FINDING PATTERN BEHAVIOR IN TEMPORAL DATA USING FUZZY CLUSTERING

RICHARD E. HASKELL
Oakland University

DARRIN M. HANNA
Oakland University

PING LI
Oakland University

KA C. CHEOK
Oakland University

GREG HUDAS
Oakland University

ABSTRACT

A clustering technique based on a fuzzy equivalence relation is used to characterize temporal data. Data collected during an initial time period are separated into clusters. These clusters are characterized by their centroids. Clusters formed during subsequent time periods are either merged with an existing cluster or added to the cluster list. The resulting list of cluster centroids, called a cluster group, characterizes the behavior of a particular set of temporal data. The degree to which new clusters formed in a subsequent time period are similar to the cluster group is characterized by a similarity measure, q. This technique has been applied to the problem of detecting driver behavior.

INTRODUCTION

Many different clustering techniques have been used for analyzing multivariate data [1]. These methods have been applied to problems in knowledge discovery and data mining [2]. The standard clustering algorithms assign each data sample to one of many clusters in which all samples in a particular cluster are similar in some sense. Fuzzy clustering algorithms do not insist that each sample must belong to only one cluster, but rather samples can belong to more than one cluster to varying degrees. The most well known fuzzy clustering algorithm is the fuzzy c-means algorithm [3] that requires that the number of cluster centers, c, be given. A different clustering approach that does not require the number of clusters to be known beforehand is based on the use of fuzzy equivalence relations [4, 5]. In this method a fuzzy compatibility relation matrix, Q, is formed in which each entry in the matrix represents the degree to which two different samples are close to each other. A value of 1 (on the main diagonal) represents the degree to which a sample is close to itself, while a value of 0 represents samples separated by the largest possible distance in the data set. The transitive closure of Q will induce crisp partitions of the data (resulting in different numbers of clusters) by choosing different α-cuts of a fuzzy set [5]. Clusters formed in this manner will be used in this paper to characterize pattern behavior in temporal data.

This research was motivated by the desire to characterize a driver's behavior by monitoring signals that are already being measured by the car's computer system. Many techniques have been suggested for monitoring the alertness of a driver [6]. Suppose that the computer in an automobile could tell who was driving the car from the way the car was being driven, or more generally, the type of driver (aggressive, cautious, inattentive, etc.). With this type of information the car might be able to adjust its control algorithm to optimize fuel consumption, minimize wear, or adjust an active suspension system to improve safety. Once recognizing a particular driving pattern, an unexpected deviation from this pattern might suggest a drowsy driver or some other type of abnormal behavior.

This paper will focus on the use of clusters formed from a fuzzy equivalence relation matrix in an attempt to characterize a driving pattern as a function of time. A different approach was taken by Peltier and Lajon [7]. They describe a fuzzy pattern recognition algorithm in which they use a neural net based architecture together with a real-time algorithm for detecting abrupt changes in a driving pattern.

When a driver starts to drive a car, data are collected for an initial amount of time, Δt_1, (say less than a minute). These data are clustered to form c_1 clusters. During the next interval of time, Δt_2, new data are collected and clustered to form c_2 clusters. If the driving activity during Δt_2 is similar to the driving activity during Δt_1 then the c_2 clusters (as specified by their centroids) will be similar to the c_1 clusters. On the other hand, if the driving activity has changed somewhat, some of the c_2 clusters will be different from the c_1 clusters. Clusters in c_2 that are close to clusters in c_1 will be merged with the corresponding clusters in c_1. The remaining clusters in c_2 will be added to the c_1 clusters to form a new, larger c_1 cluster set. This process is continued for each successive interval of time, Δt_n, with new clusters either being merged with an existing cluster or added to the cluster set. After a while, the resulting group of clusters, called a *cluster group*, will characterize a particular driver. This method will allow a large amount of data collected over a period of time to be clustered in a manner that uses a manageable amount of data at each clustering step. The same process can be carried out for a different driver. A different cluster group will characterize this second driver.

