A System of Intelligent Route Prediction and Security for Passenger Cars based on Fuzzy Clustering

Gowtham Raj G

Robert Bosch Engineering and Business Solutions Private Limited

CHIL-SEZ, Keeranatham Village, Coimbatore 641 035, India

ABSTRACT

Cars are being upgraded continuously to pave the path for intelligent transportation. This paper utilizes data mining for automatic route prediction in a passenger car through unsupervised machine learning. The use of Fuzzy Clustering Algorithm makes this concept probabilistic. It groups different routes based on the passengers in the car and provides suggestion for the driver during next use. This concept additionally provides a valuable security system that ensures owner of the car safe belonging of his/her car.

General Terms

Pattern Recognition, Fuzzy Clustering, Security, Data Mining.

Keywords

Fuzzifier, Fuzzy Clustering, Principal Component Analysis, Eigenfaces.

1. INTRODUCTION

Machine learning algorithms can figure out how to perform important tasks by generalizing from examples. In machine learning, the problem of unsupervised learning is that of trying to find hidden structure in unlabeled data [1]. The purpose of this paper is to utilize the concept of Fuzzy clustering; a data mining algorithm to cluster together the different routes travelled using the particular vehicle and provides intelligent suggestion during the future rides. This enables intelligent control of a car [2] with future being assured of driverless cars.

The whole concept besides the paper can be categorized into two sections:

1. Image recognition by the use of a camera system and a controller.

2. Process of Fuzzy Clustering algorithm to identify the routes as data and associate with a label (for particular user). Then use of learned values to suggest the route in future use.

For processing, a dataset containing a family of four members is considered (Mother, Father, Daughter, and Son).The faces of each of them is initially stored in under the tags Mother, Father, Daughter and Son in the microcontroller. The controller keeps track of usual routes travelled by each individual along with time details and make continuous unsupervised learning using Fuzzy Clustering to form clusters of data and create labels for different clusters such as Mother, Father, Daughter, and Son. The Clustering is further enhanced to identify combination of passengers such as Mother and Father or Mother and son, etc. In cases such as Father and Mother, the route with more cluster

fuzziness to the clusters Father and Mother and least or zero with clusters daughter and Son is considered. The controller for machine learning is centrally connected to the vehicle ECU. It is necessary for centralized control and co-ordination among the different ECUs in a vehicle [3].

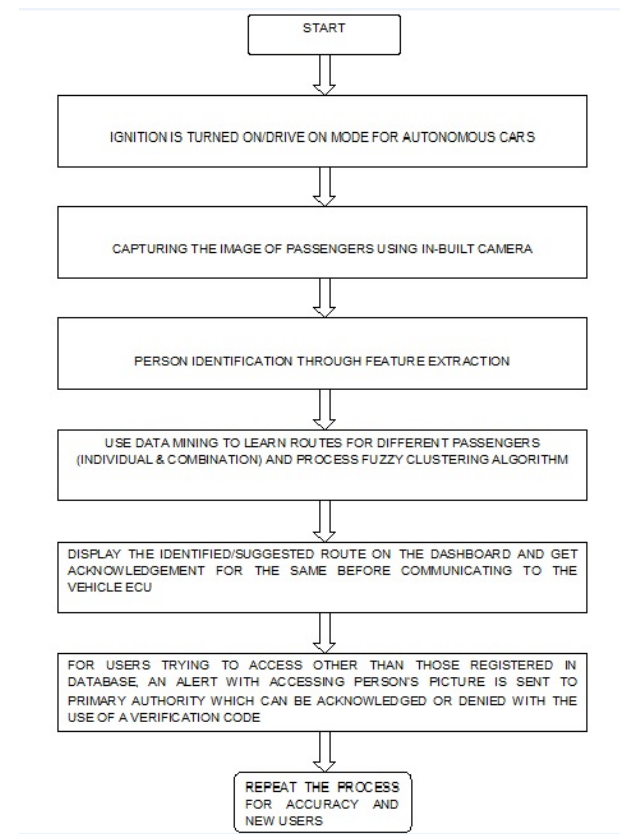The following flowchart provides the sequence of actions necessary to implement the data mining concept.

