Expert Systems with Applications 37 (2010) 5788–5794

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Data mining application on crash simulation data of occupant restraint system

Zhijie Zhao *, Xianlong Jin, Yuan Cao, Jianwei Wang

School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, PR China

ARTICLE INFO

Keywords:
Data mining
Occupant restraint system
Attribute selection
Decision tree
Crash simulation

ABSTRACT

This article presents an application of data mining method on finite element data and crashworthiness result data of an occupant restraint system. According to the characteristics of the CAE (Computer-Aided Engineering) data, a framework for data preparation is developed based on object-oriented programming concepts. Training sets are built from data recorded in 98 crash simulations that adhere to FMVSS208, the America occupant crash protection testing standard. Relationship between design parameters and system effectiveness is implied in these data sets. Decision tree using C4.5 algorithm and attribute selection method based on attribute's estimated importance are introduced to perform data mining on the building of training sets. The result yielded by data mining endows us with a deeper insight into the interrelations between the key design parameters and the performance of the occupant restraint system in crash simulations. Finally, the learned rules are tested on the real crash simulation data sets. The result of the testing shows that these rules are proper, and can been used as a guidance for the design of the occupant restraint system.

1. Introduction

With the arrival of computer age and the rapid development in high performance computing area, the finite-element-based CAE simulation has been gradually overtaking the traditional mechanical experiments as the main means for research and development (R&D) in automobile industry. The general utilization of CAE tools has remarkable advantages in cutting the R&D expenditure and shortening product cycles, yet it also brings the challenge to the management of massive data. Huge data sets are generated in R&D process, since a large number of crash simulation experiments are carried out to verify the design. The intrinsic relationship between design parameters and car performance is usually implied in the massive data, and can hardly be identified by hand. To uncover the underlying relationship and reuse the data as the guidance for the design, the data mining technology is brought into usage.

Data mining is defined to be the exploration and analysis, by automatic or semiautomatic means, of large quantities of data stored either in databases, data warehouses, or other information repositories to discover interesting knowledge including meaningful patterns and rules (Berry & Linoff, 1997; Han & Kamber, 2001). It has been widely used in fields like finance, market analysis, biological and medical research, telecommunication, traffic accident assessment, etc. (Alonso, Caraça-Valente, González, & Montes, 2002; Bayam, Liebowitz, & Agresti, 2005; Chen & Du, 2009; Huang, Chang, & Wu, 2009; Jeng, Chen, & Liang, 2006; Turhan, Kocak, & Bener, 2009). The CAE analysis is a complicated process that consists of a series of steps, including pre-processing of the FE model, the solving stage for the simulation, and post-processing of the result files. CAE data have different formats in different stages. In pre-processing phase, data is usually contained in FE (Finite Element) model files in ASCII format. In the result files, there are generally two kinds of files, the ASCII files which specify motions and force of nodes and elements, and the binary files which illustrate the kinematic and dynamic features of the results by animation. In the present paper, attention is given to the data mining on FE model files and result files in ASCII formats, with the goal of obtaining the relationship between design parameters and the occupant safety level in crash simulations.

The application of data mining is mainly achieved by three steps. The first step is data preparation, when necessary data is extracted from the original data pool and pre-processed to get target data. The second step is data mining, which includes performing knowledge discovery for target data, and ultimately forming the knowledge. The last is knowledge verification, when knowledge is applied in practical cases to verify its correctness. The details of these three steps are discussed in this paper. Besides, most of the data mining algorithms adopted in this paper are based on WEKA (Waikato Environment for Knowledge Analysis) package (Witten & Frank, 2005). WEKA includes several standard machine learning methods which can help users derive

* Corresponding author. Address: Room 225, Advanced Manufacture Building, School of Mechanical Engineering, Shanghai Jiao Tong University, No. 800, Dongchuan Road, Shanghai, 200240, PR China. Tel.: +86 21 34206099; fax: +86 21 34206338.

E-mail address: zhijie@sjtu.edu.cn (Z. Zhao).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.02.029

useful knowledge from data sets that are otherwise too large to be analyzed by hand.

2. Data preparation

Typically, data preparation takes up 80–90% of the total workload for data mining process (Klosgen, Zytkow, & Zyt, 2002). Original data must be pre-processed, otherwise, useful knowledge can rarely be obtained from data mining. The pre-processing of data includes extracting and generating parameters from FE models and simulation results, and converting them to standard formats so that they conform to the data mining algorithm. Data that has been processed by data preparation procedure exists in the form of instance sets, each of which is composed of several attributes. For the data mining on the occupant restraint system, each crash simulation result is considered as an instance. Class attribute is obtained from simulation results, while other features are got from the vehicle FE models. In the following paragraphs, the generating process of these two kinds of features will be discussed.