Once a cluster group is formed, new data that are collected during a subsequent interval of time, Δt_n, will form its own cluster set. In this paper we will introduce a similarity measure, q, that will measure the degree to which this new cluster set "fits in" with the cluster group. By using this similarity measure one can determine which of several cluster groups provides the best fit for a particular cluster set. If different cluster groups represent different drivers, then by voting over several Δt_n intervals it should be possible to determine which driver is driving the car. By the same token, once the cluster group has stabilized for a particular driver and a new cluster appears that is significantly different from what has been seen before, this may indicate some type of change such as a lack of alertness on the part of the driver.

CLUSTERING BASED ON A FUZZY EQUIVALENCE RELATION

This section will describe a clustering technique in which multivariate data are used to form a fuzzy equivalence relation matrix. Different α-cuts of this fuzzy set will produce a different number of clusters of the original data. Given a data set $\{(x_{11},\ldots,x_{1p}),\ldots,(x_{n1},\ldots,x_{np})\}$ with n samples over a p-dimensional feature space, P, a fuzzy compatibility relation matrix Q with dimension $n \times n$ is computed. Define M_k and m_k as the maximum and minimum data point x_{jk} for each feature k in P as

and

$$M_k = \max_{j \in \{1,..,n\}} x_{jk}$$

$$m_k = \min_{j \in \{1,..,n\}} x_{jk}$$

Define the i,j-th entry in Q as

$$q_{ij} = 1 - \frac{1}{p} \sum_{k=1}^{p} \left(\left| \frac{x_{ik} - x_{jk}}{M_k - m_k} \right|^s \right)^{1/s} \qquad (1)$$

to form the fuzzy compatibility relation matrix Q. Each q_{ij} represents the degree to which data point x_i is close to data point x_j. The distance measure in Eq. (1) will be the Hamming distance for $s = 1$ and the Euclidean distance for $s = 2$. In our experiments we will use a value of $s = 1$.

The matrix Q is symmetric and reflexive. However, the generalization of transitivity to fuzzy relations is not unique [5]. One common definition is to say that a fuzzy relation Q is transitive if and only if

$$Q(x,z) \geq \max_{y \in Y} \min\left[Q(x,y), Q(y,z)\right]$$

The right-hand side of this inequality represents the composition of relation Q with itself, $Q \circ Q$. The transitive closure, T, of Q can be computed by the following algorithm.

Transitive Closure:

do

$T' = Q;$

$Q = T' \circ T';$

$while(Q \neq T')$

$T = T'$

From the transitive closure matrix T with elements t_{ij}, a collection of clusters, C, is formed for a specific membership degree α. Set $C_k \in C$ such that

$$\forall i, j \in C_k, t_{ij} \geq \alpha$$

forms a fuzzy equivalence class. Define a fuzzy equivalence cluster W_i by $\{(x_{w1},\ldots,x_{wP}) \mid w \in C_i\}$.

Multiple Cluster Sets

Define the centroid, a_i, of cluster W_i by

$$a_i = \frac{1}{|C_i|} \sum_{j \in C_i} x_j$$

where $|C_i|$ denotes the cardinality of cluster set C_i. Let $A_\alpha = \{a_1, \ldots, a_m\}$ be the set of fuzzy cluster centroids resulting from data collected during a time interval, Δt_{k1} and $B_\alpha = \{b_1, \ldots, b_n\}$ be a second set of fuzzy cluster centroids resulting from data collected during a time interval, Δt_{k2} where $k2 > k1$ with accumulative maximums and minimums for each feature, max_{Ak}, max_{Bk}, min_{Ak}, and min_{Bk}, where, without loss of generality,

$$\max_{Ak} = \max_{i} \max_{j \in C_i} x_{jk}$$

and

$$\min_{Ak} = \min_{i} \min_{j \in C_i} x_{jk}$$

and weight vectors $W_{a\alpha} = \{w_{a1}, \ldots, w_{am}\}$ and $W_{b\alpha} = \{w_{b1}, \ldots, w_{bn}\}$ given, without loss of generality, by

$$W_{a\alpha} = \{w_{ai} \mid w_{ai} = |C_i|\}$$

Define r_k as the global range,

$$r_k = \max(\max_{Ak}, \max_{Bk}) - \min(\min_{Ak}, \min_{Bk})$$

Form the fuzzy relation matrix, Z, by

$$z_{ij} = 1 - \frac{1}{p} \sum_{k=1}^{p} \left(\left| \frac{(a_{jk} - b_{ik})}{r_k} \right|^s \right)^{1/s} \qquad (2)$$

