Journal of Contemporary Issues in Business and Government Vol. 27, No.5,2021
https://cibgp.com/

P-ISSN: 2204-1990; E-ISSN: 1323-6903

MARKET CLASSIFICATIONS FROM 2020 TO 2030

[1]Dr. Sumedha Tuteja, [2]Mr. Kenate Beyane , [3]Dr. Rashmi Jain, [4]Dr. G. Radha Krishna Murthy , [5]Dr. Abdul Razak

[1]Associate Professor Indira Institute of Management Pune,
[2]Head, Department of Management, COBE Wollega
University Nekemte Ethiopia
[3]Associate Professor (Marketing)
Chetana's Institute of Management and Research Mumbai
[4]Professor Department of Management, COBE
Wollega University Nekemte Ethiopia
[5]Assistant Professor
School of Business
SR University, Warangal

Abstract

This study introduces a novel pattern recognition method. Hidden Markov Models are used to build the system (HMMs). We modify the Hidden Markov Models Toolkit (HTK) to address the pattern recognition problem in this paper. HTK was created for the purpose of voice recognition research. The mean of feature vectors is used to define patterns at first. By adding headers and displaying them in consecutive frames, such feature vectors are transformed to HTK format. A Windowing function is applied to each one. HTK then uses feature vectors for training and recognition tests.For the trials, we used 1600 randomly generated patterns from sixteen different consumer groups. The obtained findings demonstrate the efficacy of the suggested method.

Keywords

Pattern classification; Hidden Markov Models; Hidden Markov Models Toolkit

I. INTRODUCTION

A subset of machine learning is pattern recognition. It is the process of identifying things using qualities such as form, colour, temperature, and so on. Automatic techniques are commonly used in the identifying procedure. The authors of [1] [2] [3] [4] offer an overview of pattern recognition, including its concept and connections to other fields. Regression, labelling, and parsing are all classic pattern recognition instances. In the first, each input is given a real-valued output [5]. In the second, each set of input values is allocated to a certain class, similar to how voice tagging is done [6]. Parsing is the process of examining a string of symbols and using a parse tree to describe its syntactic structure [7]. This article is about pattern categorization. For supervised learning [11], neural networks [8, support vector machines [9], and linear discriminant analysis-based classifiers [10], as well as k-means [12] and kernel principal component analysis classifiers [13] for unsupervised learning [14], several probabilistic classifiers have been suggested. In this article, we offer an off-line classification method based on an HMM. The Hidden Markov Models Toolkit [15] was created at the Cambridge University Engineering Department's Voice and Robotics division and has mostly been utilized in speech recognition research. In two phases, the system may be broken down. In the first, feature vectors characterizing the previous 100 events are used. Purchases are made for each consumer. Each feature vector is regarded as a signature of the database's associated pattern (customer). The second phase is carried out using HTK, with Baum-Welch re-estimation [15] being used to find optimal HMM parameters. The Viterbi algorithm [15] is then used to decode the data. Section 3 will go through this in greater depth. The following is a breakdown of the paper's structure: The second section deals with data representation. Section 3 describes the proposed system. In section 4, the outcomes of the experiments are presented. Finally, in part 5, we bring this work to a close.

P-ISSN: 2204-1990; E-ISSN: 1323-6903

II. DATA REPRESENTATION

We have sixteen different types of clients in our project. A list of 100 purchases is used to describe each one. The first 25 decide whether or not he is a vegetarian customer. If the number of vegetarian purchases surpasses the number of meat-based purchases, the consumer is deemed vegetarian. The next 25 purchases will decide whether or not he buys organic items. If he buys fresh veggies or not depends on the next 25 purchases. Finally, his recent 25 purchases decide whether or not he buys ethnic items.

• A: vegetarian consumer • B: carnivorous customer • C: organic items • D: conventional products • E: fresh veggies • F: frozen vegetables • G: ethnic products

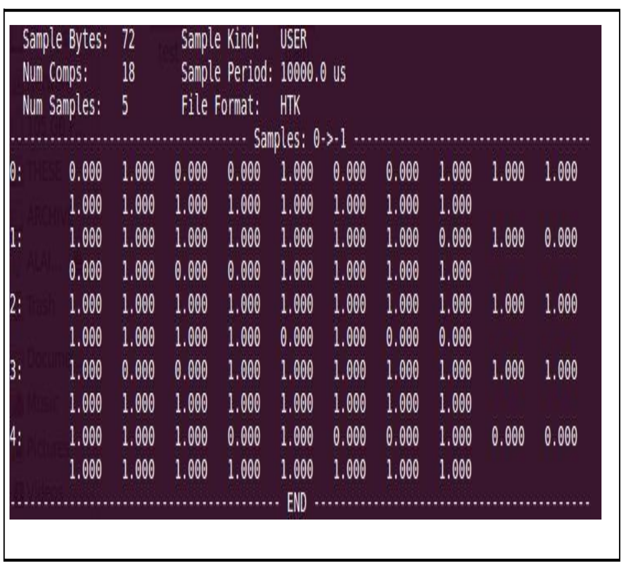
• H: purchases non-ethnic goods

For example, a consumer from Class 3 is a vegetarian. He shops for organics, frozen veggies, and ethnic foods.

