIS enables the analysis of traffic safety at specific sites, along the whole network as well as on the area-wide transportation system. Another advantage of a GIS is its ability to retrieve rapidly relevant information from various sources, such as traffic flow, land use, environmental and socioeconomic databases. Other capabilities of a GIS include: spatial analysis, user-friendly visual representation, thematic mapping and charting. Furthermore, most GIS software provides the capability of interfacing external programs for decision support and database management. For example, a Visual Basic for Application (VBA) development kit packaged with the ArcGIS software produced by the Environmental Systems Research Institute (ESRI) allows users to call external stand-alone software within a GIS environment (Lang 1999).

A review of recent traffic safety studies and projects shows that safety engineers primarily apply GIS technology in investigations that fall into three, overlapping categories: 1: Data collection and access; 2: Data processing and analysis; and 3: Decision support.

5.2.1 Data Collection and Analysis

Locating traffic collisions. In many jurisdictions, including Alberta, collision data is manually transcribed from paper collision report forms and then georeferenced as a separate operation. In this process the quality and quantity of the original data may affect accuracy in the developed GIS database. The employment of GIS techniques, such as batch processing, intersection matching and conflation, as well as the incorporation of road attributes from multiple sources, increases the percentage of identified and located collisions (Carreker and Bachman. 2000; Thompson 2001).

Transportation network data. Safety engineers must have accurate data on the physical characteristics of a transportation network and its traffic characteristics. For a road network this would include: roadway geometry, roadway condition, traffic control devices, roadside features and traffic volumes. Conventionally, this information is stored in spreadsheets, or database tables. In these spreadsheets and tables, locational information can only be described by text. GIS enable the acquisition and display of roadway characteristic data *visually*, and hence, *efficiently*. GIS also allow simultaneous access to multiple data layers, as well as the ability to interface with external information such as census tracts and municipal land use zoning areas. (Panchanathan and Faghri 1995). An example of a GIS that performs such a role is the on-going development of a road network inventory in the City of Calgary (Nelson 2002).

5.2.2 Data Processing and Analysis

Traffic safety studies, in this category, emphasize the analytical capabilities of GIS. These capabilities range from simple buffer and overlay operations to spatial query (Spring and Hummer 1995), time series analysis (Arthur 2002), network analysis (HSIS Summary Report 1999) and dynamic segmentation (Bayapureddy 1996). A review of the literature and recent projects shows five trends in the use of GIS to analyze accident data: (1) identifying the high-accident locations, (2) identifying the accident characteristics, (3) determining the causes, (4) determining the countermeasures, and (5) evaluating the countermeasures (Bayapureddy 1996).

Identification of hazardous roadway locations. The identification of hazardous or high-accident locations is one of the most commonly used GIS approaches in traffic safety studies. Spring and Hummer (1995) state that the *sole* use of traditional statistical methods for the determination of high-accident locations presents several limitations. One of which is that the determination is based on experiential data. Thus problem locations are not revealed until a significant number of accidents has occurred. Another limitation is the “regression to the mean” phenomenon and high accident levels may be due to this statistical anomaly, and not to a road network problem (Spring and Hummer 1995). The third limitation is the “accident-matching” problem whereby differently formatted data can cause distortion and misunderstanding when displaying high-accident locations on the network (Bayapureddy 1996). Within the GIS software, different types of data are easily related, either graphically or in report forms, thus making the data more easily accessible and providing a more intuitive and flexible user interface (Arthur 2002; Bayapureddy 1996). By providing consistent access to a common pool of accident and roadway data, GIS can “identify and predict hazardous highway locations before accidents occur” (Spring and Hummer 1995, p. 83).

Accident pattern analysis and remediation. GIS has the potential to improve traffic accident analysis in several significant ways. First, GIS provides visual acquisition and display of accident locations. This equates to a minimum one-year advance in knowledge of accident information, since in the past it would have taken up to one year to enter and geocode these data (Hall et al. 2000). Additionally, the individual accident characteristics become immediately available for site analysis. Rapid access to individual accident information enhances the appropriate application of countermeasures for accident remediation. GIS facilitate analysis of accident locations by integrating previously disparate data elements. For example, incorporation of accident types and conditions in conjunction with road network characteristics, such as lane width, surface type and friction index number, allow substantially more multivariate analyses (Hall et al. 2000).

Network analysis. The Linear Referencing System (LRS) provides GIS with network analysis capability, allowing multiple routes to be linked together in a manner that enables the analyst to assess the overall safety performance within a transportation network. A common use of network analysis is to examine truck crashes along designated truck routes. On a network-wide basis, a particular element may have a high accident rate and thus it may be more cost-effective to make a network-wide correction of a common element than to correct only a high-accident location (HSIS Summary Report 1999).

5.2.3 Decision Support

Another potential application of GIS in traffic safety is to support decisions for policy makers, transportation planners and safety engineers. Such decision support systems (DSS) are useful to traffic safety management organizations. Efforts have been made to integrate GIS and other DSS to manage traffic safety in the Federal Department of Transportation (DOT) in the United States. For example, Ozbay and Mukherjee (2000) developed a web-based expert GIS, called WIMSI, to provide operators with high-level analyses and recommendations concerning incident response. This system integrated a rule-based reasoner, the incident and network database in MS Access and the web-enabled GIS developed in ArcView IMS to provide a real-time, incident management DSS. Panchanathan and Faghri (1995) developed a knowledge-based GIS system to manage safety at rail-highway at-grade crossings in Delaware. The advanced functionality of a GIS, including spatial query, multi-source data overlay and network analysis, make it a powerful platform and database to support an intelligent system for decision makers.

5.2.4 A Synopsis of the Impact of GIS in Traffic Safety Studies

Although GIS have been successfully employed in recent traffic safety studies, Smith (2000) noted that, except for reporting a picture of crash locations, the majority of safety engineers were not using the full range of GIS capabilities. According to Smith the reasons why GIS is not popular for safety analyses with many safety engineers include: the need for a GIS champion; the need for progressive communications between the GIS leader and safety engineers; a lack of GIS knowledge; a lack of resources for GIS development; and not using GIS to manage the linear referencing system for the transportation network (Smith 2000, p. 12).

Smith (2000) also addressed some obvious barriers in implementing GIS-based safety applications within the traffic safety authorizations in the United Sates. These barriers included: not seeing the benefits that GIS has to offer in safety analyses; not using a common technical language, for the needs, and requirements, of both the highway safety engineers and the GIS systems engineers; not having, or supporting, standards for data consistency; an absence of appropriate data, such as transportation network inventory, accident data and work zone, necessary to perform safety analyses; and problems understanding the various linear referencing systems (Smith 2000, p. 3).

GIS has proven to be a promising technology in traffic safety studies. It provides functionality to display visually and analyze spatially distributed safety data by their actual locations. Applications of GIS in traffic safety research are still at an initial stage. The slow diffusion of GIS in traffic safety research might possibly be explained by the complexity of these systems and the misconstrued perception that they are not qualitatively different from the older computer aided design and drafting (CADD) packages. Increased productivity and well-trained personnel are the two most important variables in applying GIS in traffic safety research.

5.3 Applications of Case-Based Reasoning in Traffic Safety Analysis

The term “case-based reasoning” (CBR) is used in both cognitive science, which studies the human behaviours of reasoning and learning and in artificial intelligence (AI), in which CBR is implemented to make AI systems more efficient. In the context of artificial intelligence, CBR is normally applied in the development of expert systems (ES) (Waters 1988). However, CBR tools reason differently from rule-based expert systems that draw conclusions by chaining together generalized rules, starting from scratch. In CBR new solutions are generated by retrieving the most relevant cases from memory and adapting them to fit the new situations (Leake 1996). CBR systems retrieve previous cases that are similar to the current problem and attempt to reuse, or adapt, relevant solutions in the new situation (Kolodner and Leake 1996).

This approach is based on two tenets about the real world: first, the world is regular, similar problems have similar solutions; second, the types of problems an agent encounters tend to recur (Leake 1996). The diffusion of ES was not as broad and as rapid as predicted. One crucial reason is the so-called "knowledge elicitation bottleneck" (Waters 1989). In a rule-based ES, much time and effort are required in order to elicit and generalize a series of rules, thus creating a "bottleneck."

Roger Schank first proposed CBR, in 1982, as a model for human reasoning processes (Joh 1997). Since the early 1980s, CBR research has matured and the methodology has been widely used in various research programs and practical applications (Capus and Tourigny 2000; Kolodner 1996). Compared to a rule-based system, a case-based system has advantages in the following five aspects:

Knowledge acquisition. The first step in building a rule-based system is to provide rules for the inference engine. In some domains, however, rules are difficult to formalize or become unmanageably large (Capus and Tourigny 1998; Leake 1996). CBR systems utilize previous cases to infer a new solution, thus making it unnecessary to decompose experiences and to elicit rules to support the reasoning processes. CBR avoids the "knowledge elicitation bottleneck" problem in rule-based systems and is especially useful when conceptual knowledge is limited (Khattak and Renski 1999).

