Identification of High Crash Road Segment using Genetic Algorithm and Dynamic Segmentation

Amin Mirza Boroujerdian [1], Masoud Fetanat [2], Vahid Abolhasannejad [3,4]

Received:21.02.2015 Accepted: 13.02.2016

Abstract

This paper presents an evolutionary algorithm for recognizing high and low crash road segments using Genetic Algorithm as a dynamic segmentation method. Social and economic costs as well as physical and mental injuries make the governments perceiving to road safety indexes in order to diminish the consequences of road accidents. Due to the limitation of budget for safety improvement of all parts of the road, the road segments with more accidents should be recognized for safety budget assignment. So, considering this fact it's important to identify the segments with high and low number of accidents to optimize the road safety program. In this study, a novel chromosome coding method and a fitness function which are consistent with Genetic Algorithm are proposed. The proposed methodology is also validated by using two mathematical parameters so that the results confirm that the proposed modeling works properly. Afterward, the proposed dynamic segmentation method is compared with the other static segmentation methods along 51 km of Shahrood–Sabzevar highway. The proposed method may have more advantages comparing to static segmentation methods for all of the performance indexes which were considered in this study. The proposed method has a variance about two times higher than the one for accident density in comparison with the other static segmentation methods. About 62% and 34% improvement is achieved in average of segments accident density and total segments density respectively in comparison with the other fixed methods.

Keywords: Genetic Algorithm, dynamic segmentation, road accident segments

Corresponding author E-mail: boroujerdian@modares.ac.ir

1 Department of Civil & Environmental Engineering, Tarbiat Modares University, Tehran, Iran

2 Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran

3 Department of Civil Engineering, Birjand University of Technology, Birjand, Iran

4 School of Transportation, Southeast University, Nanjing, China

1. Introduction

Because of the complexity of the accident causes and dependence on the location of accident, as well as other effective parameters, recognizing high and low crash road segments is a challenging problem. The total cost resulting from accidents is much bigger than the improvement and maintenance cost of special sections of the roads, so identification and prioritization high crash road segments is of high importance. In order to identify the high and low accident segments of a road, the road of interest should be divided into some segments properly by defining some suitable functions, right variable or suitable performance indexes to predict the probability of accident risk using the data including physical and traffic characteristics of the road. Some of the segmentation methods are explained in the literature review section.

2. Literature Review

Safety experts are able to determine the special section of the roads in which the accidents happen to identify high and low crash road segments. In many countries, the previous research in the field of road segmentation methods has shown that identification of high and low crash road segments has been implemented by dividing the road into some segments with equal lengths and considering number of accidents in each segment. The lengths of segments are different from country to country, but the specific lengths of the road can be the same or different from each other.

Federal Highway Administration reported some of fundamentals regarding road segmentation by which the length of high crash road segments is considered 0.3 mile [Federal Highway Administration, 1981]. Vistisen in 2002, identified high crash road segment by dividing the road segments into different forms based on the road sections and intersections in Denmark [Vistisen, 2002]. An experiment based on the Poison distribution is conducted to identify high crash road segments. The lowest accident number for considering a segment as a high crash road segment is four accidents during 5 years. However, the length of the segments depends on the number of normal accidents in each segment not on the other form of unusual accidents [Vistisen, 2002]). Kanono and Aleri studied on the degree of road safety by dividing some part of the road into 2-mile segments in order to identify high crash road segments[Kononov and Allery, 2003]. In another research, the length of road segments is equal to 0.25 miles and each segment should not be too long or too short. The segments with length of 0.25 miles or less and between 0.25 to 0.5 miles can be defined as the segments with special conditions (Pant et al, 2003). Texas Transportation Institution has proposed a segmentation method for identification of high crash road segments so that the length of each segment should be at least 0.1 mile [Bonneson and Zimmerman, 2006]. Geurts in 2006 proposed a method for identification of high crash road segments in Belgium based on the Police report. Using his method, the segment with at least three accidents during three years are categorized as a high crash road segment. Geurts also proposed two definitions to identify high crash road segments in 2006. In regions which are far from the residential areas, high crash road segment is defined while there is at least four accidents during three years in a segment with the length of less than 1 Km, but in residential areas, high crash road segments are defined if at least four accidents happen within three years in that segment with the length of less than 0.1 Km [Geurts, 2006]. Troche in 2007 proposed a road segmentation method by which the road is divided into 0.2-Km segments that may include highway, intersection or other segments of the road. Furthermore, static segmentation considering a fixed length of 2.5 Km was studied in Austria [Troche, 2007]. Saffarzadeh et al. proposed a cause-based model for identifying hazardous locations for pedestrians. In this method, by taking the accident frequency, accident severity and accident cause into account and using one of the multi attribute decision making methods, the mentioned segments were prioritized [Saffarzedeh et al. 2007] . The segments which satisfy the specific standards of high accident segments considered as high crash road segments. Elvik in 2008 as well as some other European researchers suggested the approach of high crash road segment management and road network analysis [Elvik, 2008]. In Iran, road is divided into 1-Km segments and then the accidents happening within one year are considered. Table 2 summarizes the common definition of high crash road segments in each state [Research Institute of transportation, 2008]. Boroujerdian et al. developed a new methodology for prioritizing high crash road segments based on the accident cause [Boroujerdian et al. 2009]. According to this model, a segment may be compared with another and identified to be highly accident-prone with the same cause, whereas the total number of accidents that occurred in the former may be less than that in the latter. In 2014, identifying the length of high crash road segments is performed with a dynamic-based model by wavelet theory. This method was multi-scale segmentation. In fact, this model identifies high crash road segments with different lengths (short and long). Running the model into a real case for identifying 10–20 percent of high crash road segments showed an improvement of 25–38 percent relative to existing methods [Boroujerdian et al. 2014].

High crash road segments with variable lengths are one of differences between static and dynamic methods. Start point in segmentation methods is another difference between them. Some countries use the fixed segments to analyze their accident data, but some others use the floating segment for identification of high crash road segments. It can be said that floating fixed segment is more accurate in comparison with the other fixed methods.

One of the ways by which the segments are classified is to divide them based on the fixed or variable lengths of the segments. In static segmentation method, the length of each section is constant. Dividing the road into some segments having fixed length recognizes high crash road segments along that road and the segments with high accident density are identified as high accident-prone segments. Some deficiencies of static segmentation methods are as follows:

a- Distributing accidents within a few segments and poor identification of segments with high accident density;

The boundary of some segments may be fallen within the high accident density sections and it leads dividing the segments so that bunch of accidents are located within two different segments and consequently the results of segmentation model is not perfectly reliable.

b- Poor matching between the length of segments in static segmentation methods and the length of the real high crash road segments.

If the length of high or low crash road segments is more or less than the length of each fixed segment, the suitable length of each segment can't be identified by the current static methods. Thus, there will be existing considerable errors for identification of high crash road segments by fixed-length segmentation method.