where p is the dimension of the feature space and $s = 1$ for a Hamming distance or $s = 2$ for a Euclidean distance. Form the projections

$$\rho_i^A = \max_i z_{ij} \qquad (3)$$

and

$$\rho_i^B = \max_j z_{ij} \qquad (4)$$

Merging Cluster Sets

Using the centroid relation matrix, Z, and ρ_i^B, a new collection of clusters $\aleph_{t+1}$ is constructed with threshold, β. For all $\rho_i^B > \beta$, a_j is replaced as follows,

$$\left.\frac{w_{aj} a_j + w_{bi} b_i}{w_{aj} + w_{bi}}\right|_{z_{ij} = \rho_i^B} \qquad (5)$$

Accordingly, w_{aj} is updated with $w_{bi} + w_{aj}$.
Finally, form

$$\aleph_{t+1} = \left\{b_i \middle| \rho_i^B \le \beta\right\} \cup A_\alpha \qquad (6)$$

for a new collection of clusters representing time-series Δt_{k1} and Δt_{k2}. This method is repeated for each successive time interval, Δt.

Cluster Similarity Measure

Let A_α be the set of clusters formed by adding and merging cluster sets over a number of time intervals. Let B_α be the cluster set during a new time interval, Δt_n. A fuzzy relation matrix Z can be computed from Eq. (2) and the projection ρ_i^B is given by Eq. (4). The similarity measure q, is defined as the degree to which cluster set B_α is similar to cluster set A_α as

$$q = \frac{1}{n} \sum_{i=1}^{n} \rho_i^B \qquad (7)$$

Examples of using this similarity measure will be given in the following section.

EXPERIMENTAL RESULTS

To test the theory given in the previous section we conducted two experiments. The first experiment used the well-known Iris data as surrogate driving data in order to show the behavior of the algorithms in a simple situation. The second experiment used data collected from a driver simulator.

Iris Data

The well-known iris data [8, 9] contains 50 samples for each of three species of iris plants: *Iris setosa*, *Iris versicolor*, and *Iris virginica*. Each measurement contains four features representing sepal length, sepal width, petal length, and petal width. It is well known that the *setosa* class is linearly separable from the other two classes. The clusters associated with the *versicolor* and *virginica* classes overlap somewhat and are not linearly separable. While most clustering algorithms try to cluster these three classes into three clusters, our interest is to use sequences of this data to simulate different temporal data.

To simulate a possible driving situation we took the first 15 samples of *setosa* together with the first 15 samples of *versicolor* to represent driving during Δt_1. We would expect this to form two distinct clusters representing two different driving situations. The next 15 samples of *setosa* and *versicolor* represented the data in Δt_2 and the last 20 samples of *setosa* and *versicolor* represented the data in Δt_3. We call this the *setver* data set and it represents Driver 1. A similar set of data (called *setvir*) was made from the *setosa* and *virginica* data sets and represents Driver 2. Note that both data sets contain the same *setosa* data and therefore should have a common cluster. This might represent driving along a straight road at a specific speed. However, the clusters associated with *versicolor* and *virginica* are slightly different but can be distinguished using the similarity measure, q, given in Eq. (7).

The Δt_1 samples (set *S1*) from *setver* (Driver 1) were clustered using the clustering algorithm described above with a value of $\alpha = 0.83$. This produced the two clusters labeled *S1* in Table 1. The first cluster contains the *setosa* data and the second cluster contains the *versicolor* data. The Δt_2 and Δt_3 samples (sets 2 and 3) produced the two-cluster sets labeled *S2* and *S3* in Table 1. The two clusters in set 2 were then merged with the two clusters in set 1 using Eq. (6) with a value of $\beta = 0.8$ to produce the three-cluster group labeled *G12* in Table 1. Note that the two *setosa* clusters were merged while the two *versicolor* clusters remained unique. A similar set of clusters was formed from the *setvir* data set representing Driver 2.