Knowledge base maintenance. A conventional rule-based system requires the definition of a complete and "perfect" knowledge base. However, initial understanding of the problem is usually imperfect and circumstances and requirements may change over time. In CBR, the knowledge base is scalable and adaptable on an ongoing basis (Wisdo 1997). CBR systems "can be operated with only a partial case base, and are always expected to add new cases" (Clayton and Waters 1999, p. 279) or update existing cases. The knowledge base in a CBR system can evolve and grow simply by adding new experiences.

Problem-solving efficiency. Reuse of prior solutions in CBR systems greatly reduces the need to repeat prior effort, thus increasing problem-solving efficiency. In addition, reuse can avoid potentially problematic solutions by saving failed cases.

Solution quality and consistency. Rules can be imperfect and unreliable when the principles of a domain are not well understood. In this situation, solutions suggested by actual cases might be more accurate and reliable since they reflect real world circumstances. Consistent solutions are guaranteed for problems under similar situations (Leake 1996; Richter 1998).

User acceptance. Most rule-based systems produce unsatisfactory results because of their limited explanatory capabilities. End users of the system often find it difficult to understand, and accept, the solutions suggested by chains of rules. In CBR, solutions and explanations are given based on actual prior cases rather than on the basis of generalized rules (as in rule-based systems) or by "black boxes" (as in a neural network). These solutions and explanations are more acceptable and interpretable to the users of the system (Boury-Brisset and Tourigny 2000; Burkhard 1998; Leake 1996).

5.3.1 Applications of CBR

Since the early 1980s, CBR methodology has been used by many prominent corporations such as Lockheed, GTE, DEC, Boeing and Martin Marietta (Clayton et al. 1998). As the technology becomes more mature, CBR has been successfully applied in various fields such as medical diagnosis, software development, planning and decision support. Recently, CBR has proven to be quite a useful and efficient tool for electronic commerce (E-commerce), such as customer support, hotlines and helpdesks. Software companies that provide CBR tools and consulting services include eGain Communication Corporation, US (Knowledge product family) and, in Europe, empolis Knowledge Management GmbH (CBR-Works) (see University of Kaiserslautern 2003 for a detailed list). CBR applications can be sorted by different task types (Richter 1998; University of Kaiserslautern 2003):

Diagnosis, Classification, and Decision Support. This category contains the most common applications of CBR, such as medical diagnosis, information classification (Richter 1998), legal reasoning (University of Kaiserslautern 2003), troubleshooting (Lenz et al. 1998), tutoring (Weber and Schult 1998) and helpdesk (Haley Enterprise Inc. 2003). Algorithms to calculate the nearest neighbour in the knowledge base appear crucial for applications in this category.

CBR Supported Planning and Design. Experience plays an important role in planning and design activities. However, planning and design activities introduce variable degrees of creativity. Solutions taken from a case base almost always need modification before reuse in an actual planning or design problems (Richter 1998). Hence, the adaptation capability of a CBR is an important issue in this group of applications. Example projects include *Aircraft Conflict Resolution – CBR support for Air-Traffic Control* by Artificial Intelligence Group at Trinity College Dublin and *Bioplan – Planning of Bioprocess Recipes* by Bioprocesses Group at VTT Biotechnology and Food Research (University of Kaiserslautern 2003). Reuse of parts, software and knowledge also falls into this category.

Information Retrieval. Applications in this group include E-commerce on line catalogue sales (Vollrath et al. 1998), intelligent Internet search (Vollrath et al. 1998) and hotline support (Lenz et al. 1998). The most essential issue for this type of task is the inexact matches of textual information. CBR is a natural problem solving technique that makes reference to the specific context of a particular problem. For example, when one searches the Internet for a certain product only the intended use of the project and not its various features are known (Richter 1998).

5.3.2 Challenges and Limitations of CBR

The major limitation of CBR is that an optimal solution cannot be guaranteed. Solutions provided by a CBR system are restricted to its case library and thus have only a certain degree of flexibility. The efficiency of the solution provided by a CBR system relies on the coverage of the case base, thus making it very difficult to be evaluated *quantitatively* (Leake 1996). Another challenge is associated with designing the adaptation rules. The definition of adaptation rules can be a complicated task that depends on the knowledge domain. Although several techniques have been applied in CBR for adaptation, most commercial CBR tools remain only case retrieval systems with adaptation being left to human intervention (Watson & Marir 1994).

The third challenge involves finding an appropriate indexing scheme (Haddad 2003). Relevant cases in human memory can be retrieved quickly when needed due to the efficient methods of indexing in a human brain. Although many algorithms have been developed for automated indexing methods in a case-based reasoner, Kolodner (1993) believed that people tend to be better at choosing indices than algorithms, and therefore, for practical applications, indices should be chosen by hand.

5.3.3 CBR Applications in Traffic Safety Analysis

Waters (2002) noted that traffic accidents and traffic safety studies are complex phenomena in a technical sense. Accordingly, existing regulations and safety measures tend not to reflect reality nor do they necessarily protect the driver. One of the first uses of CBR in road safety analysis (ROSAC – Road Safety Analysis with Cases) was proposed and developed by Capus and Tourigny (1998). ROSAC is a knowledge-based system that provides safety improvement solutions at road intersections based on previous, similar cases. Domain knowledge and expertise is represented by individual cases instead of rules. The system has two main functions: the first function is to manage a case base of road intersections as a conventional database while the second function allows the system to search for cases most similar to the situation encountered and, as needed, to reuse, adapt and save them as new cases. The developers of ROSAC found that the results of tests on hypothetical cases and a small sample of real cases were “satisfactory” (Capus and Tourigny 1998, p. 7).

Further to the development of this prototype, researchers at Laval University have designed an organizational memory (OM) using a hybrid approach that combines rule-based reasoning and case-based reasoning for road safety analysis (Boury-Brisset and Tourigny 2000). An OM consists of the integration of different knowledge assets in an organization, including both theoretical knowledge and practical know-how. This system is called SICAS (System with Intelligent and Cooperative functions to help in the Analysis of Sites). It shares domain expertise of road safety analysis among experts and analysts within the organization. It assists the analysts by identifying safety problems on a given site, determining possible causes and identifying proper correcting actions. Within SICAS, domain expertise is stored in a knowledge base while individual cases are stored in a case base. The developers realized that the domain expertise was efficient for solving “easy cases.” However, as it grew, users could more effectively exploit the case base as a DSS for handling more complex cases. The developers also realized that the CBR approach enabled communication between experts and end users. Solutions to a complex case were proposed by the experts according to the features of the case reported by the end users. Moreover, detailed follow-up studies were facilitated and made possible since both the features *and* solutions of the case were saved in the case base (Boury-Brisset and Tourigny 2000). Lin et al. (2003) state that the CBR approach is useful in estimating safety benefits of road improvements. In order to evaluate the effectiveness of countermeasures that have been applied to a problematic road site, collision reduction factors (CRF) were calculated. Lin et al. argued that the estimation of CRF should also account for specific circumstances (e.g. location characteristics and surrounding environment, etc.) and the random nature of collisions. They reviewed 450 previous safety studies that described the effectiveness of various safety improvements. The reported CRF of these cases together with their physical and traffic attributes, and countermeasure types were documented in a CBR system, named ISECR. The effectiveness of safety countermeasures on a new study project can then be estimated based on similar previous cases in the case base. The developers tested the system and found that it “does provide valid results” (Lin et al. 2003, p. 389). CBR applications in traffic safety analysis are still limited to the research stage. Potential challenges of applying CBR to traffic safety analyses in the real world are: first, problems associated with scaling up the case base from test bed systems (Leake 1996) and, second, gaining user acceptance (Boury-Brisset and Tourigny 2000).

5.3.4 Summary of the Importance of CBR in Traffic Safety Studies

CBR is an information-retrieval methodology that uses the most similar prior cases as references for suggesting solutions to solve new problems (Leake 1996). The information-retrieval method, adaptation capability, and learning process are the major advantages of CBR compared to other database and information retrieval systems (Waters 2002). In addition, CBR systems are more interactive and allow communication between the experts and the end users (Boury-Brisset and Tourigny 2000).

5.4 Integrating Case-Based Reasoning and Geographic Information Systems

Making decisions is a complex activity that often involves the interaction of many disciplines. It is agreed that the key to useful computer-based DSS is integration (Abel et al. 1996; Cortes 2000). Tools obtained from different sources should be interoperable and, at best, combined in a common framework. GIS provides a rich solution for simultaneously integrating, analyzing and distributing data. Data manipulation operations (e.g. overlay) in most GIS applications overcome the heterogeneity of data coming from different sources with different levels of precision (Cortes 2000; ESRI 2003; Cusack 2003). Further, the mapping capability in GIS is optimal for visualizing data distributed across a network.