Considering above discussion, the current methods for segmentation of high crash road sections are not properly working. Therefore, instead of fixed lengths segmentation method, variable length segmentation model should be utilized based on the length, physical characteristic and statistical data of the road.

3. Methodology

The idea of this paper is to model and identify the road segments with high crashes and variable length. In order to make an identification model, road mathematical equations related to number of road accidents, length of segments, physical characteristics of the roads and other effective parameters are needed. As there are no such equations in real world, they should be modeled in a right way. Furthermore, the final goal of the paper is to find segments with highest and lowest crashes. Changing the length of each segment makes the problem more challenging to be solved. In addition, the length of one segment affects the length of other segments and will also influence on the number of accidents within other segments. Therefore, the problem is converted to an optimization problem to find optimal number of segments and location of the boundaries of segments. Some of the optimization problems have many variables and can't be solved using ordinary optimization methods easily. In addition, the Genetic Algorithm (GA) can be used for solving multi-dimensional, non-differential and discontinuous optimization problems. In this paper, it is supposed to solve an optimization problem with high dimension, non-differential and discontinuity. It should be also mentioned that while modelling a natural process with mathematical equations, one of the approaches is to design a chromosome encoding method to formulate the cost function (fitness function) and constraints. Chromosome encoding method is an approach that converts real concepts and numbers to binary strings in a meaningful way in order to be used in a fitness function properly. If the chromosome encoding method is designed to model a natural process, it's easy to use the genetic algorithm for solving the optimization problem since the genetic algorithm is deal with binary strings. The approach of the paper is to model and identify the high crash road segments with variable length in a real road. To do so, a methodology is proposed that converts the location of accidents, boundaries of segments, number and density of accidents within each segment to binary strings and fitness function. Therefore, genetic algorithm and a method for dynamic recognition of road accident segments are proposed in this section. This section includes some sections: section of chromosome coding design, section of proposed methodology, section of fitness function, section of genetic operators and section of performance indexes. GA is known as an iterative stochastic global search algorithm, which can solve the optimization problems by inspired processes from natural evolution. GA serves a large number of candidate solutions for a problem that are called population. Every solution is named as chromosome which is usually coded as a binary string. In fact, the population is made of a group of chromosomes. After chromosome has been decoded to real numbers, the fitness function of each chromosome is calculated by a proposed fitness function. One of the most important problems in GA is to find the right performance function which is called fitness function. The other important challenge in GA is to find genetic encoding algorithm to make a perfect problem which is explained more in the next sections.

3.1. Chromosome Coding Design

Chromosome coding algorithm is a challenging part of the GA. In this section, a method is designed to convert the real numbers to binary numbers so as to use for evaluation of the fitness function. It is also needed to convert the binary numbers to real numbers to see the optimal solution. In binary numbers, there are only "0" and "1". A method should be defined to show each segment and their boundaries. In our proposed method, "0" in the chromosomes means the start or end point of each segment and "1" means the continuity of each segment. "0" in binary strings means a segment is finished or started. Number "1" plus one determines the length of each segment. To clarify more, an example is brought here; the binary code like "011000101101110" has a segment with length of 3, then a segment with length of 1, a segment with length of 1, a segment with length of 2, a segment with length of 3, and a segment with length of 4 respectively and totally has 6 segments with total length of 14. The number of accidents in each segment can be computed by finding the locations of each accident corresponding to the segments after the length of each segment is determined by the mentioned approach. The accident density is calculated through dividing the number of accidents in each segment by the length of each segment. The binary strings are used with length of 52 (a string includes 52 "0" and "1") 2 points at the start and end of the binary string are always "0" since the road has 51 points in 51 Km. [Boroujerdian et al, 2015].

3.2 Proposed Method

The first phase of the proposed methodology is to collect the road accident data as well as the location of road accidents and initialize the Genetic Algorithm. It means some binary strings are produced to define the road segments randomly by ones and zeros of chromosomes. The second phase is to compute the boundaries and the length of each segment with the solution of GA. The boundaries and the length of each segment is calculated by chromosome coding design (section 3.1). Determining the number of accidents in each segment is done for the third phase. When the boundaries and the length of each segment are computed and compared with the number of accidents happening in a special location of the road, the number of accidents of any segment can be calculated easily. In phase four of the flowchart, the accidents density in each segment is estimated. It can be computed by a simple function (dividing the number of accidents by the length of that segment) for each segment. In the next phase, the proposed fitness function is evaluated to find the high crashes road segments with variable lengths. Stop condition in phase 6 is an iteration limit in the proposed method or value of the improvement of the solution in next iterations compared with previous iterations. The last stage will be implemented when the stop condition is satisfied. GA algorithm continues if the stop condition is not satisfied and the process goes to phase 2 to calculate the boundaries and the length of each segment. This iteration loop is repeated over and over to find the global optimal solution. In this research, stop condition is a certain number of iteration. In last stage, the optimal solution is found that is the boundary of the segment. Then the accident density, variance and average of the accident density are calculated for each segment to compute the fitness function. Figure 1 shows the procedure of the proposed method.

3.3 Fitness Function

One of the most important challenging sections in the optimization problem is to find a suitable fitness function. Therefore, it should be chosen correctly to find a global optimal solution. However, it should be a simple fitness function to consume less time for calculation of the fitness function. In our problem, the time is not an important issue since the optimization process is done in off-line mode. After one iteration in the proposed method, the chromosomes are ranked in descending order based on the fitness function values. The chromosomes with worse fitness function values are arranged in the lower position of populations. Here, the chromosome are chosen with lowest fitness function value as a best solution since a minus is added in the proposed fitness function and the chromosome with the lowest fitness function value is ranked in the first position of solutions. [Goldberg, 1989; Engelbrecht, 2007; Huang et al, 2009; and Samanta, 2004].

The goal of this research is to find the high and low crash road segments, therefore a fitness function is proposed that is suitable in this regard. In fact, it's desirable to maximize the accident density among some of high crash road segments and minimize the accident density among the other segments. Therefore, if variance of accident density in the proposed model is higher, it will lead to maximizing accident density within high crash road segments and minimizing accident density within other segments. In addition, if it is supposed to improve the road safety while there is budget limitation, it would be desirable to improve the safety of the segment having the shorter length and higher number of accidents. Thus, the segments with maximum accident density among all segments are of higher priority to choose as the most important sections for safety improvement. So, it's desirable to maximize the variance and summation of accident density, so the minus sign was added to change the maximization problem to a minimization problem. The following equations presents the fitness function:

$$F = -\left(\sum_{i=1}^{N} AD_i + Var\,(AD)\right) \tag{1}$$

$$Var(AD) = \frac{\sum_{i=1}^{N}\left(AD_i - \overline{AD}\right)^2}{N} \tag{2}$$

$$\overline{AD} = \frac{\sum_{i=1}^{N} AD_i}{N} \tag{3}$$

Where F is the fitness function, AD_i is the accident density of ith segment, Var is variance of all the accident densities, $\overline{AD}$ is average of the accident density in all segments and N is the number of segments. In fact, F is a summation of all accidents density and variance of the accident density.