Fig.1 Flowchart of the complete Process

2. IMAGE RECOGNITION

Face recognition is a set of two tasks:

Face Identification: Given a face image that belongs to a person in a database, tell whose image it is.

Face Verification: Given a face image that might not belong to the database, verify whether it is from the person it is claimed to be in the database.

For Feature extraction, the classical algorithm of PCA is utilized.

2.1 The Process of PCA:

1. A training set consisting of total M image is considered.
2. Normalize the face vectors
   a. Calculate the average face vectors
   b. Subtract average face vector from each face vector
3. Calculate the Eigenvectors from reduced covariance matrix.
4. Select K best Eigenfaces such that K<+M and can represent the whole training set.
5. Convert lower dimension K eigenvectors to original face dimensionality.
6. Represent each face image a linear combination of all K eigenvectors.

The goal of PCA from [4, 5] is to reduce the dimensionality of the data while retaining as much information as possible in the original dataset and generate Eigen vectors. Though there are many methods developed for face recognition, the algorithm of eigenfaces remain to be the most dependent and accurate algorithm [6].

To generate a set of eigenfaces, a large set of digitized images of human faces, taken under the same lighting conditions is required, in this case the images of users of vehicle. They are then all re-sampled at the same pixel resolution. Eigenfaces can be extracted out of the image data by means of principal component analysis (PCA).Eigen faces uses statistical approach to recognize unknown faces using a set of training face which we also refer to as the database of the members who use the car .The eigenvectors are derived from the covariance matrix of the probability distribution over the high-dimensional vector space of face images.

Eigenfaces can be considered a set of "standardized face ingredients", derived from statistical analysis of many pictures of faces. Any human face can be considered to be a combination of these standard faces. Each face image is matrix of pixels which is viewed as a vector of length equal to the number of elements in the matrix [7].

The idea is to build a face space which better describes the face and the basis vectors of this face space are known as principal components.
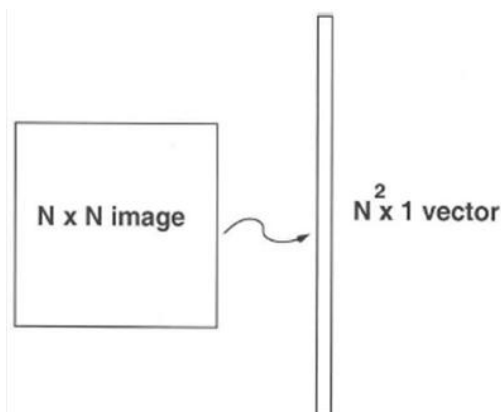

Fig.2 Image Space

The Eigen-vectors can be thought of as a set of features which characterize the variation between face images. The number of Eigen-vectors is equal to the number of training face images in the database. The face images in the training set are then re-constructed by a weighted sum of a small set of characteristic images. All the training face images are projected onto the face space. The test face images are also projected onto the face space and the distances between the test image and all the training face images are calculated. The Euclidean distance is used as the basis for calculating the distance between the face images. The training face image for which the Euclidean distance is the minimum is recognized as the unknown face. On complete recognition of travelers, the process of data mining begins.

3. FUZZY CLUSTERING PROCESS FOR ROUTE PREDICTION

Data mining also called knowledge discovery is the process of analyzing data from different perspectives and summarizing it into useful information and cluster analysis being the main task of exploring data mining [8]. Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters).

The goal of clustering is to

   1. Group data points that are close (or similar) to each other

   2. Identify such groupings (or clusters) in an unsupervised manner

        a. Unsupervised: no information is provided to the algorithm on which data points belong to which clusters.

If data objects are distributed in well-separated groups, then a crisp classification of the objects into disjoint clusters seems like an ideal approach. However, in most cases, the objects in a data set cannot be partitioned into well-separated clusters, and there will be certain arbitrariness in assigning an object to a particular cluster.