The three cluster sets, *S1*-*S3*, from each of the two data sets, *setver* (Driver 1) and *setvir* (Driver 2), were compared with the *G12* clusters from both *setver* (Driver 1) and *setvir* (Driver 2). The resulting similarity values, q, from Eq. (7) are shown in Figure 1. One can think of sets 1 and 2 as the “training” sets for *setver* (Driver 1) and *setvir* (Driver 2), and the unknown set 3 as the test samples. In both cases the values of q for the same driver are larger than the values of q for the other driver, making the recognition possible.

Table 1 Clusters formed from the *setver* data set

Set/ Group	No. of Samples	Cluster Centroids				Class
S1	15	4.9	3.3	1.4	0.2	Setosa
	15	6.0	2.8	4.2	1.3	Versicolor
S2	15	5.1	3.6	1.5	0.3	Setosa
	15	6.2	2.8	4.4	1.4	Versicolor
S3	20	5.0	3.4	1.5	0.2	Setosa
	20	5.7	2.7	4.2	1.3	Versicolor
G12	15	6.0	2.8	4.2	1.3	Versicolor
	30	5.0	3.5	1.5	0.2	Setosa
	15	6.2	2.8	4.4	1.4	Versicolor

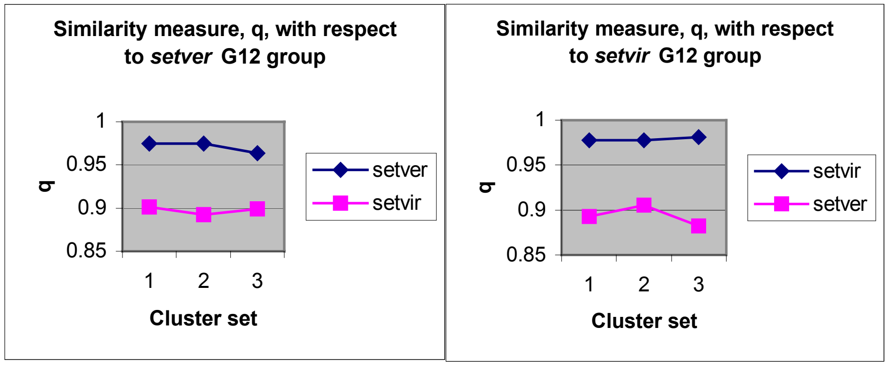

Figure 1 Similarity measurements with respect to the *setver* and *setvir G12* cluster groups

Driving Simulator Data

To test the feasibility of using fuzzy clustering to characterize driver behavior a series of experiments were conducted on a Virtual Vehicle System Simulator (VVSS) [10, 11] at Oakland University. The VVSS is a networked distributed modular real-time driving simulator that immerses a human operator in various driving scenarios. It provides visual, audio, and motion feedback to the operator as the virtual car is being driven. Validated dynamic models for the vehicle are employed for the simulation. Simulated data for the vehicle states can readily be collected from the "virtual driving test runs."

The shape of the driving track used in the experiments is shown in Figure 2. Three separate drivers drove around the track. Real-time measurements of speed, acceleration, and steering angle were collected at a rate of 5 samples per second for approximately two complete trips around the track. The data for each driver were divided into ten time segments with cluster sets formed for each time segment using a value of $\alpha = 0.9$. After removing clusters made from a single data point the number of clusters remaining in these data sets ranged from 1 to 4. The data from the first five time segments (representing approximately the first time around the track) were merged with a value of $\beta = 0.9$ to form a cluster group for each of the three drivers. The total number of clusters in these cluster groups was 6 for driver 2 and 9 for drivers 1 and 3.

The cluster sets 6-10 (representing approximately the second time around the track) were used to test the recognition capability of these cluster groups. These five cluster sets for each driver were compared with each of the three training cluster groups (from the first time around the track) using the similarity measure, q, given in Eq. (7). The resulting values of q ranged from 0.79 to 0.97. For each segment the three driver cluster sets were compared with the three driver cluster groups in an attempt to identify the driver. For example, in segment 7 the cluster set for driver 3 was compared with the cluster groups for drivers 1, 2, and 3. The value of q for the driver 1 group was 0.85, for the driver 2 group was 0.84, and for the driver 3 group was 0.97. Therefore, the predicted driver in this segment is driver 3, which is correct. The predicted drivers in segments 6-10 for each of the three drivers are shown in Table 2. By taking a vote over all five segments, each driver is correctly identified the second time around the track.