3.4 Genetic Algorithm

Genetic Algorithm (GA) is a population based optimization algorithm that is inspired from natural selection process. The initial solutions of GA which are called population, are produced randomly. Population is made of group of binary strings (named as chromosomes or individuals). The basis of GA is to survive the fittest population based on the ranking of the fitness function for all the populations. In each generation, some genetic operators lead the solution to have better fines function. The initial step in GA is to produce initial population randomly as feasible solutions. In the proposed GA, four stages are implemented until it reaches to the stop condition. The stop condition is defined as maximum number of iterations in the proposed method. In the first stage, the selection operator chooses the fittest individuals based on the fitness function so as to find the best chromosomes to apply crossover operator on them. In the second stage, crossover operator is applied to selected chromosomes from the previous section. Third stage is to run mutation operator over the whole population by chance. In the fourth stage, fitness function is evaluated for each individual and individuals are ranked based on their fitness function and the fittest individual is selected and kept for final decision regarding the best solution. Finally, after the stop condition is satisfied, the best solution is chosen among the solutions of each iteration based on the fitness function of individuals. The flowchart of proposed GA is shown in Figure 2. Table 1 shows the types and values of Genetic Algorithm's parameters. The population size and maximum number of iteration are selected equal to 160 and 200 respectively. (Goldberg, 1989).

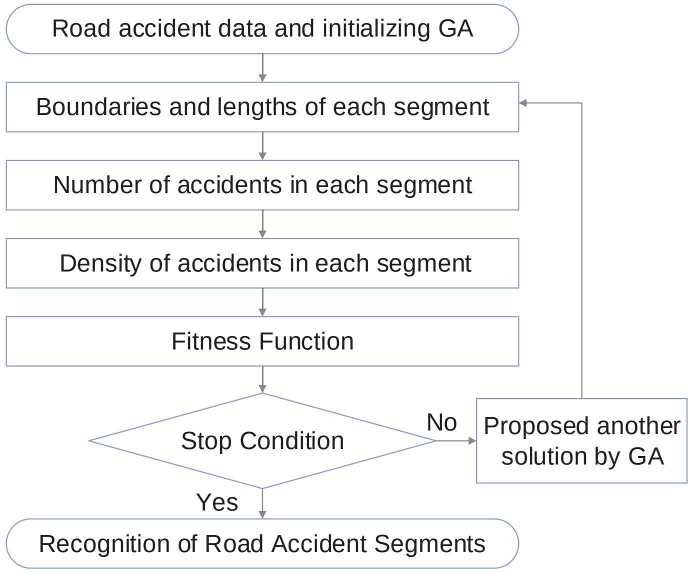

Figure 1. Flowchart of proposed Algorithm

3.4.1 Genetic Operators

Solutions are changed by these genetic operators in every iteration: selection, crossover and mutation operators. Selection operator makes chromosomes with better fitness function appear with higher probability in the next generation. Crossover operator is done between two selected chromosomes (as parents) to exchange some bits of their binary strings. Crossover points can be chosen randomly in several points of the parents' strings. Mutation operator is performed on problem solutions to lead to a global optimal solution of GA. This approach is done by changing the zeros to ones and ones to zeros at random locations on the populations. The GA is finished if the algorithm reaches to its maximum iteration as it's defined in the first step of the GA (Goldberg, 1989; Engelbrecht, 2007; Huang et al, 2009, and Samanta, 2004).

3.4.1.1 Selection Operator

Roulette wheel selection method is used to choose the fittest individuals within population for the next iterations. In roulette wheel selection method, the individuals that have better fitness function should have a greater chance to survive for the next iterations. The method assumes that the probability of selection is proportional to the fitness of individuals. The fitness function of each individual will be normalized to create the probability for choosing the right individuals. A corresponding section will be dedicated in a wheel based on the probability of individuals. Selection of an individual is to choose a point randomly in the wheel and locating the corresponding section for a special individual.

3.4.1.2 Crossover Operator

Crossover is the procedure of combining the chromosomes and passing the new chromosomes to the next iteration. The single point crossover method is used with the crossover probability of 0.90. Single point crossover method is an approach that uses only one breaking point in order to combine two random chromosome from a random breaking point. Crossover probability shows the fraction of the populations on which the crossover operator is implemented in each iteration.

3.4.1.3 Mutation Operator

Mutation operator changes some of the genes (binary bits) in a chromosome. Bit string mutation method is utilized with the mutation probability of 0.2. Bit string mutation method is an approach that converts the zeros to ones and ones to zeros in genes of populations. Mutation probability shows the fraction of the populations that mutation operator is implemented on some random genes in some random populations in each iteration.

3.5 Performance Indexes

Some indexes are introduced to compare the results between the proposed method and previous researches. These indexes are as follows:

1- Number of segments: This parameter shows the number of segments to have a sense to the number of high and low crash road segments.

2- Maximum density of all segments: It shows maximum accident density in all segments in order to identify the segment with highest accidents

3- Total segments density: It presents a summation of all accidents density so as to compare this number to other methods as a critical index.

4- Average of segments density: This parameter combines two previous indexes (the number of segments and total segments density) to make a new index to show the average accident density of each segment.

5- Variance of segments density: It illustrates the variance of the segments density. It is supposed to maximize this index in order to distinguish the segment based on the accident density.

Table 1. Types and values of Genetic Algorithm's parameters

Genetic Algorithm	Operators			Population Size	Maximum Iteration	Initial population
	Crossover Operator	Selection Operator	Mutation operator			
Parameters	Single Point Crossover Crossover probability = 0.90	Roulette wheel Selection	Bit string mutation Mutation probability = 0.2	160	200	Random