Consider an object that lies near the boundary of two clusters, but is slightly closer to one of them. In many such cases, it might be more appropriate to assign a weight to each object and each cluster that indicates the degree to which the object belongs to the cluster. Fuzzy clustering techniques are based on fuzzy set theory [9] and provide a technique for providing a clustering in which membership weights have a natural interpretation.

3.1 Algorithm of Fuzzy C-means Clustering:

1. Initialize $U = [u_{ij}]$ matrix, $U^{(0)}$

2. At k-step: calculate the centers vectors $C^{(k)} = [c_j]$ with $U^{(k)}$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m \times \chi_i}{\sum_{i=1}^{N} u_{ij}^m} \qquad (1)$$

3. Update $U^{(k)}$, $U^{(k+1)}$

$$\frac{1}{\sum_{k=1}^{C} (\frac{|\chi_i - c_j|}{|\chi_i - c_k|})^{\frac{2}{m-1}}} \qquad (2)$$

4. If $\| U(k+1) - U(k) \| < \varepsilon$ then STOP; otherwise return to step 2.

Here, m is the fuzzifier constant; c is the centroid of the clusters and χ is the node considered for calculation (route).

m is simply chosen as 2 since it simplifies the weight update formula [10].

For the concept of route prediction and calculation simplification, following simplified test cases are considered as base cases:

- Mother, Father, Daughter and Son travel to and from work and study places respectively on a weekday (all four in car).
- Father uses car on two more occasions in a day.
- On weekends, all four travel to holiday spots and individual uses for unique purposes on different occasions of a day.

Here routes are considered as numerical values for compiling using statistical software. With initially four clusters being labeled from the results centered around four members of the family. Successive calculations based on unsupervised learning improve the results for individual clusters [11] and clusters formed when two or more members travel at the same time.

Following is the fuzzy clustering report generated with the statistical software:

3.2 Fuzzy Clustering Report:

Dataset   Test Data
Variables   A, B
Distance Type   Euclidean
Scale Type   Standard Deviation
Fuzzifier Constant   2

TABLE I
Membership Summary Section for Clusters = 4

Row (A)	Cluster (B)	Cluster Membership
4	1	0.8225
3	1	0.7910
6	2	0.8999
7	2	0.5233
5	3	0.7204
1	3	0.5099
8	3	0.4621
2	3	0.4277
12	4	0.9987
11	4	0.9987
10	4	0.9987
9	4	0.9987

TABLE II
Membership Matrix Section

Row	Cluster	Prob in 1	Prob in 2	Prob in 3	Prob in 4
1	3	0.2496	0.1525	0.5099	0.0881
2	3	0.2049	0.2518	0.4277	0.1156
3	1	0.7910	0.0484	0.1205	0.0401
4	1	0.8225	0.0368	0.1124	0.0283
5	3	0.1501	0.0717	0.7204	0.0578
6	2	0.0238	0.8999	0.0409	0.0355
7	2	0.0930	0.5233	0.1434	0.2403
8	3	0.2004	0.1595	0.4621	0.1779
9	4	0.0003	0.0006	0.0004	0.9987
10	4	0.0003	0.0006	0.0004	0.9987
11	4	0.0003	0.0006	0.0004	0.9987
12	4	0.0003	0.0006	0.0004	0.9987

In Table 1, Cluster B represents person at different time and Row A represents various routes travelled. Here the person and routes are mentioned numerically for statistical manipulation.

For cluster B, 1-Father, 2-Mother, 3-Daughter, and 4-Son.

The provision for numerical values to people is not included in the calculation.

In Table 2, for different person their membership or probability in other clusters is calculated.

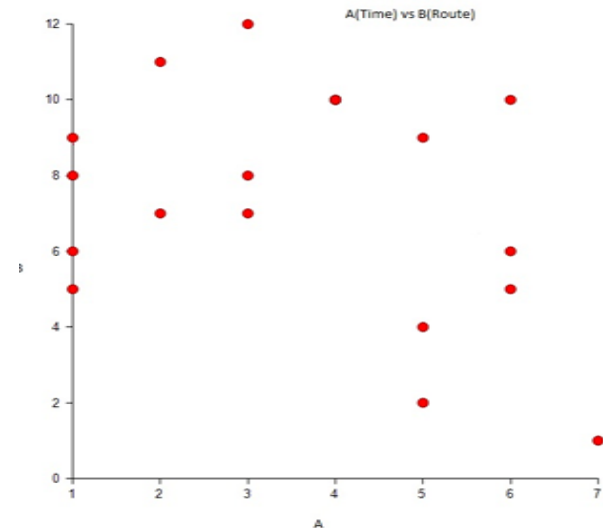The following figure shows the initial clustering graphically,