New research has sought to incorporate knowledge-based systems with GIS in order to create intelligent information systems. A knowledge-based GIS for safety analysis at rail-highway crossings in Delaware was developed on the TransCAD platform (Caliper Corporation 2003) to store and retrieve safety-related information at state-wide crossings (Panchanathan and Faghri 1995). Abel et al. (1996) used a KBS as a programming tool to perform queries while GIS worked as a "toolbox" and a database to enable spatial analysis. They also discussed the "re-use" of prior experiences as an intelligent component in the next generation of DSS. According to Cortes's (2000) review, the integration of GIS and ES can benefit an environmental decision support system (EDSS) with respect to data integration, data mining, problem diagnosis and decision support. The integration of CBR and GIS also shows promising results in some recent research. Examples include Clayton et al.'s (1998) work of integrating Traditional Environment Knowledge (TEK) in Gwich'in, Northern Territories within a GIS and Khattak and Renski's (1999) project of reusing and adapting prior cases for High-Occupancy-Vehicle (HOV) lane planning.

5.4.1 Options for Integrating CBR Tools with a GIS

There are several options when integrating CBR tools with GIS (Clayton and Waters 1999): 1. attach GIS layers at the end of CBR actions or answers; 2. operate both the CBR and the GIS through a common interface on a network; 3. have the GIS operate as a client program for CBR, sending requests and initiating the CBR when required; 4. run CBR as a client program in the GIS context using VBA; or 5. implement the CBR methodology and algorithms directly within a GIS environment. Clayton and Waters (1999) explored the first two options in their study of integrating traditional environment knowledge within a GIS. They found that the best option was to attach the GIS layers as "browsable" files over a common network because it "does not require the user to be an expert in GIS" (Clayton and Waters 1999, p. 299). A web-based expert GIS was also found to be highly useful and quite user-friendly for applications that involve multiple users (Ozbay and Mukherjee 2000). However, there are a number of drawbacks to this approach. First, the user's access to the GIS files is limited since the GIS files are always the end result of the search. Second, it takes significant efforts to link the database and case base without a powerful relational database server. The feasibility of having the GIS operate as a client program greatly depends on the interoperability of the CBR software. Most readily usable CBR programs are commercial applications; they are usually specific and offer very limited capability of accommodating other software. On the other hand, most GIS software has the capability of interfacing external software such as CBR tools. Khattak and Renski (1999) developed a CBR tool for HOV lane analysis in the ESRI's ArcView 3.0 environment (www.esri.com). They applied CBR methodology and algorithms using the Avenue programming language to query historical HOV cases. The direct application of CBR algorithms in GIS is a truly intelligent system that enables the immediate query and analysis of historical cases, adaptation of previous solutions and the display of final results all in one database and on one software platform.

5.4.2 The Potential for Integrating CBR and GIS in Traffic Safety Research

The above discussion outlined how GIS and expert systems have been applied to solve traffic safety problems, and how the integration of case-based reasoning and GIS can provide new opportunities for the design and application of intelligent systems for traffic safety analysis. Recent efforts have contributed to applying new AI technologies such as case-based reasoning (Boury-Brisset and Tourigny 2000; Capus and Tourigny 1998) and integrating AI with other technologies such as GIS (Ozbay and Mukherjee 2000) to build more efficient intelligent systems for traffic safety analysis. As Guariso and Werthner (1994) pointed out, all these intelligent systems cannot, and will not, do the work that remains to be done by humans. *Better computer support does not necessarily imply a better decision*; it is still the safety professional’s responsibility to validate decisions provided by these systems.

5.5 A Case Study: Integrating GIS Network and CBR Approaches for Traffic Safety

The Calgary Light Rail Transit (LRT) system has been serving Calgarians for more than two decades and will continue to serve as a major public transit system in the City. The Calgary LRT system is a relatively safe system and in recent years has produced few collisions. Light-rail vehicle (LRV) related collisions from 1996 to 2000 averaged only 2.4 per annum. However, collisions involving an LRV are generally more severe, affect more people and take more time to clear up. Safety engineers, transit planners and transit organizations seek to reduce potential traffic collisions and conflicts among LRV, motor vehicles and pedestrians along LRT corridors and to provide a safer and more reliable service.

Based on the potential conflicts between LRV and motor vehicles, bicycles and pedestrians, light rail alignments can be classified into three categories (TCRP Report 69 2001):

Exclusive: light rail and roadways are fully grade separated or at-grade without crossings; examples in Calgary include short sections along the northwest line, where both below-grade and elevated structures are used;
Semi-excusive: light rail and roadways use separate rights-of-way, or shared rights-of-way protected by barriers, this occurs in various parts of the Calgary network including just east of the downtown area;
Non-exclusive: mixed traffic operations or LRT/pedestrian malls, such as in the central downtown portion of the network.

Semi-exclusive and non-exclusive systems experience variable degrees of mixed operation with other traffic and with pedestrians. Seventy percent of the total track length of the Calgary LRT is classified as semi-exclusive or non-exclusive alignment (TCRP Report 69 2001). The main advantage of semi-exclusive or non-exclusive LRT systems is their cost efficiency. However, the at-grade design of light rail can result in traffic collisions or conflicts among trains, motor vehicles, bicycles and pedestrians.

For example, all 12 LRV-related collisions 1996 to 2000 occurred at rail-roadway grade crossings, including one fatality and four injuries (Fig. 5.1). Questions occur when implementing safety strategies along existing LRT, and planning future LRT systems: how safe are the existing light rail/roadway at-grade crossings in Calgary? Did they experience higher collision rates and/or more severe consequences due to the presence of LRT? Are there ways to improve the safety at these crossings? Or is it essential to grade-separate light rail and roadways in the City? This case study aims at implementing a methodology to help traffic safety professionals to answer these, and related, questions.

5.5.1 Methodology

Traffic safety studies typically involve analyzing collision history, addressing existing safety problems and providing countermeasures. This methodology requires a robust sample size of collision data in order to minimize variations due to external effects (Hamilton-Finn 2000) and is thus questionable for analyzing LRT safety since LRV-related collisions are relatively rare. When LRV-related collision data are insufficient, analyzing the patterns of non LRV-related collisions at rail/roadway grade crossings might provide useful information to predict potential LRV-related collisions, and to examine safety issues at these crossings. However, the occurrence of a traffic collision is a complex function of three major factors: human, vehicle and environment, each considered before, during and after the collision (Haddon et al. 1964). The variations in collision patterns are the results of the differences of these three major factors among sites and time. Thus the relationship between non LRV-related collisions and LRV-related safety issues at a given site cannot be modelled using a simple equation. Rather, another, more holistic reasoning approach that takes into account site-specific characteristics should be considered. As noted above, the CBR approach that values the details of each individual case may be used to explore the relationships among such complex events.

The purpose of this case study was to review traffic safety at Calgary’s light rail/roadway grade crossings, by employing GIS technology and CBR tools, to analyze the collision history and site-specific characteristics at these crossings. In this study, GIS played an important role as a framework enabling the manipulation and analysis of collision data, as well as the retrieval of other relevant information. CBR tools were employed to assist in the identification of potential safety problems at rail/roadway at grade crossings based on the analysis of collision histories and the site-specific network characteristics.

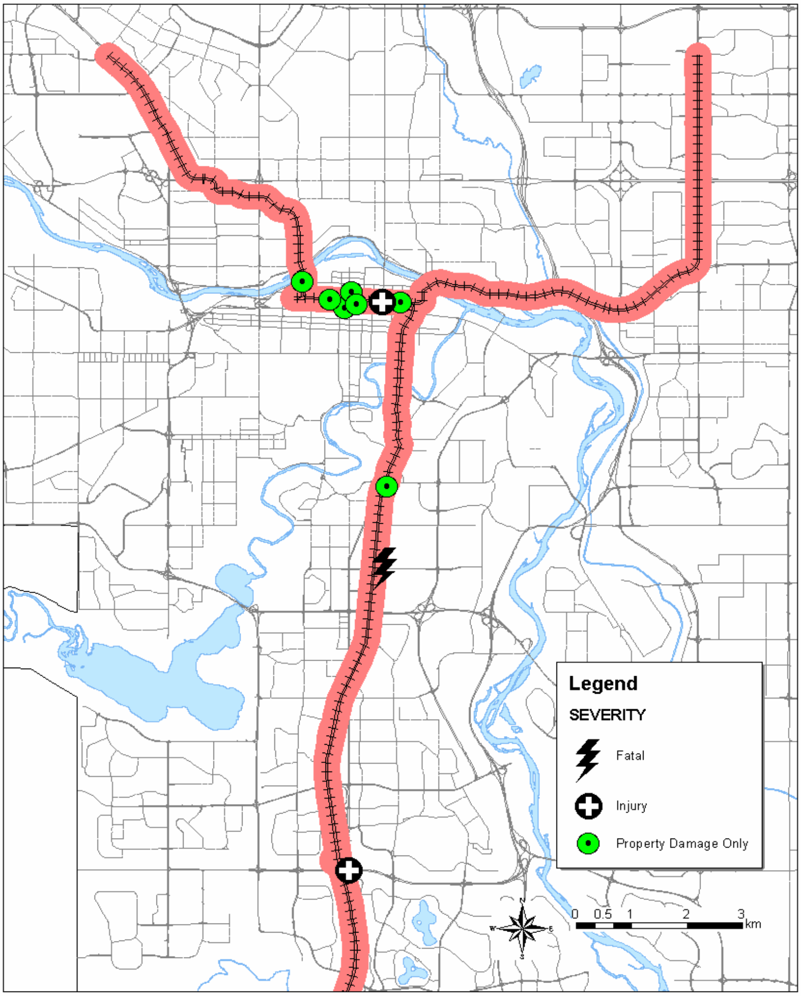


Fig. 5.1. LRV-related collisions from 1996 to 2000

The CBR system was then used to recommend possible solutions.