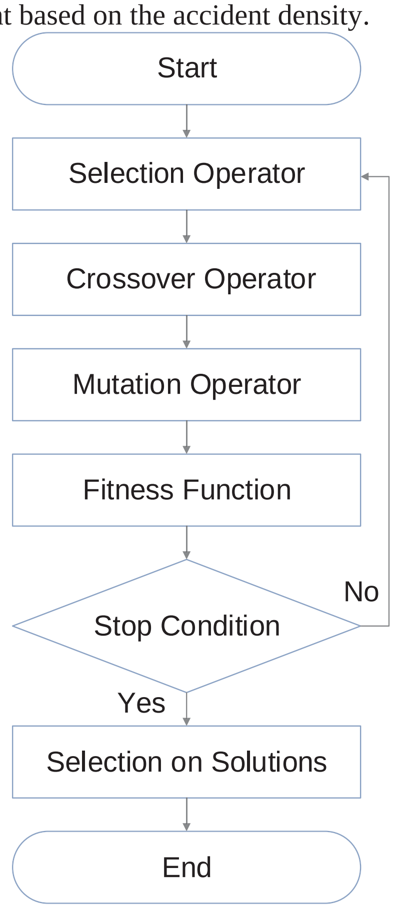

Figure 2. Flowchart of GA

4. Results and Simulations

In this part, the results of the proposed method and the comparisons of performance indexes with other methods have been shown. The first example of the paper was shown in Figure 3 so as to test the capabilities of the proposed method to display the distribution of accidents along the road. Figure 4, Figure 5, and Figure 6 show other examples in order to verify the suggested method. Figure 3 shows the method that can identify the segments with zero accident. Figure 4 depicts the method which can recognize the segments with the same number of accidents. Figure 5 illustrates that the method can identify the peaks and lows in the accident segments. Figure 6 demonstrates that the method can compound the segments which are almost equal in number of accidents. The manual analysis of these examples can show the general evaluation in capabilities of the proposed method to identify the correct boundaries of the segments. The vertical axis displays the number of accidents versus the location of accidents and the horizontal axis shows the number of samples (length of sampling unit is 100 m). The blue points show the position of accidents along the road and the red lines show the boundaries of the segments by the proposed method. As it is concluded from Figure 3, Figure 4, Figure 5 and Figure 6, the proposed method can identify the high and low crash segments corresponding to distribution of the accidents along the road, so these examples can verify the reliability and capability of the proposed method because these examples have various and comprehensive distributions of accidents along the road. The proposed method is implemented on the real road (Shahrood–Sabzevar highway) so as to find the segments with variable lengths that have most and least accidents. Figure 7 shows the data are extracted from about 51 km of Shahrood–Sabzevar highway in Iran. In this figure, the number of accidents versus the locations has been shown within about 51 kilometers. It's obvious that the boundaries of the segments can't be extracted by observation correctly. The values of the performance indexes are estimated for the road of interest in order to find the global optimal boundaries of the segment.