2.1. FE model of the occupant restraint system and its parameter extraction process

The CAE data used in this article are based on the FE models of the vehicle which includes an occupant restraint system as illustrated in Fig. 1. The FE model consists of about 600 parts, 160,000 nodes, and 160,000 elements. The occupant restraint system contains a Hybrid III dummy, a three-point seat belt, and a foldable airbag. The FE model for seat belt also includes seat belt elements, sliding ring elements, seat belt pretensioner, seat belt retractor, and the sensing elements that trigger the seatbelt pretensioner and retractor. The airbag model (http://www.ncac.gwu.edu/vml/archive/ncac/interior/DRVR-Airbag-V2.key.gz) is built according to the norm of National Crash Analysis Center of America, with the pressure of the airbag being uniform.

The occupant restraint system consists of many components and involves a large number of design parameters. As suggested in references (Sinha, 2007; Thiele, 2006), nine variables coming from the seat belt and the airbag are chosen as parameters. Table 1 lists the meanings and units of these parameters.

For confidentiality of their key techniques, car companies do not usually share a large amount of data for research. In this paper, a total of 98 cases of experimental data for crash simulation are obtained by way of orthogonal experiment design (Chen, Zhang, Yang, & Niu, 2007), which includes nine influencing factors and seven levels as illustrated in Table 2.

FE models are usually in the form of ASCII files with the influencing factors being interrelated, which makes it possible that the parameters as shown in Table 1 be generated automatically by programming. In this paper, we rebuilt the class objects of seat belt and airbag using JAVA language based on the object-oriented concept. Fig. 2 shows the class structure in the developing environment Eclipse. In the class structure, member variables include material ID, structure ID, units and other base information of seat belt and airbag; member functions stand for the algorithms adopted in the generation of parameters as shown in Table 1. For example, to obtain the parameter sb_loader as shown in Fig. 3, we can firstly obtain the loading curve ID according to the retractor type, and then find the discrete number group that define the loading curve, and finally calculate the area confined by this curve and the time axis.

2.2. Assessment of occupant safety in frontal impact and generation of class attribute

The result data was obtained from a 100% rigid wall frontal impact simulation test with the collision speed being 50 km/h as illustrated in Fig. 4. The test was in accordance to FMVSS208, the America occupant crash protection testing standard. The simulation lasted 120 ms, and the solver being used was LS-DYNA (SMP, Symmetrical Multi-Processing, version).

The impact of parameters of occupant restraint system to the occupant protection in crash simulation is analyzed in this article. To evaluate the occupant injury under frontal impact, we adopt the Weighted Injury Criterion (WIC). The value of WIC coming from the crash result reflects the degree of the occupant injury. A larger value of WIC indicates more severe injuries. In the American standard FMVSS208, WIC is defined as:

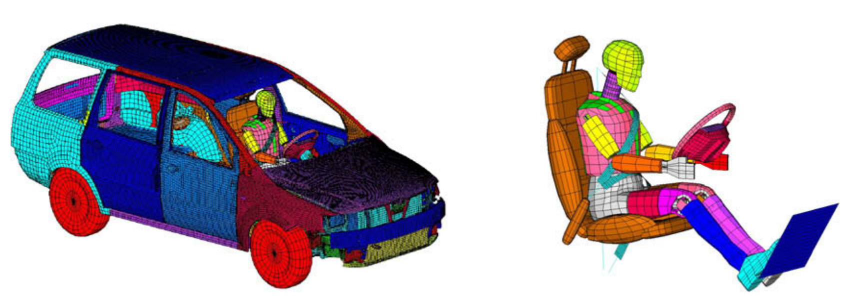
$$WIC = 0.6\left(\frac{HIC}{1000}\right) + 0.35\left(\frac{C_{3ms}}{60} + \frac{C_{comp}}{0.0762}\right) \Big/ 2.0 + 0.05\left(F^{femur}_{keft} + F^{femur}_{right}\right) \Big/ 20.0 \tag{1}$$

Here, HIC (Head Injury Criterion) stands for head injury criterion whose value is as following

$$HIC = Max\left[\left(\frac{1}{t_2 - t_1}\int_{t_1}^{t_2} a_{Head}\,dt\right)^{2.5}(t_2 - t_1)\right] \tag{2}$$

Here, a_{Head} is the resultant acceleration of head center; t_1, t_2 stands for the starting time and termination time during the process of maximizing HIC. Generally, when the time interval is 36 ms, $HIC \leqslant 1000$. C_{3ms} is the value of chest 3 ms norm; the maximum linear acceleration sustained for 3 ms or longer in the center of the chest should not exceed 60 g (TNO, 2004). C_{comp} is the chest contraction (unit: m); should be less than 75 mm. F^{femur}_{left} is the maximum axial force on left thighbone; should be less than 10 kN. F^{femur}_{right} is the maximum axial force on right thighbone; should be less than 10 kN.

Using object-oriented programming language as described in Section 2.1, HIC can be obtained from the simulation result of head



(a) Vehicle FE model (b) Occupant restraint system FE model

Fig. 1. Occupant restraint system and vehicle FE model.