The case-study objectives were: 1. to import, manipulate and analyze the collision data in a GIS environment; 2. to retrieve relevant site-specific information from multiple sources using GIS analysis; 3. to determine potential safety problems at these light rail/roadway grade crossings and to recommend solutions using CBR tools.

5.5.2 Data Acquisition and Manipulation

ESRI’s ArcGIS ArcView 8.2 was chosen as the GIS software to manipulate the spatial and non-spatial data used in the project. Table 5.1 lists the descriptions and sources of the available data used in this project. Figure 5.2 illustrates the major procedures used in ArcView 8.2 to import collision data, and to retrieve desired information at light rail/roadway grade crossings in the City of Calgary.

The Collisions_2000.dat, a text file containing records of collisions occurring within the City of Calgary in year 2000, was provided by the Calgary Police Services. This text file was first imported to MS Access database (Fig. 5.2). Simple queries were then performed to separate information concerning the occurrence, involved objects and involved vehicles into three tables. These three tables were linked by the field of ‘Complaint Number’, which is unique for each individual collision. The table of ‘Occurrence’ contains information on the collision occurrence, such as the date and time of occurrence, type of collision, severity level, and location of occurrence, among others.

Table 5.1. Data sources

Data file name	Descriptions	Source	# of records
Collision_2000.dat	Text file. Each record represents one reported collision occurring in 2000. Each record contains one identification number and 62 attributes with information on occurrence, objects and vehicles involved (e.g. date and time of occurrence, object type and vehicle make, etc.).	Calgary Police Services. 2001	Occurrence: 34,825
Roadnet_2000.shp	ArcView polyline shapefile. Road network in the City of Calgary, with traffic flow information of 2000 on major roads.	City of Calgary, 2000	Polylines: 76,774
Signalized_ Intersection.shp	ArcView polygon shapefile. Major signalized intersections in the City of Calgary	City of Calgary, 2003	Points: 778
Lightrail.shp	ArcView polyline shapefile. Existing LRT tracks in the City of Calgary (by May, 2003)	City of Calgary, 2003	Polylines: 13,598
Landuse.shp	ArcView polygon shapefile. Land use type information within the City limit	City of Calgary, 2000	Polygons: 5,549
LRT_crossings.shp	ArcView point shapefile. 47 light rail/roadway crossings along the City’s existing LRT tracks.	Digitized by Li, K., 2002	Points: 47
LRT_stations.shp	ArcView point shapefile. 31 existing LRT stations by December, 2001.	Digitized by Raza, W., 2001	Points: 31

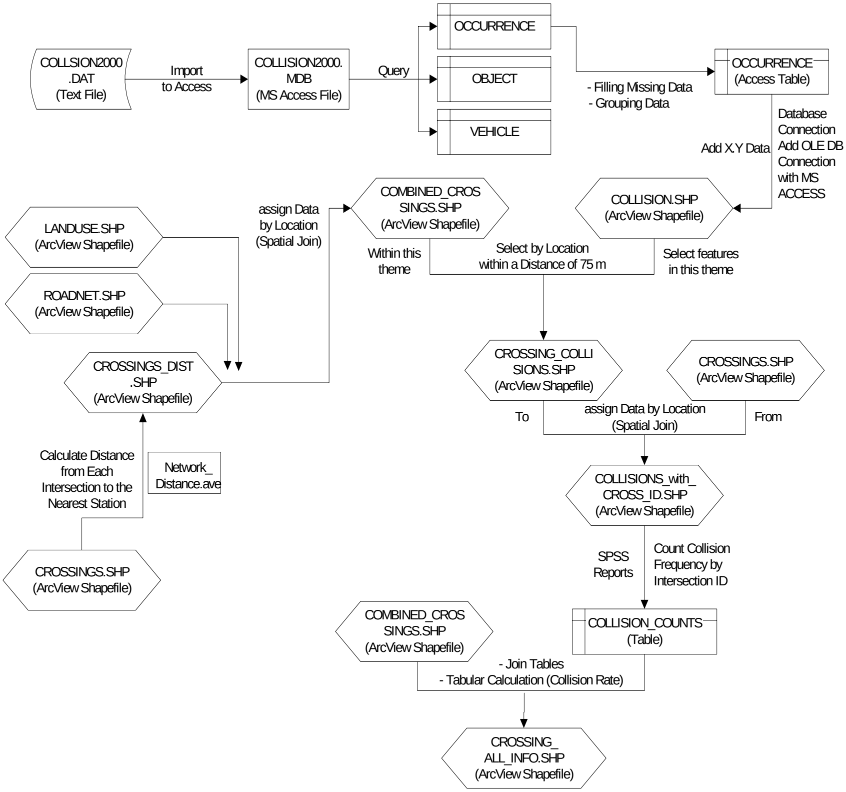


Fig. 5.2. Procedures for importing and retrieving information in GIS

Collision locations are represented by X,Y coordinates, and thus can be displayed geographically in a GIS. 30,717 collisions with valid X,Y coordinates were converted to the ArcView shapefile format. LRT crossings were digitized based on the Calgary Transit Map 2002. Attributes, such as type of light rail alignment and gate treatment, were added to each LRT crossing. Eight of the 47 LRT crossings in Calgary were classified as exclusive alignments with underpass structures, while the other 39 are at-grade crossings. Since the goal of this project was to examine the safety of those at-grade crossings, only the 39 at-grade crossings were extracted and analyzed. Thirty one LRT stations were digitized based on the same transit map. Most of the data used in this project were provided by the City of Calgary. Therefore, the City's standard coordinate system was used in this project. The City of Calgary currently uses datum NAD83, projection 3TM (a modified UTM system) to represent transportation data.

5.5.3 Retrieving Desired Information in the GIS Framework

Based on the review of LRT safety studies (Capus and Tourigny 1998; Lin et al. 2003; TCRP Report 69 2001), three types of information are considered useful for analyzing the safety of an LRT crossing: the collision history, the traffic and physical characteristics and the context information at this given crossing. Table 5.2 lists the three categories of data for LRT crossing safety analysis.

Table 5.2. Three categories of data for LRT crossing safety analysis

Collision History	Traffic and Physical Characteristics	Context Information
• Collision Frequency • Collision Rate • Collision Severity • Collision Patterns	• LRT Alignment Type • Gate Treatment • Annual Average Daily Traffic (AADT)	• Land Use Information • Distance to the Nearest Station

This information was either acquired by conducting a site survey, or retrieved from the available data sets using various GIS functions, and then attached as the attributes of each crossing in the LRT crossing table.

5.5.3.1 Collision Variables

- Collision Frequency

Collision frequency refers to the number of crashes which occur over a given time period. It is a reflection of the overall exposure at an intersection, and often increases with traffic volume. Collisions occurring in 2000 were first selected by locations within a radius of 75 meters of each LRT at-grade crossings.

- Collision Rate

Collision rate indicates the collision risk relative to the traffic volumes at a given location. It is an estimate of the chance that a vehicle at any point in time will be involved in a crash. It is usually expressed in terms of the number of collisions per million entering vehicles. The collision rate for each LRT at-grade crossing was calculated using the following equation:

$$CollisionRate = \frac{CollisionFrequency}{(AADT * 365)/1{,}000{,}000} \tag{5.1}$$

where AADT is the average annual daily traffic volume in 2000.

- Collision Severity

Collision severity is classified into three categories: fatal, injury and property damage only (PDO). Safety professionals are typically interested in the fatal and injury rate at a study site.

- Collision Patterns

Collision patterns represent general trends and common characteristics of collisions occurring at a given site. These trends and characteristics can be useful in identifying collision causes and can reveal safety problems at the collision site. The determination of variable(s) in the collision dataset for pattern analysis is discussed below.