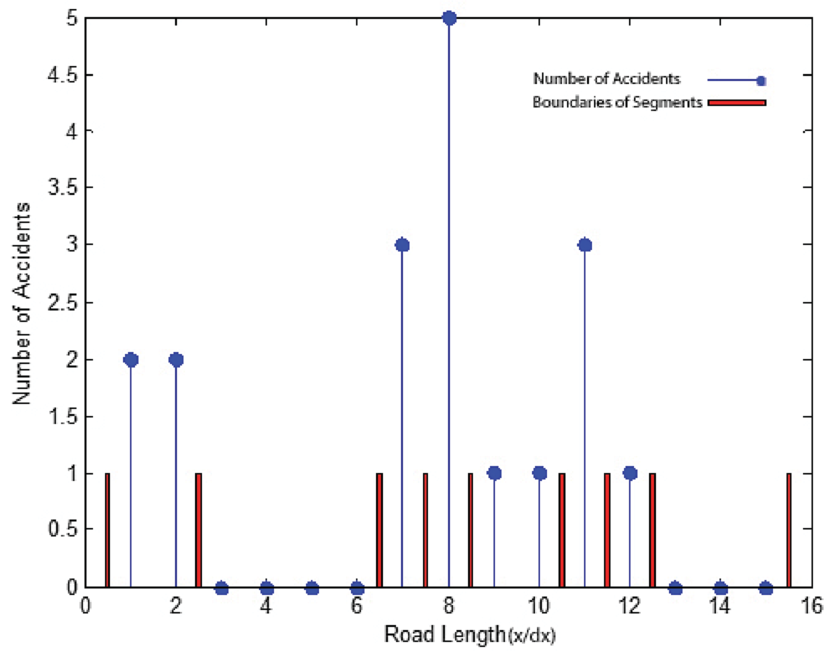

Figure 3. Example 1 for validation of proposed algorithm

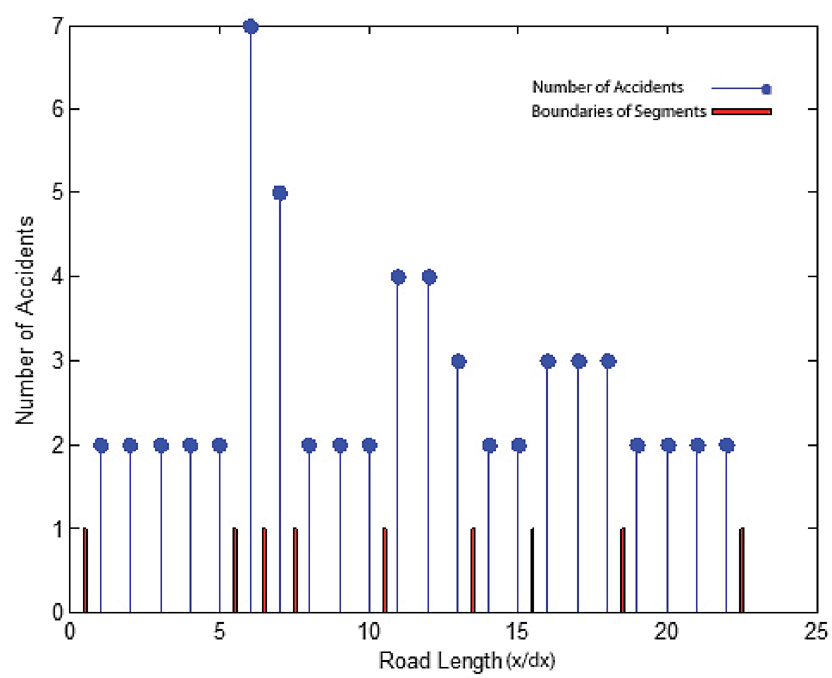

Figure 4. Example 2 for validation of proposed algorithm

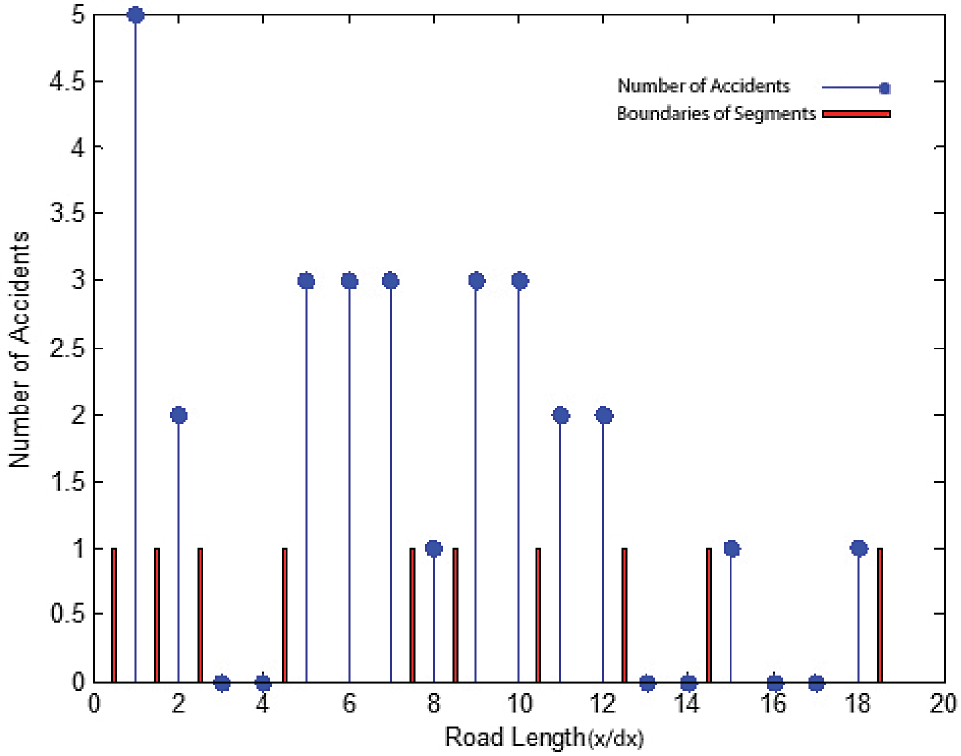

Figure 5. Example 3 for validation of proposed algorithm

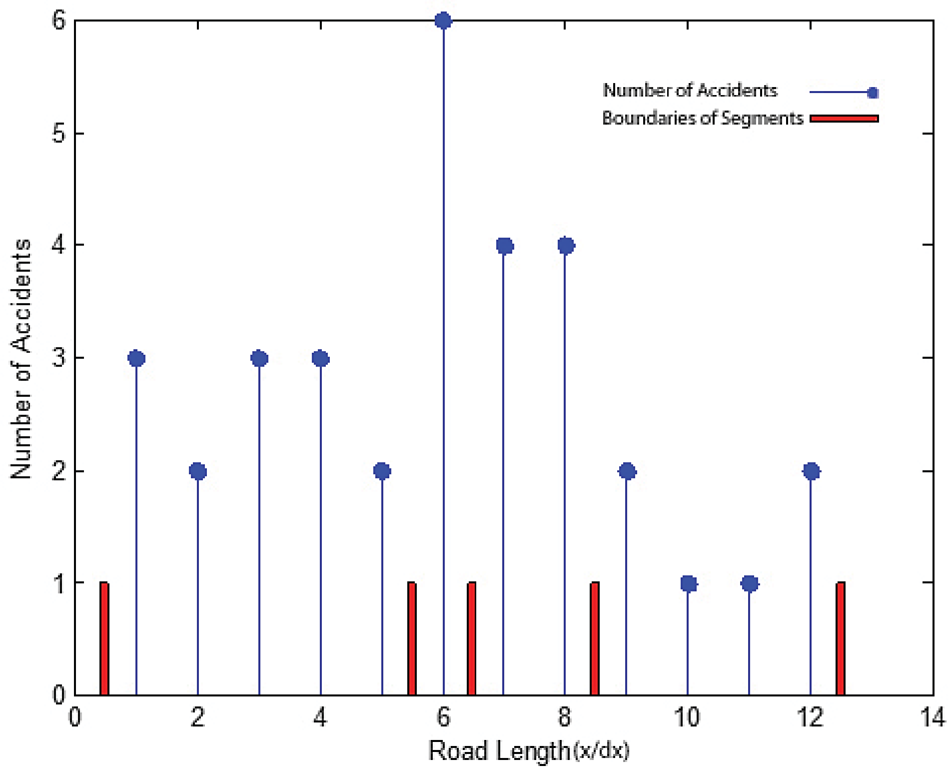

Figure 6. Example 4 for validation of proposed algorithm

Table 2. Number of accidents in diverse segmentation methods for each segment [Boroujerdian, 2011]

<table>
<tr><th rowspan="3">Km</th><th rowspan="3">Number of Accidents per Kilometer</th><th colspan="5">Number of Accidents of Each Segment</th></tr>
<tr><th colspan="4">Fixed Segmentations</th><th>Dynamic Segmentation</th></tr>
<tr><th>2 Km Segment</th><th>3 Km Segment</th><th>5 Km Segment</th><th>2 Km segments (Floating)</th><th>Proposed Algorithm using GA</th></tr>
<tr><td>1</td><td>0</td><td rowspan="2">0</td><td rowspan="3">0</td><td rowspan="5">46</td><td>0</td><td rowspan="3">0</td></tr>
<tr><td>2</td><td>0</td><td rowspan="2">0</td></tr>
<tr><td>3</td><td>0</td><td rowspan="2">25</td></tr>
<tr><td>4</td><td>25</td><td rowspan="3">66</td><td rowspan="2">46</td><td>25</td></tr>
<tr><td>5</td><td>21</td><td rowspan="2">41</td><td>21</td></tr>
<tr><td>6</td><td>20</td><td rowspan="5">96</td><td>20</td><td>20</td></tr>
<tr><td>7</td><td>14</td><td rowspan="2">49</td><td rowspan="3">56</td><td rowspan="2">49</td><td rowspan="2">49</td></tr>
<tr><td>8</td><td>35</td></tr>
<tr><td>9</td><td>7</td><td rowspan="2">27</td><td>7</td><td>7</td></tr>
<tr><td>10</td><td>20</td><td rowspan="3">40</td><td rowspan="2">36</td><td rowspan="2">36</td></tr>
<tr><td>11</td><td>16</td><td rowspan="2">20</td><td rowspan="5">62</td></tr>
<tr><td>12</td><td>4</td><td rowspan="2">28</td><td>4</td></tr>
<tr><td>13</td><td>24</td><td rowspan="2">28</td><td rowspan="3">42</td><td>24</td></tr>
<tr><td>14</td><td>4</td><td>4</td><td>4</td></tr>
<tr><td>15</td><td>14</td><td rowspan="2">22</td><td rowspan="2">22</td><td rowspan="3">28</td></tr>
<tr><td>16</td><td>8</td><td rowspan="3">61</td><td rowspan="5">76</td></tr>
<tr><td>17</td><td>6</td><td rowspan="2">53</td><td rowspan="2">53</td></tr>
<tr><td>18</td><td>47</td><td>47</td></tr>
<tr><td>19</td><td>1</td><td rowspan="2">15</td><td rowspan="3">50</td><td>1</td><td>1</td></tr>
<tr><td>20</td><td>14</td><td rowspan="2">49</td><td rowspan="2">49</td></tr>
<tr><td>21</td><td>35</td><td rowspan="2">42</td><td rowspan="5">113</td></tr>
<tr><td>22</td><td>7</td><td rowspan="3">57</td><td>7</td><td>7</td></tr>
<tr><td>23</td><td>41</td><td rowspan="2">50</td><td rowspan="2">50</td><td>41</td></tr>
<tr><td>24</td><td>9</td><td>9</td></tr>
<tr><td>25</td><td>21</td><td rowspan="2">33</td><td rowspan="3">39</td><td rowspan="2">33</td><td rowspan="3">39</td></tr>
<tr><td>26</td><td>12</td><td rowspan="5">61</td></tr>
<tr><td>27</td><td>6</td><td rowspan="2">40</td><td rowspan="2">40</td></tr>
<tr><td>28</td><td>34</td><td rowspan="3">43</td><td>34</td></tr>
<tr><td>29</td><td>5</td><td rowspan="2">9</td><td rowspan="2">9</td><td rowspan="5">14</td></tr>
<tr><td>30</td><td>4</td></tr>
<tr><td>31</td><td>1</td><td rowspan="2">2</td><td rowspan="3">5</td><td rowspan="5">6</td><td>1</td></tr>
<tr><td>32</td><td>1</td><td rowspan="2">4</td></tr>
<tr><td>33</td><td>3</td><td rowspan="2">3</td></tr>
<tr><td>34</td><td>0</td><td rowspan="3">1</td><td rowspan="2">1</td><td rowspan="4">1</td></tr>
<tr><td>35</td><td>1</td><td rowspan="2">1</td></tr>
<tr><td>36</td><td>0</td><td rowspan="5">0</td><td rowspan="2">.</td></tr>
<tr><td>37</td><td>0</td><td rowspan="2">0</td><td rowspan="3">0</td></tr>
<tr><td>38</td><td>0</td><td rowspan="2">.</td><td rowspan="4">1</td></tr>
<tr><td>39</td><td>0</td><td rowspan="2">0</td></tr>
<tr><td>40</td><td>0</td><td rowspan="3">1</td><td rowspan="2">1</td></tr>
<tr><td>41</td><td>1</td><td rowspan="2">1</td><td rowspan="5">1</td></tr>
<tr><td>42</td><td>0</td><td rowspan="2">.</td><td rowspan="7">1</td></tr>
<tr><td>43</td><td>0</td><td rowspan="2">0</td><td rowspan="3">0</td></tr>
<tr><td>44</td><td>0</td><td rowspan="2">.</td></tr>
<tr><td>45</td><td>0</td><td rowspan="2">0</td></tr>
<tr><td>46</td><td>0</td><td rowspan="3">1</td><td rowspan="5">1</td><td rowspan="2">.</td></tr>
<tr><td>47</td><td>0</td><td rowspan="2">1</td></tr>
<tr><td>48</td><td>1</td><td rowspan="2">1</td></tr>
<tr><td>49</td><td>0</td><td rowspan="2">0</td><td rowspan="3">0</td><td rowspan="3">0</td></tr>
<tr><td>50</td><td>0</td><td rowspan="2">.</td></tr>
<tr><td>51</td><td>0</td><td></td><td></td></tr>
</table>