Fig.3 Initial Cluster Plot

The cluster plot shows the route travelled initially. Each point is a route travelled by a person at different time. Clustering is done on these data points. Corresponding person(s) is/are matched at all times. This is updated on further calculations and learning. The cluster is labeled with appropriate person name and matched with corresponding picture in the database on all occasions. The process is coded appropriately in microcontroller and it is in continuous communication with vehicle ECU. During successive travels, the clusters' centroid is updated automatically and new clusters with improved efficiency are formed [12]. When a new person travels along with family members, the process prompts for a new addition which can either be accepted or declined.

The security pattern is not associated with fuzzy process but still is programmed along with fuzzy logic. The communication with vehicle ECU enables locking of ignition system from enabling on attempt to access by an unauthorized person. This feature can also be made as an add-on feature based on customer requirements.

4. FUTURE WORK

The concept has been developed to aid passenger cars. It is now been extended to autonomous cars. Future Work is channeled to enable this process available for commercial vehicles and mass transportation systems through improved process efficiency. A point is also made to improve the computation time of Fuzzy algorithm which considerably lags behind K-means method [13].

5. CONCLUSION

Use of Fuzzy Clustering technique to establish data mining in automobile dashboard gives a valuable addition to the automation of automotive features. It is one of the key concepts which provide user a sense of luxury and security. The reliable and robust nature of Fuzzy algorithm has made this concept possible in the wake of intelligent transportation system becoming the future.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] Murphy, Kevin P. *Machine learning: a probabilistic perspective*. MIT press, 2012.

[2] Sugeno, M. A., and M. Nishida. "Fuzzy control of model car." *Fuzzy sets and systems* 16.2 (1985): 103-113.

[3] Dietsche, K., Maria Klingebiel. *Bosch Automotive Handbook*, 2007.

[4] Vohra, Siddharth, Sagar Khattar, and Arvind Kumar. "Microcontroller based Automatic Face Recognition Attendance System." *vol* 4: 32-37.

[5] Wold, Svante, Kim Esbensen, and Paul Geladi. "Principal component analysis." *Chemometrics and intelligent laboratory systems* 2.1 (1987): 37-52.

[6] Yang, Jian, et al. "Two-dimensional PCA: a new approach to appearance-based face representation and recognition."*Pattern Analysis and Machine Intelligence, IEEE Transactions on* 26.1 (2004): 131-137.

[7] Gonzalez, Rafael C. *Digital image processing*. Pearson Education India, 2009.

[8] Tan, Pang-Ning, Michael Steinbach, and Vipin Kumar. *Introduction to data mining*. Vol. 1. Boston: Pearson Addison Wesley, 2006.

[9] Bezdek, James C., Robert Ehrlich, and William Full. "FCM: The fuzzy c-means clustering algorithm." *Computers & Geosciences* 10.2 (1984): 191-203.

[10] Mohri, Mehryar, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of machine learning*. MIT press, 2012.

[11] Quigley, C. P., et al. "Predicting journey parameters for the intelligent control of a hybrid electric vehicle." *Intelligent Control, 1996., Proceedings of the 1996 IEEE International Symposium on*. IEEE, 1996.

[12] Morris, Brendan Tran, and Mohan M. Trivedi. "Learning, modeling, and classification of vehicle track patterns from live video." *Intelligent Transportation Systems, IEEE Transactions on* 9.3 (2008): 425-437.

[13] D'silva, Mitchell, and Deepali Vora. "Comparative Study of Data Mining Techniques to Enhance Intrusion Detection." *International Journal of Engineering Research and Applications (IJERA) ISSN* (2013): 2248-9622.