5.5.3.2 Traffic Data

The only available traffic data are the AADT volumes from the ROADNET_2000 shapefiles. AADT ranges from 1,000 to 10,600 at the LRT at grade crossings. The information concerning the LRT alignment type and gate treatment type was digitized based on field surveys.

5.5.3.3 Context Information

Land use information was retrieved by locations, and attached as attributes of the LRT at-grade crossings. Land use types at these crossings, together with the number of occurrences in parentheses, are: agricultural (1), commercial (7), direct control (13), industrial (9), public education and recreation (4), public services (1) and urban reserve (4).

An Avenue script was used to calculate the distance from each at-grade crossing to the nearest LRT station (Wang 2000). This script allows the user to calculate the nearest network distance from one point data set (crossings) to another point data set (stations). The shortest distance between an at-grade and an LRT station is 25 meters, occurring in downtown.

5.5.3.4 Comparing Samples

To allow a comparison between signalized intersections and the at-grade LRT crossings, the same types of attributes were retrieved for each signalized intersection, and attached in the intersection table using similar procedures. Signalized intersections with the same land use type and falling into the same AADT range as that of the at-grade crossings were extracted for comparison.

5.5.4 Multiple Correspondence Analysis: Exploring the Collision Data Structure

There are 24 variables describing each collision occurrence in the data set. These variables include date of occurrence, time of occurrence, location of occurrence, collision type, collision consequence and environmental condition, among others. Some variables contain redundant information; it would be inefficient to include all these variables in the analysis. Hence, it is necessary to explore the data structure and reduce the data dimensions before further analyzing the collision data. Correspondence analysis (CA) is an exploratory technique for categorical and higher level data. The results of a CA provide statistics that are similar to those produced by factor analysis techniques. CA allows the exploration of the structure of categorical variables, and is a useful way to reduce redundant information in the data set, producing fewer data dimensions. Multiple correspondence analysis (MCA) or optimal scaling (SPSS 2000) may be considered an extension of simple correspondence analysis to more than two variables (StatSoft Inc. 2003, Fellenberg et al. 2001). A multiple correspondence analysis was carried out using SPSS v.11.5 to explore the data set of collisions occurring at LRT grade crossings. Figure 5.3 shows the 'Eigenvalues' of each dimension. The eigenvalues measure "how much of the categorical information is accounted for by each dimension" (SPSS 2000, p. 181). They are a measure of the total variance accounted for by the model. The higher the eigenvalue, the larger the amount of the total variance among the variables that loads on to that dimension. The largest possible eigenvalue for each dimension is 1.0. Ideally, the first two or three dimensions contain high eigenvalues (e.g. higher than 0.5) while the others express much lower values. Data dimensions are therefore reduced to the first two or three according to the eigenvalues. In our case the maximum eigenvalue in the first dimension was only 0.222, and there was not a significant "cut-off point" to determine which dimensions could be removed.

Dimension	Eigenvalue
1	0.222
2	0.165
3	0.150
4	0.121
5	0.111
6	0.106
7	0.100
8	0.094
9	0.089
10	0.087

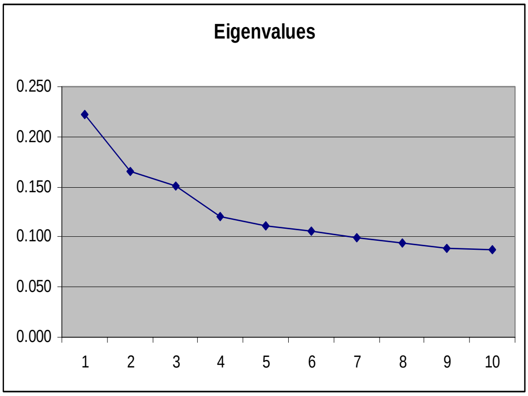


Fig. 5.3. Eigenvalues for the Intersection-collision data

The similarly low eigenvalues in each dimension indicated that the variables in the collision data are heterogeneous and all carry, to some extent, unique information. Reducing any of the variables might result in losing useful information concerning the observations. The heterogeneity of the collision variables reflects the random nature of collision occurrence. Traffic collisions can occur anywhere, anytime.

Table 5.3 shows the 'Discrimination Measures' of each variable in each dimension.[1] A discrimination measure, which can be regarded as a "squared component loading", is the variance of the quantified variable in that dimension (SPSS 2000, p. 184).

Large discrimination measures indicate that the categories of a variable are better separated along that dimension. Similar discrimination measures for different variables in the same dimensions indicate that these variables are related to each other. Related variables provide redundant information and therefore, some of them can be removed.

Since the eigenvalues are not very large in this analysis, discrimination measures in all ten dimensions were examined. The main findings were:

First, the variables 'Surface Condition', 'Environment Condition' and 'Month of the Year' were highly related as they all have high discrimination measures in dimension 3 and 4, but relatively low discrimination measures in other dimensions. These related variables described the winter road conditions in Calgary, when it is snowing (Environment Condition) and the roads are wet, slushy, snowy or icy (Surface Condition).

Second, the variables of 'Special Facility', 'Road Class' and 'Collision Location,' were related since they had similar discrimination measures in dimensions 1, 2 and 5. These variables described the locational information of collision occurrence.

Table 5.3. Discrimination measures for the Intersection_collision data

	Dimension									
	1	2	3	4	5	6	7	8	9	10
PRI_EVEN	0.460	0.229	0.125	0.008	0.112	0.370	0.438	0.552	0.335	0.331
SEVERITY	0.139	0.044	0.004	0.009	0.009	0.009	0.033	0.019	0.026	0.004
HIT_RUN	0.192	0.086	0.005	0.004	0.025	0.053	0.167	0.060	0.016	0.001
SCENE	0.408	0.239	0.013	0.001	0.001	0.005	0.025	0.006	0.004	0.000
DIAGRAM	0.118	0.107	0.006	0.003	0.003	0.011	0.028	0.013	0.004	0.000
SUR_CON	0.044	0.138	0.693	0.698	0.009	0.008	0.005	0.002	0.006	0.002
ENV_CON	0.024	0.086	0.442	0.743	0.013	0.024	0.015	0.024	0.036	0.025
SPE_FAC	0.521	0.189	0.040	0.006	0.203	0.114	0.030	0.069	0.151	0.422
RD_AL_A	0.070	0.072	0.035	0.003	0.291	0.073	0.012	0.036	0.015	0.044
RD_AL_B	0.064	0.072	0.015	0.004	0.320	0.126	0.023	0.020	0.001	0.004
RD_CLAS	0.375	0.278	0.053	0.004	0.347	0.102	0.157	0.327	0.049	0.023
LOCAT	0.673	0.623	0.116	0.019	0.183	0.534	0.147	0.020	0.034	0.022
MONTH	0.014	0.123	0.505	0.170	0.026	0.049	0.061	0.137	0.417	0.226
GRP_HOUR	0.011	0.030	0.049	0.017	0.019	0.008	0.256	0.028	0.152	0.113

---

[1] Further details on this analysis are available upon request from the authors.

The relatively homogeneous nature of these variables indicated that collisions tend to occur at certain types of location, such as, on undivided two-way roads (Road Class), at intersections (Collision Location), on divided roads (Road Class) and on highway interchange ramps (Special Facility). Third, the discrimination measures for the variable of ‘Primary Event’ (type of collision) were high in dimension 1, 6, 7, 8, 9 and 10. Conclusions can be drawn that this variable accounted for significant variance in the data set. The type of collision was related to the location of collision occurrence (dimensions 1, 6, 8 and 10) and time of occurrence (dimensions 7 and 9).

Fourth, discrimination measures appeared low in almost all dimensions for the binary variables of ‘Hit and Run’, ‘Scene Visit’ and ‘Diagram Available’ (the only exception was the relatively high value for the variable of ‘Scene Visit’ on the first dimension). These variables might be considered as noise in the dataset, and were ignored.

5.5.5 Crosstabulation, Confirming the Significant Variable

According to the results of a multiple correspondence analysis, the “Primary Event” variable appeared to be the most important variable in the collision dataset. A crosstabulation table was used to examine if the pattern of values for the “Primary Event” variable was associated with the two types of intersection (i.e. LRT or non-LRT crossing). Crosstabulation can be used to examine frequencies of observations that belong to specific categories on more than one variable (StatSoft Inc. 2003). By examining these frequencies, we can identify relations between crosstabulated variables. In this analysis, 13 categories of “Primary Event” (column variable) were crosstabulated with the two types of intersection (row variable). The Pearson Chi-square was 98.152. The significance value (0.00) indicated that there was a highly significant relationship between the variables of “Primary Event” and “Type of Intersection.”

Although the Pearson Chi-Square showed a significant relationship between the tested variables, the phi value of 0.094 in this test indicated that the relationship was quite weak and this was probably due to the large number of accidents.