Table 3. Comparison of diverse segmentation methods in accident density of each segment [Boroujerdian, 2011]

<table>
<tr><th rowspan="3">Km</th><th colspan="5">Accident Density of Each Segment</th></tr>
<tr><th colspan="4">Fixed Segmentations</th><th>Dynamic Segmentation</th></tr>
<tr><th>2 Km Segment</th><th>3 Km Segment</th><th>5 Km Segment</th><th>2 Km Segment (Floating)</th><th>Proposed Algorithm using GA</th></tr>
<tr><td>1</td><td rowspan="2">0.0</td><td rowspan="3">0.0</td><td rowspan="5">9.2</td><td>0</td><td rowspan="3">0</td></tr>
<tr><td>2</td><td rowspan="2">0</td></tr>
<tr><td>3</td><td rowspan="2">12.5</td></tr>
<tr><td>4</td><td rowspan="3">22.0</td><td rowspan="2">23</td><td>25</td></tr>
<tr><td>5</td><td rowspan="2">20.5</td><td>21</td></tr>
<tr><td>6</td><td rowspan="5">19/2</td><td>20</td><td>20</td></tr>
<tr><td>7</td><td rowspan="2">24.5</td><td rowspan="3">18.7</td><td rowspan="2">24.5</td><td rowspan="2">24.50</td></tr>
<tr><td>8</td></tr>
<tr><td>9</td><td rowspan="2">13.5</td><td>7</td><td>7</td></tr>
<tr><td>10</td><td rowspan="3">13.3</td><td rowspan="2">18</td><td rowspan="2">18</td></tr>
<tr><td>11</td><td rowspan="2">10.0</td><td rowspan="5">12.4</td></tr>
<tr><td>12</td><td rowspan="2">14</td><td>4</td></tr>
<tr><td>13</td><td rowspan="2">14.0</td><td rowspan="3">14.0</td><td>24</td></tr>
<tr><td>14</td><td>4</td><td>4</td></tr>
<tr><td>15</td><td rowspan="2">11.0</td><td rowspan="2">11</td><td rowspan="3">9.33</td></tr>
<tr><td>16</td><td rowspan="3">20.3</td><td rowspan="5">15.2</td></tr>
<tr><td>17</td><td rowspan="2">26.5</td><td rowspan="2">26.5</td></tr>
<tr><td>18</td><td>47</td></tr>
<tr><td>19</td><td rowspan="2">7.5</td><td rowspan="3">16.7</td><td>1</td><td>1</td></tr>
<tr><td>20</td><td rowspan="2">24.5</td><td rowspan="2">24.5</td></tr>
<tr><td>21</td><td rowspan="2">21.0</td><td rowspan="5">22.6</td></tr>
<tr><td>22</td><td rowspan="3">19.0</td><td>7</td><td>7</td></tr>
<tr><td>23</td><td rowspan="2">25.0</td><td rowspan="2">25</td><td>41</td></tr>
<tr><td>24</td><td>9.33</td></tr>
<tr><td>25</td><td rowspan="2">16.5</td><td rowspan="3">13.0</td><td rowspan="2">16.5</td><td rowspan="3">13</td></tr>
<tr><td>26</td><td rowspan="5">12.2</td></tr>
<tr><td>27</td><td rowspan="2">20.0</td><td rowspan="2">20</td></tr>
<tr><td>28</td><td rowspan="3">14.3</td><td>34</td></tr>
<tr><td>29</td><td rowspan="2">4.5</td><td rowspan="2">4.5</td><td rowspan="5">2.80</td></tr>
<tr><td>30</td></tr>
<tr><td>31</td><td rowspan="2">1.0</td><td rowspan="3">1.7</td><td rowspan="5">1.2</td><td>1</td></tr>
<tr><td>32</td><td rowspan="2">2</td></tr>
<tr><td>33</td><td rowspan="2">1.5</td></tr>
<tr><td>34</td><td rowspan="3">0.3</td><td rowspan="2">0.5</td><td rowspan="4">0.25</td></tr>
<tr><td>35</td><td rowspan="2">0.5</td></tr>
<tr><td>36</td><td rowspan="5">0.0</td><td rowspan="2">0</td></tr>
<tr><td>37</td><td rowspan="2">0.0</td><td rowspan="3">0.0</td></tr>
<tr><td>38</td><td rowspan="2">0</td><td rowspan="4">0.25</td></tr>
<tr><td>39</td><td rowspan="2">0.0</td></tr>
<tr><td>40</td><td rowspan="3">0.3</td><td rowspan="2">0.5</td></tr>
<tr><td>41</td><td rowspan="2">0.5</td><td rowspan="5">0.2</td></tr>
<tr><td>42</td><td rowspan="2">0</td><td rowspan="7">0.14</td></tr>
<tr><td>43</td><td rowspan="2">0.0</td><td rowspan="3">0.0</td></tr>
<tr><td>44</td><td rowspan="2">0</td></tr>
<tr><td>45</td><td rowspan="2">0.0</td></tr>
<tr><td>46</td><td rowspan="3">0.3</td><td rowspan="5">0.2</td><td rowspan="2">0</td></tr>
<tr><td>47</td><td rowspan="2">0.5</td></tr>
<tr><td>48</td><td rowspan="2">0.5</td></tr>
<tr><td>49</td><td rowspan="2">0.0</td><td rowspan="3">0.0</td><td rowspan="3">0</td></tr>
<tr><td>50</td><td rowspan="2">.</td></tr>
<tr><td>51</td><td></td><td></td></tr>
</table>