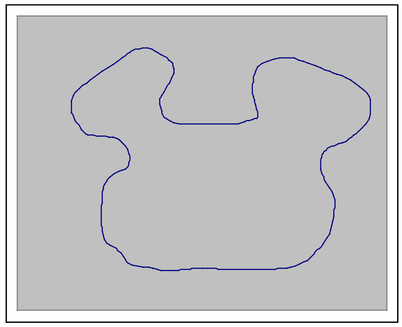

Figure 2 Shape of Driving Track

Table 2 Predicted Drivers for Segments 6-10

Segment	Driver 1	Driver 2	Driver 3
6	Driver 3	Driver 2	Driver 2
7	Driver 1	Driver 2	Driver 3
8	Driver 1	Driver 3	Driver 2
9	Driver 1	Driver 3	Driver 3
10	Driver 3	Driver 2	Driver 3
Predicted Driver	Driver 1	Driver 2	Driver 3

CONCLUSIONS

In this paper we have shown how clustering based on a fuzzy equivalence relation can be used to find pattern behavior in temporal data. Its possible application to detecting driver behavior was discussed. One advantage of the method is that large amounts of data that are arriving in real time can be absorbed by clustering small segments of data and merging the clusters into cluster groups that characterize a particular time sequence. A similarity measure, q, was introduced that measures how closely a new cluster set matches (fits into) a larger cluster group. This can be used to determine if the new cluster set belongs to a particular cluster group (e.g. is this the same driver?). The individual components of the fuzzy projection used to calculate q could be used to indicate when a new cluster that is different from existing clusters appears unexpectedly. In the case of driving data this might indicate a change in driver behavior related to fatigue, for example.

Future research should collect a wider range of real-time data in an actual car. This data should include signals that are likely to be different for different driving patterns. In addition to detecting driver behavior, the behavior of the car itself can be monitored using this method. Such an intelligent vehicle might be able to warn the driver of needed maintenance before the car breaks down on the highway.

REFERENCES

1. B. S. Everitt, *Cluster Analysis*, 3rd Ed., Edward Arnold, London, 1993.
2. D. Wishart, "Efficient hierarchical cluster analysis for data mining and knowledge discovery," Computer Science and Statistics, 30, pp. 257-263, 1998.
3. J. D. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.
4. G. J. Klir, U. St.Clair and B. Yuan, *Fuzzy Set Theory – Foundations and Applications*, Prentice Hall PTR, Upper Saddle River, NJ, 1997.
5. G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic – Theory and Applications*, Prentice Hall PTR, Upper Saddle River, NJ, 1995.
6. S. Sarkar, I. K. Sethi and G. Bochenick, “A Survey of Alertness Monitoring Methods for Drivers,” in *Intelligent Engineering Through Artificial Neural Networks*, Vol. 9, Eds: C. H. Dagli, et. al., pp. 845-849, ASME Press, New York, 1999.
7. M. Peltier and M. Lajon, “A Fuzzy Diagnosis Process for the Detection of Evolution of a Car Driver’s Behavior,” Proceedings of the Intelligent Vehicles Symposium, pp. 19-24, Paris, France, October 1994.
8. E. Anderson, "The Irises of the Gaspe Peninsula, Bull. Amer. Iris Soc., Vol. 59, pp. 2-5, 1935.
9. R. A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems, Ann. Eugenics, Vol. 7, pp. 179-188, 1936.
10. G. E. Smid, Ka C. Cheok & T. K. Tan, "Multi-CPU Real-time Simulation of Vehicle Systems," Proc. 7th Int'l Conf. Intelligent Systems (ICIS '98), Melun Fointain Fontainebleau, July 1-2, 1998.
11. G. E. Smid and Ka C. Cheok, "Human Integration in Driving Simulator," Proc. IEEE Advanced Motion Conrol (AMS) '98, Coimbra, Portugal, June 29-July 1, 1998, pp. 554-558.