5.5.6 Interpretations

Multiple correspondence analysis is a useful statistical technique to explore the structure of categorical data and reduce the data dimensions. The MCA results showed that some of the variables, such as the variables of ‘Surface Condition’, ‘Environment Condition’ and ‘Month of the Year’, in the data set provide exactly the same information, and could therefore be removed from the analysis. These results can also be used to structure a more efficient case base.

The crosstabulation analysis further confirmed that the variable “Primary Event” was associated with the variable “Type of Intersection”. LRT crossings experienced a different pattern of collision types. The pattern of the “Primary Event” variable at each crossing will be reviewed therefore, to determine the safety attributes of the various crossings.

5.5.7 Case Base Construction

The construction of a case base is crucial, since prior cases are the fundamental elements of a CBR system. New solutions are generated by retrieving the most relevant cases from memory and adapting them to fit the new situations. In CBR, knowledge is organized by cases. The CBR tool used in this study is the Knowledge software family from eGain Communication Corporation (eGain 2003). Knowledge Central functions as the administrator that manages the case bases and their users. Knowledge Author is the key tool for case input, modification and adaptation. Search results of the most relevant cases are presented in Knowledge Agent. In eGain Knowledge, the knowledge, or the descriptions of a case, are represented by “questions” and then grouped by “clusters.” Answers to these questions are used to refine the search results, or to lead to another question or cluster by actions. Answers in eGain Knowledge can be formatted as text (e.g. “LRT at-grade crossings”), HTML links, lists (e.g. a list of safety issues at a given LRT at-grade crossing), tables (e.g. numeric tables), and controls (to enable actions within the eGain Knowledge Software).

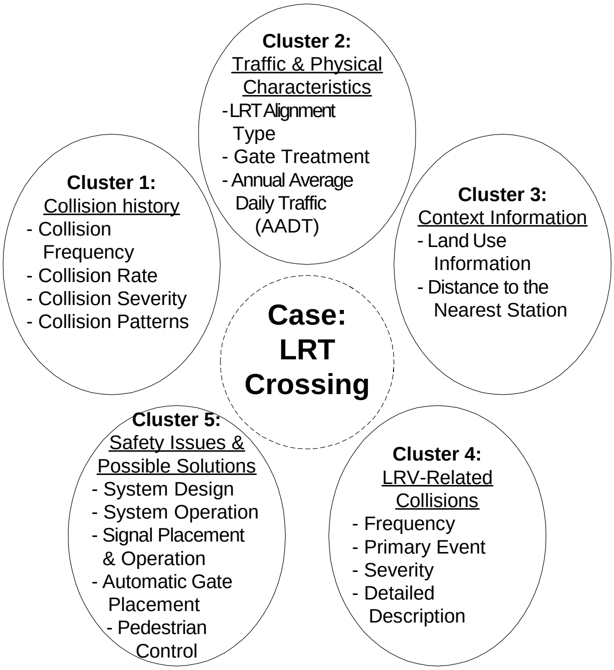


Fig. 5.4. Five clusters in the LRT crossing safety case base

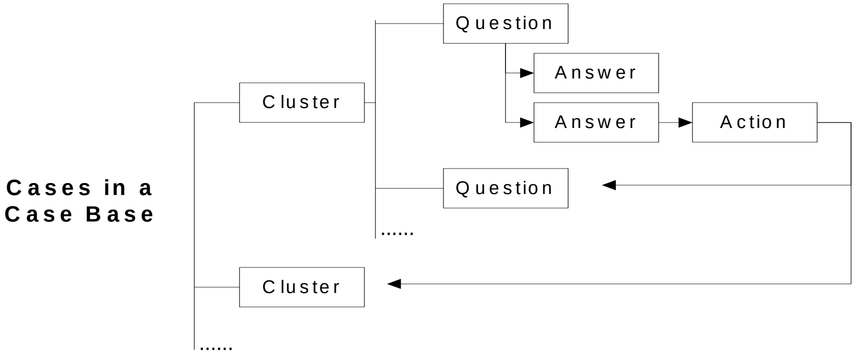


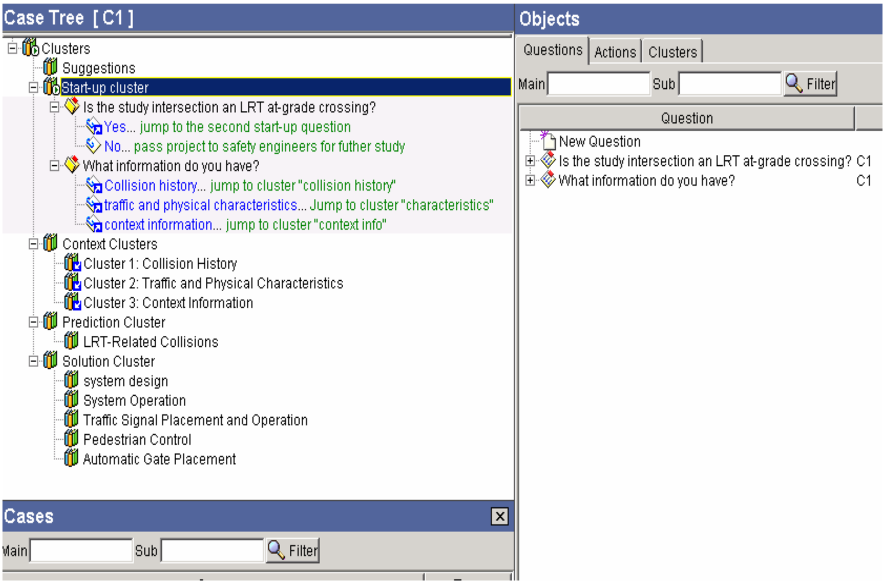
Fig. 5.5. Case base structure in eGain knowledge

The structure of a case base in eGain Knowledge is shown in Fig. 5.4.

In this project, traffic safety studies at the LRT at-grade crossings were considered cases. Five clusters were defined in the case base (Fig. 5.4). Cluster 1–3 contain criteria for searching matched cases, while Cluster 4 and 5 provide LRV-related collision prediction and possible solutions to enhance traffic safety at the study crossing. Figure 5.5 shows the case tree structure of these clusters in eGain Knowledge.

5.5.8 Case Retrieval

When traffic safety professionals conduct a safety study for an existing LRT crossing, or design a new LRT crossing, the prior cases in the case base can be used to provide relevant experiences and to recommend countermeasures. The case retrieval process begins by answering questions about the collision history, the traffic and physical characteristics and the environmental information concerning past collisions of a given crossing (Fig. 5.6). Based on the similarity of variables in these clusters, the eGain knowledge reasoner is able to search for the most relevant cases in the case base according to the matching scores. Potential safety issues at the given LRT crossing, and possible solutions to enhance the safety at this crossing can then be suggested according to the experiences retrieved from prior cases. Although case based reasoning tools can be operated with an incomplete and small case base, an inadequate case base reduces the chances of finding perfectly matched cases for a new situation. Due to the limited number of accident cases in this study, hypothetical "new situations" or "new crossings" were generated in order to retrieve a matched case with a satisfactory score.



Case Tree [C1]

- Clusters
  - Suggestions
  - Start-up cluster
    - Is the study intersection an LRT at-grade crossing?
      - Yes... jump to the second start-up question
      - No... pass project to safety engineers for futher study
    - What information do you have?
      - Collision history... jump to cluster "collision history"
      - traffic and physical characteristics... Jump to cluster "characteristics"
      - context information... jump to cluster "context info"
  - Context Clusters
    - Cluster 1: Collision History
    - Cluster 2: Traffic and Physical Characteristics
    - Cluster 3: Context Information
  - Prediction Cluster
    - LRT-Related Collisions
  - Solution Cluster
    - system design
    - System Operation
    - Traffic Signal Placement and Operation
    - Pedestrian Control
    - Automatic Gate Placement

Cases

Objects

- New Question
- Is the study intersection an LRT at-grade crossing? C1
- What information do you have? C1

Fig. 5.6. A portion of the case tree in eGain author

Table 5.4 shows the case of 7th Avenue and 4th Street S.W. intersection. Detailed descriptions of the traffic collision history, traffic and physical characteristics and context information of the crossing were stored in the criterion clusters (Clusters 1, 2 and 3).

Table 5.4. One case example for 7th Avenue and 4th Street

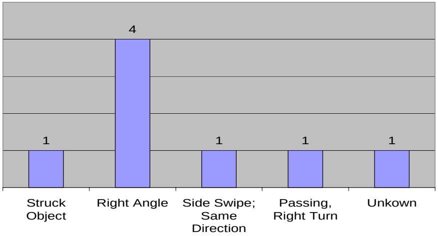
Cluster 1: *Collision History (2000)*	Collision Frequency	8
	Collision Rate	1.37
	Severity	Injury: 1 (12.5%); PDO: 7
	Primary Event Pattern	
Cluster 2: *Traffic and Physical Characteristics*	LRT Alignment Type	Type c, Non-Exclusive, Transit Mall
	Gate Treatment	No Gate, Traffic Signal
	AADT (2000)	16,000
Cluster 3: *Context Information*	Land Use Type	Direct Control
	Distance to Station	30 Meters, to 4th Street SW Station

Table 5.4. (cont.)