Table (2) shows the diverse segmentation methods for each segment. In this table, the number of accidents in static and dynamic segmentation is shown in 51 kilometers of the road. Static segmentations include fixed 2-km segmentation, fixed 3-km segmentation, fixed 5-km segmentation and 2-km (floating) segmentation. Table 3 depicts the comparison of all static and dynamic segmentation methods in accident density of each segment. Table 4 shows the comparison between the performance indexes of static and dynamic segmentation methods. It can be inferred from Table 4 that the proposed dynamic segmentation using the genetic algorithm has the highest maximum accident density, the highest total accident density for all segments, the highest variance of accident density for all segments, and the highest average of accident density for all segments in comparison with the other static segmentation methods. Two mathematical parameters were used to validate the proposed dynamic segmentation by GA. One of them is variance of accident density and the other is number of accidents in the segments with maximum accident density among all the segments. In fact, it’s desirable to maximize the accident density among some of the high crash road segments and minimize the accident density among the other segments. Therefore, if the variance of accident density in the proposed model is higher, it leads to maximization of accident density in high crash road segments and minimization of accident density in other segments. In addition, if it is to improve the safety of the roads and there is budget limitation, it’s desirable to improve the safety of shorter road segments with higher number of accidents. Thus, the segments having maximum accident density among all segments have higher priority to identify as the most important sections to be improved. As it can be seen in Tables 3 and 4, two mentioned parameters for validation of the proposed model is greatly better than the previous methods and it shows that the proposed segmentation method is working properly so that they can validate the modelling process. The results are compared between best outcome in fixed segmentation and the dynamic segmentation. The results show that GA has maximum density of 47, while other static methods have lower maximum density. The best maximum density for static methods belongs to 2-km (floating) segmentation with maximum density of 26.5. Although GA has total segments density of 336.78, 2-km (floating) segmentation has 250.5 in the mentioned subject. It is also figured out that variance of segments density in GA is 180.76 comparing to 94.4 for 2-km (floating) segmentation. The suggested method has an average of segment density of 14.64 in comparison with 3-km segmentation which is about 9.05. Table 4 implies that the variance value in GA is about two times higher than 2-km (floating) segmentation method. About 62% and 34% improvement is achieved in average of segments density and total segments density respectively in comparison with the other fixed methods. Figure 7 illustrated the data of 51 km of Shahrood–Sabzevar highway; the number of accidents versus the road length. Figure 8 shows dynamic segmentation of Shahrood–Sabzevar highway by the proposed method. The blue points show the number of accidents at the location of the accidents along the road and the red lines show the boundaries of the segments. It is concluded from Figure 8that the boundaries of the segments and the locations with highest or lowest accidents recognized properly. This result helps the government to develop new regulations for segmentation of the roads and perform the safety improvement for the sections before their situation or the others segments situations become worse so that the safety of all road segments remain acceptable.

5. Conclusion

Segmentation is a helpful method in order to locate the accidents occurring along the road through specific road segmentation method and identify high and low crash road segments. The problem was divided into two parts conceptually; modeling and optimization of the segments with variable lengths. The optimization problem has many variables and is also non-differential and discontinuous problem. Therefore, genetic algorithm was chosen to solve the optimization problem. The results illustrate that genetic algorithm can present an excellent solution for optimization of the problem well. One of the application of the approach is to identify high and low crash road segments with variable lengths and their priorities so as to the safety of special road sections are improved before entering to the worse situation. Furthermore, the budget can be optimized for improving the safety level of limited number of segments. The proposed method shows that it may identify the high crash road segments along the road more accurately in comparison with other static algorithms. The proposed algorithm has better result; about two times higher than the 2-km floating method in variance of accident density. Total segments density and average of segments density is considerably higher than the other static methods. For future work, the researchers suggest a new fitness function that can be derived from a linear or nonlinear function of statistic and physical characteristic of the roads. Since the most important part is to define the problem through math equations it's a challenging work to find high crash road segments with variable lengths. Furthermore, some new optimization algorithms can be used for evaluation and finding the best solution based on the new fitness function. If the fitness function has many local solutions, solutions can be placed in local solutions so that more reliable optimization algorithm is needed for this approach.

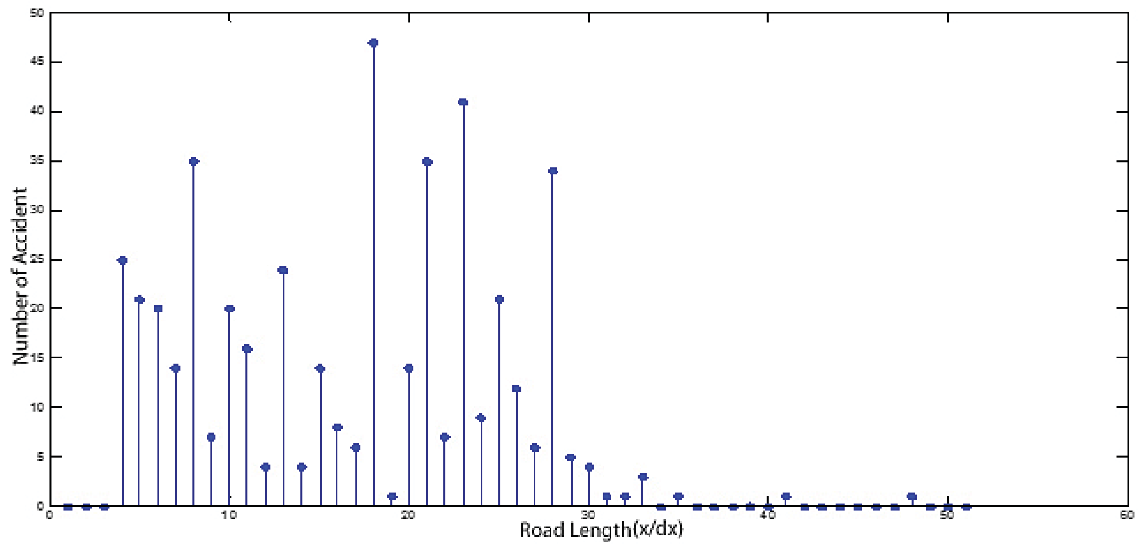

Figure 7. The Number of Accidents in Road Length for 51 km of Shahrood–Sabzevar Road [Boroujerdian et al. 2014]

