A decision-making model for explaining driver behavior.

MSc. Marjana Cubranic - Dobrodolac

Review of PhD Thesis

The aim of the author is to propose a model enabling the estimation of driver´s behavior based on empirically determined psychological data. It is primarily an estimate of the driver's predisposition to commit an accident. Based on an extensive study of the literature dealing with this topic (which is summarized in the first part of the dissertation), the author decided to use four questionnaires. She assumed that they carried the most information about the possibility of traffic accidents of certain driver. These are questionnaires carrying psychological data on the aggressiveness and impulsiveness of drivers (ADBQ, BIS-11), the Manchester Driver Attitude Questionnaire (DAQ) and the Questionnaire for Self - Assessment of Driving Ability. The answers to the questions in the individual questionnaires were evaluated up to 6 levels, the total score of about 20 questions for the driver was taken as the value of the input variable for the proposed system. The aim was therefore to design a system whose inputs will be relevant variables generated by these four questionnaires and the output will be an estimate of propensity road traffic accidents.

The author conducted extensive empirical research and managed to collect data for 305 drivers. For many reasons, the set is not very homogeneous, the predominance of professional drivers, the predominance of male drivers, etc. and therefore the data are supplemented by data from the Demographic and driving history questionnaire. There are also data on the number of accidents that the driver has committed during his or her career and which actually form the output of the proposed system.

Based on this data, the author proceeded to design the required system. She chose two basic approaches. A model based on linear regression and model using fuzzy logic. The regression model basically assumes that the output, is the estimate of propensity road traffic accidents, is a linear combination of quantities obtained from psychological questionnaires. The linear regression model gives surprisingly good results, although the author does not give any reasons why a linear model should suffice. (See comments below).

The second variant of the proposed system is a fuzzy model. The author used the classic Mamdanian system, based on rules like

$$\text{If } x1 = A \text{ and } x2 = B \text{ then } y = C$$

She performed fuzzification using different versions of triangular fuzzy sets, first evenly distributed on the definition domain of variables. Furthermore, using various optimization methods and with the help of the evolutionary method Bee colony optimization, modified the membership functions of individual variables. She infered the rules from the data using the Wang-Mendel method. She created a set of knowledge bases for various models, based on individual variables and their combinations. IT is not mentioned in the work how the defuzzification was performed.

Finally, she tested different types of fuzzy models and compared them with a regression model. It turned out that the best results were obtained by an optimized fuzzy model using all four input variables, but that they did not differ much from the results provided by the regression model as a linear combination of all four input variables.

Now for comments. Regarding the regression model, the author assumes that any dependence can be approximated by a linear combination of input variables. It is not stated what led her to this. (However, this is usually the first step in using regression in empirical research). At the end of the work, Figures 35-38 show that this was a legitimate assumption that the dependence between the individual quantities and the output can be approximated by a straight line. Furthermore, for the use of regression methods, it is assumed that the individual quantities have a normal distribution. The author did not test the normality, but due to the number of samples (< 300) the central limit theorem works, and the condition is probably satisfied.

Note the conclusions on page 63 regarding the correlation coefficient can be applied only in the case of a linear dependence. If there is no dependence between the quantities, the correlation coefficient is zero. It is not true, but the statement that when the correlation coefficient is zero, there is no dependence between the quantities. The correlation coefficient can be zero even if the quantities are in a nonlinear functional dependence. If we want to find out what information the input quantity (or their combination) carries for the output quantity, we must use the mean mutual information

$$T(X: Y) = H(X) + H(Y) - H(X, Y)$$

Where H (X), H (Y) are the entropies of the respective variables H (X, Y) their associated entropy. It works regardless of whether the dependence is linear or nonlinear.

Next to the fuzzy model. Used rules are in the form of a logical implication. In classical logic we can express the implication in several ways, in fuzzy logic in much more ways. The thesis do not state how the implication was interpreted, how the logical product between antecedents was interpreted and how the individual fuzzy sets generated by individual firing (active) rules were composed. That is, how do we get the output fuzzy set for the active rule and how do we compose the outputs for several rules operating simultaneously. In fact, a description of the inference mechanism is missing. There is no mention of how defuzzification was performed. I assume the most popular method of the center of area. The author had to use all the mentioned steps and probably used them successfully, given that the resulting optimized fuzzy system works very well, but for some reason they are not mentioned in the work. I suppose she will coplete it during to her defense.

Otherwise, I have no major comments. Just one formal comment. A list of used symbols would greatly benefit from the work, which would make reading easier. Overall, I consider the work to be very good. The author showed how a verbal, qualitative psychological evaluation can lead to a formal system that provides a good estimate of the propensity of road traffic accidents. The work is mostly based on the results of a set of works that the author has successfully published. The author has considerable experience with this issue, which is shown by the list of publications, where she is the author or co-author, and which are directly related to the topic. In conclusion, I would like to say that the dissertation convincingly documents the professional abilities of the author and her ability to do scientific work. I

believe that the work of Ms, Marjana Čubranić Dobrodolac MSc satisfy all the requirements for a dissertation in the sense of Act 111/98 Coll. and therefore I recommend it for the defense.

Doc. Ing. Petr Vysoký, CsC.

Ústav dopravní telematiky,

Fakulta dopravní, ČVUT v Praze