Cluster 4:	**Frequency**	**3 (2 in 1997; 1 in 1998)**
LRV-Related Collisions from 1996 to 2000	Collision 1: 19:00pm, 03-22-1997 Right Angle Property Damage Only (PDO)	#1 vehicle northbound on 4 St. SW in east Centre lane. Had stopped at red light. Proceed through light before it turned green. #2 Vehicle was westbound on 7Av. SW on tracks. Unable to stop to avoid #1 vehicle. Collision occurred.
	Collision 2: 6:58am, 04-8-1997 Right Angle Property Damage Only	Vehicle #2 westbound on 7 Av (LRT) proceeding through green light, when vehicle #1 southbound on 4 St. SW went through red light, last minute saw train, accelerated to avoid driver's side impact. Vehicle #1 stuck by train right rear quarter panel.
	Collision 3: 18:42pm, 06-27-1998 Right Angle PDO	Vehicle #1 southbound on 4 St. SW. Ran the red light at 7 Ave. Vehicle #2, a LRT, going eastbound, hit vehicle #1. Little damage
Cluster 5:	**Safety Issues**	**Possible Solutions**
	System Design Motorist disregard for regulatory signs at LRT crossings and grade crossing warning devices	Avoid excessive use of signs Photo-enforcement Maximize sight distance by limiting potential obstructions to 1.1 m (3.5 ft.) in height within about 30 to 60 m (100 to 200 ft.) of the LRT crossings (measured parallel to the tracks back from the crossing)
	Sign distance limitations at LRT crossings	
	System Operations Motorists disregard grade crossing warning devices and traffic signal	Adequately maintain LRT crossing hardware (e.g., routinely align flashing light signals) and reduce device "clutter" Public education and enforcement

This information was used to match cases with similar conditions. In Cluster 4, three LRV-related collisions and their causes were documented. This cluster provides a warning of potential LRV-related collisions at a crossing with similar characteristics. Cluster 5 listed the underlying safety issues at this crossing and the possible enhancement solutions. The information in Cluster 5 was generated based on an LRT study in 2001 (TCRP Report 69).

5.5.9 Conclusions

In this project, traffic collision records within the City of Calgary were imported to ESRI's ArcView 8.2 GIS software, where collisions were analyzed by their locations. Collision history and site-specific roadway characteristics of the LRT at-grade crossings were retrieved using various GIS functions. Traffic safety at these crossings was reviewed. The results of the studies were documented in eGain Knowledge, a case-based reasoning system, to enable more efficient and intelligent referencing for future safety studies of LRT crossings of the road network. GIS is an effective framework for integrating data from various sources, and for manipulating this multi-source data geographically. The acquisition of collision history and other site-specific characteristics of the study crossings would have been extremely difficult and time-consuming without a GIS solution. Moreover, the fully relational database architecture within ArcView GIS makes it possible to integrate with other database applications, such as most case based reasoning tools.

Case based reasoning methodology takes advantage of prior experiences when solving new situations. This reasoning approach is similar to the problem solving activities of a human brain. It also appears as a common research method in traffic safety, in which experience plays an important role. Another advantage within a case based system is that previous safety studies or projects can be documented and stored in a more interactive fashion. In such an interactive system, expertise and solutions to the previous problems are more reviewable and reusable. Thus professional knowledge can be circulated more efficiently within the traffic safety community.

Future efforts should be devoted to developing a more advanced system that implements CBR tools directly in a GIS environment, in order to monitor traffic safety at LRT crossings. Such a system would have several advantages. First, information retrieved from GIS layers could remain within the GIS database. CBR tools would provide a new indexing system to this information for performing queries to obtain matching cases. This would save the time and effort of converting database tables between the GIS and CBR systems. Second, the existing CBR tools are mostly focused on dealing with text information. Searches are triggered by entering key words, or answering questions. The integrated system would enable graphical representation of knowledge in the case base. For example, by pointing to a potential LRT crossing on a GIS map, relevant information concerning this crossing would be retrieved and the most similar existing crossings would be highlighted. Third, the fully integrated system would avoid requiring the user to switch between two applications, and thus would be more user-friendly.

Acknowledgements

The authors would like to thank Drs. Laurence Capus and Nicole Tourigny at Laval University who kindly provided a copy of their unpublished paper from the

1998 TRB conference and Mr. Cam Nelson for the use of the City of Calgary Accident database. Special thanks go to Mr. David Baines and eGain Corporation for providing the Knowledge package without which this research would not have been possible.

References

Abel D. et al. (1996) Decision support systems for sustainable development experience and potential: a position paper. Macau Workshop. Available at http://www.fes.uwaterloo.ca/Tools/docs/dssfsd_pp.pdf (Accessed on 08/19/2003)

Aha DW (2003) Case-based reasoning sources. Available at http://www.aic.nrl.navy.mil/ /~aha/research/case-based-reasoning.html (Accessed on 11/18/2003)

Alberta Motor Association (2003) Available at http://www.ama.ab.ca/mission_possible/ /mission.html (Accessed on 06/19/2003)

Arthur R (2002) Modeling hazardous locations with geographic information system. In: Rothe P (ed) Driving lessons, exploring systems that makes traffic safer. Edmonton AB: University of Alberta Press, pp 257–270

Arthur R, Waters N (1997) Formal scientific research of traffic collision data utilizing GIS. Transportation Planning and Technology 21: 121–37

Bayapureddy D (1996) Geographic information system for identification of high accident locations. Available at http://www.esri.com/library/userconf/proc96/PAP105/ /P105.htm (Accessed on 12/03/2001)

Boury-Brisset A, Tourigny N (2000) Knowledge capitalisation through case bases and knowledge engineering for road safety analysis. Knowledge-Based Systems 13: 297–305

Burkhard H (1998) Extending some concepts of CBR – Foundations of case retrieval nets. In: Lenz M, Bartsch-Spörl B, Burkhard H-D, Wess S (eds) Case-based reasoning technology: from foundations to applications. Lecture Notes in Artificial Intelligence 1400, Springer, Berlin Heidelberg New York, pp 17–50

Capus L, Tourigny N (1998) Road safety analysis: a case-based reasoning approach. Paper presented at Transpiration Research Board 77th Annual Meeting. Washington, D.C.

Carreker L, Bachman W (2000) Geographic information system procedures to improve speed and accuracy in locating cashes. Transportation Research Record 1719, National Academy Press, Washington, D.C.

Chris Naylor Research Limited (2003) Expert systems: a definition. Available at http://www.chrisnaylor.co.uk/Definition.html (Accessed on 09/04/2003)

Cirillo J, Council F, Griffith M, Hauer E, Paniati J (2000) Making safety management knowledge based. Transportation in the new millennium. Transportation Research Board, Washington, D.C.

City of Calgary (2003) ITS strategic plan, stakeholder consultation workbook. Workshop No.3, City of Calgary

Clayton D, Waters N (1999) Distributed knowledge, distributed processing, distributed users: integrating case-based reasoning and GIS for multicriteria decision making. In: Thill JC (ed) Spatial multicriteria decision making and analysis, a geographic information science approach. Ashgate, Aldershot, pp 275–304

Clayton D, Charlie R, Waters N (1998) Developing a knowledge base to incorporate traditional knowledge with GIS. Paper presented at Twelfth Annual Symposium on Geographic Information Systems. Toronto

Cortes U, Sanchez-Marre M, Ceccaroni I (2000) Artificial intelligence and environmental decision support systems. Applied Intelligence 13: 77–91

Cusack M (2003) Knowledge presentation in GIS-T for ITS decision support systems. Available at http://www.albany.edu/rcinf/students/KnowledgeRepresentationinGIS-TforITS.pdf (Accessed on 08/19/2003)

Dieng R, Giboin A, Amerge C, Corby O (1996) Building of a corporate memory for traffic accident analysis. Available at http://www-sop.inria.fr/acacia/Publications/1996/dieng-kaw96.html (Accessed on 03/25/2003)

eGain Communication Corporation (2003) Knowledge user’s manual. Available at www.egain.com

ESRI Website (2003) Available at http://www.esri.com

Fellenberg K, Hauser NC, Brors B, Neutzner A, Hoheisel JD, Vingron M (2001) Correspondence analysis applied to microarray data. In: Proceeding of the National Academic of Sciences of the United States of America, Vol.98, no.19

Glauz W, Bauer K, Migletz D (1985) Expected traffic conflict rates and their use in predicting accidents. Transportation Research Record 1026. National Academy Press, Washington, D.C.