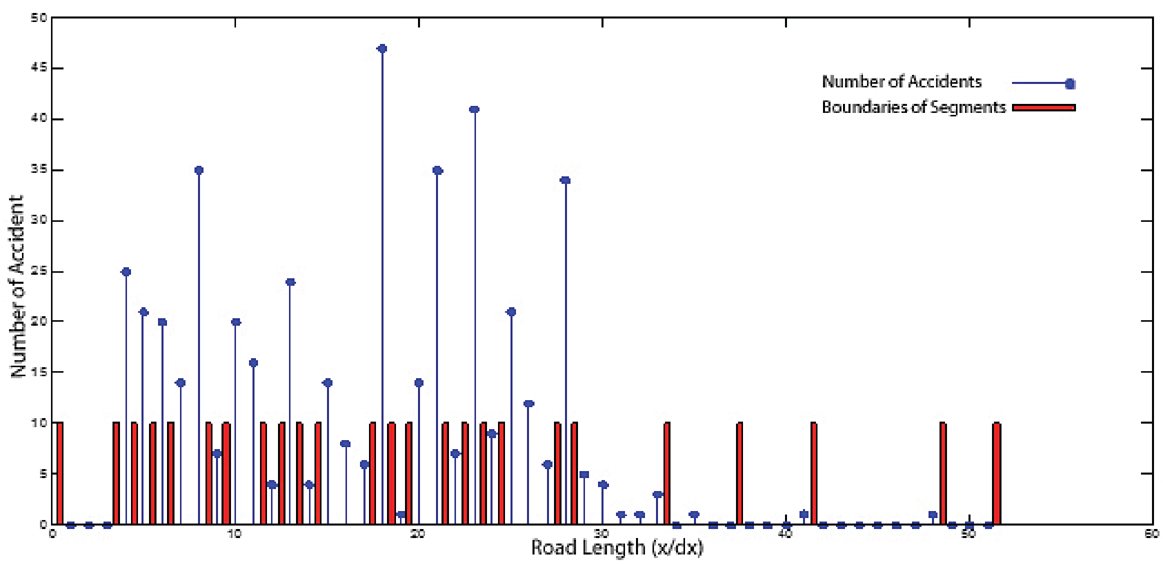

Figure 8. Dynamic Segmentation for 51 km of Shahrood–Sabzevar Road by proposed algorithm

Table 4. Performance indexes in different fixed and dynamic segmentation methods [Boroujerdian et al. 2014]

Performance Indexes	Static Segmentation				Dynamic Segmentation
	2 Km Segment	3 Km Segment	5 Km Segment	2 Km Segment (Floating)	Proposed Algorithm using GA
Number of Segments	26	17	11	30	24
Maximum Density of all segments	26.5	22.0	22.6	26.5	47
Total Segments Density	231	154	92.4	250.5	336.78
Variance of Segments Density	86.3	73.0	64.5	94.4	180.76
Average of Segments Density	8.85	9.05	8.40	8.36	14.64

6. Acknowledgement

This study is sponsored by Research Department at Birjand University of Technology, to which the authors are very grateful.

7. References

- Bonneson, J. and Zimmerman, K. (2006) "Procedure for using accident modification factors in the highway design process", Texas Transportation Institute, Report No. 0-4703-P5.

- Boroujerdian, A., Saffarzadeh M. and Abolhasannejad, V. (2009) "Developing a model for prioritizing high crash road segments", Proceedings the Institution of Civil engineers, Journal Transport, Vol. 163, pp. 19-28

- Boroujerdian, A. (2011) "Developing evaluation model of road safety based on the dynamic segmentation and caused based prioritization", Ph.D. Dissertation, Faculty of Civil and Environmental Engineering, Tarbiat Modares University.

- Boroujerdian, A. M., Saffarzadeh, M., Yousefi, H. and Ghassemian H. (2014) "A model to identify high crash road segments with the dynamic segmentation method", Accident Analysis and Prevention, Vol. 73, pp. 274–287.

- Boroujerdian, A. M., Fetanat, M and Abolhasannejad, V. (2015) "Dynamic segmentation method for identification of accident-prone road segments using genetic algorithm", Accepted 14th International Conference Traffic and On Transportation Engineering (ICTTE 2015).

- Elvik, R. (2008) "A survey of operational definitions of hazardous road locations in some European countries", Accident Analysis and Prevention, No. 40, pp. 1830–18357.

- Engelbrecht, A. P. (2007) "Computational intelligence: an introduction", 2nd edition.

- Federal Highway Administration (1981) "Highway safety improvement program", FHWA-TS-81-218, US Department of Transportation, Washington, DC, December.

- Geurts, K. (2006) "Ranking and profiling dangerous accident locations using data mining and statistical techniques", Doctoral Dissertation, Faculty of Applied Economics, Hasselt University, Hasselt.

- Goldberg, D. E. (1989) "Genetic algorithms in search, optimization, and machine learning reading", MA: Addison-Wesley.

- Huang, C. J., Chen, Y. J., Wu, C. F., and Huang, Y. A. (2009) "Application of neural networks and genetic algorithms to the screening for high quality chips", Applied Soft Computing, Vol. 9, No, 2, pp. 824–832.

- Kononov J. and Allery B. (2003) "Level of service of safety conceptual blueprint and analytical framework", Transportation Research Record 1840, TRB, Washington, D.C, pp. 03-2112.

- Pant, P. D., Rajagopal, A. S. and Cheng, Y. (2003) "Rational Schedule of Base Accident Rates for Rural Highways in Ohio (Phase II)", Report No. FHWA-OH-2003-008.

Saffarzadeh, M., Abolhasannejad V and Boroujerdian A. (2007) "Developing a cause prioritizing model to identify high accident prone segments of roads for pedestrians", Journal of The Collage of Engineering, Vol. 42, Issue 4, pp. 457-467

- Samanta, B. (2004) "Artificial neural networks and genetic algorithms for gear fault detection", Mechanical Systems and Signal Processing, Vol. 18, No. 5, pp. 1273–1282.

- The Research Institute of Transportation (2008) "The accidents records methods and identifying high crash road segments", The Ministry of Road and Transportation, Tehran, Iran.

- Troche, L. R. (2007) "Methodology to identify hazardous locations for highways in Puerto Rico", Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science.

- Vistisen, D. (2002) "Models and methods for hot spot safetywork", PhD Dissertation, Department for Informatics and Mathematical Models, Technical University of Denmark, Lyngby.