Guariso G, Werthner H (eds) (1994) Environmental Decision Support Systems. Ellis Horwood-Wiley, Chichester New York

Haddad M (2003) Case-based reasoning case study: a case-based reasoning system for automated image interpretation of myocardial perfusion scintigraphy. Available at http://escher.cs.ucdavis.edu/mdi207/CBR.pdf (Accessed on 08/07/2003)

Haddon W, Jr., Suchman EA, Klein D (1964) Accident research – Methods and approaches. Harper and Row, New York

Haley Enterprise Inc. (2003) Available at http://www.haley.com/2045129008650259/ /THE.html (Accessed on 09/16/2003)

Hall J, Kim T, Darter M (2000) Cost-benefit analysis of geographic information system implementation. Transportation Research Record 1719. National Academy Press, Washington, D.C.

Hamilton-Finn Road Safety Consultants Ltd. (2000) Calgary blackspot pilot study. Available at http://www.hamiltonfinn.ca/

Herland L, Moller B, Schandersson R (2000) Knowledge acquisition, verification, and validation in an expert system for improved traffic safety. Transportation Research Record 1739. National Academy Press, Washington, D.C.

HSIS Summary Report (1999) GIS-based crash referencing and analysis system. (http://www.tfhrc.gov/safety/hsis/99-081.pdf (Accessed on 12/04/2001)

Joh D (1997) CBR in a changing environment. In: Leake DB, Plaza (eds) Case-based reasoning research and development. Lecture Notes in Artificial Intelligence. No. 1266. Springer, Berlin Heidelberg New York

Khattak AJ, Renski H (1999) PLAN<>HOV: case-based reasoning planning tool for high-occupancy vehicle lane analysis in a geographic information system environment. Transportation Research Record 1682. National Academy Press, Washington, D.C.

Kolodner J (1993) Case-based reasoning. Morgan Kaufmann, San Francisco

Kolodner J (1996) Making the implicit explicit: clarifying the principles of case-based reasoning. In: Leake DB (ed) Case-based reasoning, experiences, lessons, & future directions. AAAI Press, Menlo Park, pp 349–370

Lang L (1999) Transportation GIS. ESRI Press, Redlands

Leake DB (1996) CBR in context: the present and future. In: Leake DB (ed) Case-based reasoning, experiences, lessons, & future directions. AAAI Press, Menlo Park, pp 3–30

Leake DB, Kolodner J (1996) A tutorial introduction to case-based reasoning. In: Leake DB (ed) Case-based reasoning, experiences, lessons, & future directions. AAAI Press, Menlo Park, pp 31–66

Lenz M, Auriol E, Manago M (1998) Diagnosis and decision support. In: Lenz M, Bartsch-Spörl B, Burkhard H-D, Wess S (eds) Case-based reasoning technology: from foundations to applications. Lecture Notes in Artificial Intelligence 1400. Springer, Berlin Heidelberg New York, pp 51–90

Lin F, Sayed T, Deleur P (2003) Estimating safety benefits of road improvements: case based approach. Journal of Transportation Engineering 129: 385–391

Miller J (2000) The unique analytical capabilities geographic information systems can offer the traffic safety community. Transportation Research Board 79th Annual Meeting. Washington, D.C.

Nelson C (2003) Personal communication. (Traffic Safety Coordinator, City of Calgary)

Nyerges TL (2004) GIS in urban and regional transportation planning. In: Hanson S, Giuliano G (eds) The geography of urban transportation. The Guilford Press, New York, pp 163–95

Ozbay K, Mukherjee S (2000) Web-based expert geographical information system for advanced transportation management systems. Transportation Research Record 1719, National Academy Press, Washington, D.C.

Panchanathan S, Faghri A (1995) Knowledge-based geographic information system for safety analysis at rail-highway grade crossings. Transportation Research Record 1497. National Academy Press, Washington, D.C.

Paniati JF, Hughes W (2000) Making safety management knowledge based. Transportation in the New Millennium, Washington, D.C.

Parker MR, Jr., Zegeer CV (1988) Traffic conflict techniques for safety and operations: engineers guide. Report FHWA-IP-88-026, FHWA, U.S. Department of Transportation, Washington, D.C.

Peled A, Haj-Yehia B, Hakkert A (1996) Arc/Info-based geographic information system for road safety analysis & improvement. Available at http://www.esri.com/library/ /userconf/proc96/TO50/PAP005/P5.htm (Accessed on 02/09/2002)

Richter M (1998) Introduction. In: Lenz M, Bartsch-Spörl B, Burkhard H-D, Wess S (ed) Case-based reasoning technology: from foundations to applications. Lecture Notes in Artificial Intelligence 1400. Springer, Berlin Heidelberg New York, pp 1–16

Riesbeck C (1996) What next? The future of case-based reasoning in postmodern AI. In: Leake DB (ed) Case-based reasoning, experiences, lessons, & future directions. AAAI Press, Menlo Park, pp 371–388

Saccomanno F, Fu L, Roy R (2003) GIS-based integrated model for road accident analysis and prediction. Available at http://www.civil.uwaterloo.ca/saccomanno/Publications/ /GIS_PapTRB.doc (Accessed on 04/28/2003)

Smith R (2000) Implementing GIS-based highway safety analyses: bridge the gap. Available at http://www.esri.com/library/userconf/proc00/professionalpapers/PAP888/ /p888.htm (Accessed on 11/28/2001)

Souleyrette R, Strauss T, Estochen B, Pawlovich M (1998) GIS-based accident location and analysis system (GIS-ALAS). Project Report: Phase 1. Available at http://www.ctre.iastate.edu/Research/gis-alas/papers/gisalasph1.pdf (Accessed on 03/26/2002)

Spring G, Hummer J (1995) Identification of hazardous highway locations using knowledge-based GIS: a case study. Transportation Research Record 1497. National Academy Press, Washington, D.C.

SPSS Inc. (2000) SPSS user's manual

StatSoft, Inc. (2003) The statistics homepage. Correspondence analysis. Available at http://www.statsoftinc.com/textbook/stcoran.html (Accessed on 06/19/2003)

Suttayamully S, Hardpriono F, Hemeth Z (1995) Knowledge acquisition, representation, and knowledge base development of intelligent traffic evaluation for prompt incident diagnosis. Transportation Research Record 1497. National Academy Press, Washington, D.C.

TCRP Report 17 (1996) Integration of light rail transit into city streets. Transportation Research Board. National Academy Press, Washington, D.C.

TCRP Report 69 (2001) Light rail service: pedestrian and vehicular safety. Transportation Research Board. National Academy Press, Washington, D.C.

Thill J (2000) (ed) Geographic information systems for transportation in perspective. Transportation Research C 8: 1–144

Thompson R (2001) Targeting traffic problems through an automated collision database. Project presented at 27th International Forum on Traffic Records & Highway Information Systems. Available at http://www.traffic-records.org/forum2001/ /Best%20Practices/Thompson.htm (Accessed on 11/10/2002)

University of Kaiserslautern (2003) Case-based reasoning on the web. Available at http://www.cbr-web.org/CBR-Web (Accessed on 06/19/2003)

Vollrath I, Wilke W, Bergmann R (1998) Case-based reasoning support for online catalog sales. IEEE Internet Computing 2: 47–54

Wang L (2000) Network distance.ave. (Downloaded from ESRI website at www.esri.com)

Waters N (1988) Expert systems and systems of experts. In: Coffey W (ed) Geographical systems and systems of geography, essays in honour of William Warntz. Dept. of Geography, University of Western Ontario, London, pp 173–187

Waters N (1989) Expert systems within a GIS: knowledge acquisition for spatial decision support systems. In: Proceedings of the National Conference for Geographic Information Systems. The Canadian Institute for Surveying and Mapping, Ottawa, pp 740–759

Waters N (1999) Transportation GIS: GIS-T. In: Geographic information systems, vol 2, Management Issues and Applications, Second Edition

Waters N (2002) Geographic information systems, case-based reasoning and system design. In: Rothe P (ed) Driving lessons, exploring systems that makes traffic safer. University of Alberta Press, Edmonton, pp 247–256

Watson I, Marir F (1994) Case-based reasoning: a review. Available at http://www.ai-cbr.org/classroom/cbr-review.html (Accessed on 08/07/2003)

Weber G, Schult T (1998) CBR for tutoring and help systems. In: Lenz M, Bartsch-Spörl B, Burkhard H-D, Wess S (ed) Case-based reasoning technology: from foundations to applications. Lecture Notes in Artificial Intelligence 1400. Springer, Berlin Heidelberg New York, pp 255–272

Webnox Corp. (2003) Hyper dictionary. Available at http://www.hyperdictionary.com/ /computing (Accessed on 09/06/2003)

Wisdo C (1997) A scalable approach for question based indexing of encyclopedic texts. In: Leake, Plaza (ed) Case-based reasoning research and development. Lecture Notes in Artificial Intelligence. No. 1266. Springer, Berlin Heidelberg New York