III. THE MARKET CUSTOMERS CLASSIFIER PROPOSED ON THE BASE OF HTK

A. Converting data to HTK format

HTK was created for the purpose of voice recognition research. In this paper, we suggest adapting it to create a pattern categorization system. The following is a description of the proposed system: Each database client has a signature, which is a 100-item vector that represents his most recent 100 purchases. Those signatures are first saved as text. They are represented as a sequence of samples preceded by a 12-byte header in HTK format. The header contains a 2-byte integer indicating the number of bytes per sample and a 2-byte integer indicating the sample type, as well as a 4-byte integer indicating the sample time in 100 ns units and another 4-byte integer indicating the number of samples in the linked HTK file. After then, the signature is converted into a series of frames. Each frame is finally multiplied by the Hamming function to provide a compact representation of its spectral properties (a windowing function). Figure 2 shows the header source information of an HTK file pertaining to a client sample.



```
Sample Bytes:  72     Sample Kind:   USER
Num Comps:     18     Sample Period: 10000.0 us
Num Samples:   5      File Format:   HTK
------------------------------ Samples: 0->-1 ------------------------------
0:  0.000  1.000  0.000  0.000  1.000  0.000  0.000  1.000  1.000  1.000
    1.000  1.000  1.000  1.000  1.000  1.000  1.000  1.000
1:  1.000  1.000  1.000  1.000  1.000  1.000  1.000  0.000  1.000  0.000
    0.000  1.000  0.000  0.000  1.000  1.000  1.000  1.000
2:  1.000  1.000  1.000  1.000  1.000  1.000  1.000  1.000  1.000  1.000
    1.000  1.000  1.000  1.000  0.000  1.000  0.000  0.000
3:  1.000  0.000  0.000  1.000  1.000  1.000  1.000  1.000  1.000  1.000
    1.000  1.000  1.000  1.000  1.000  1.000  1.000  1.000
4:  1.000  1.000  1.000  0.000  1.000  0.000  0.000  1.000  0.000  0.000
    1.000  1.000  1.000  1.000  1.000  1.000  1.000  1.000
----------------------------------- END ------------------------------------
```

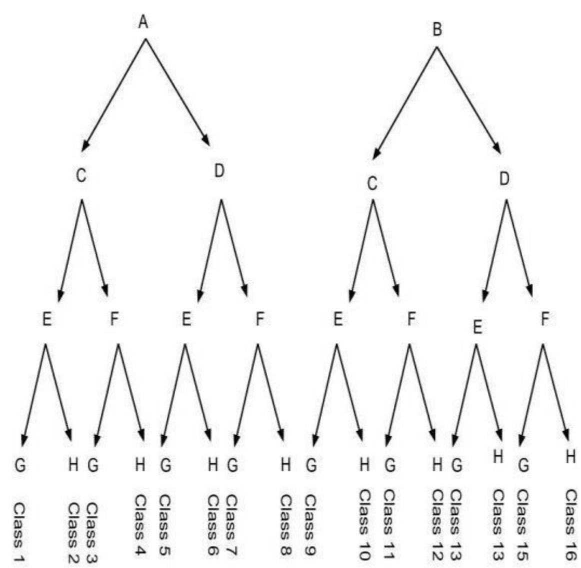


Figure 1 & 2. AHTK customer sample file

B. System Set-up

For each client segment, the first step is to construct an HMM prototype, which includes establishing the number of states, the observation function, and the transition matrix between states. Figure 3 depicts an HMM prototype sample.

P-ISSN: 2204-1990; E-ISSN: 1323-6903

```
<StreamInfo> 1 100
<VecSize> 10 <nullD> <USER> <diagC>
~h "customer_458"
<BeginHMM>

    <NumStates> 18
    <State> 2
    <Stream> 1
    <Mixture> 1 1.0

        <Mean> 100
        0.0    0.0    0.0    0.0    0.0    ...    0.0
        <Variance> 100
         1.0    1.0    1.0    1.0    1.0    ...    1.0

    <State> 3
        .
        .
        .
    <State> 17
    <Stream> 1
    <Mixture> 1 1.0

        <Mean> 100
        0.0    0.0    0.0    0.0    0.0    ...    0.0
        <Variance> 100
         1.0    1.0    1.0    1.0    1.0    ...    1.0

      <TransP> 18
      0.000e+0  1.000e+0  0.000e+0  0.000e+0  0.000e+0 .....  0.000e+0
      0.300e-0  0.400e-0  0.300e-0  0.000e+0  0.000e+0        0.000e+0
      0.000e+0  0.400e-0  0.300e-0  0.400e-0  0.000e+0        0.000e+0
         :         :         :         :         :               :
         :         :         :         :         :               :
      0.000e+0  0.000e+0  0.000e+0  0.000e+0  0.000e+0        0.000e+0

<EndHMM>
```

Figure 3. AHMM Customer Prototype sample

The state's observation function I is defined as state> I, which is a single Gaussian observation description.The variance vector of the observation function is provided by variance>i, whereas the mean vector is provided by mean>i. They should both be the same size as the observation space. The transition matrix, or Transp, determines the likelihood of transitioning from one condition to another. We will assume that there are 18 states in this work (Figure 4). State 1 and state 18 are non-emitting states, whereas each state represents a consumer class. Staying in the same state has a probability of 0.4, whereas going to the next or previous state has a probability of 0.3.

The language is then specified in terms of certain precise rules before being compiled to produce the system network.

```
$phn = customer1 |customer2 |customer3 | ... |customer16 ; (<$phn>)
```

Figure 4.The system grammar

After that, the dictionary is defined. It's a simple text file that illustrates how grammatical variables and HMM prototypes are related. Finally, HMM prototypes, dictionaries, and a system network make up the customers' recognizer system (Figure 5).

Figure5. The proposed customers' recognizer system

C. Training Phase

The HMM prototype parameters (transition matrix, mean, and variance vectors) are put up initially, with mean vector elements set to 0.0 and variance vector elements set to 1.0. In addition, we set the values of the transition matrix. During the training process, these factors vary (Figure 6). The Baum-Welch method is then used to re-estimate them.

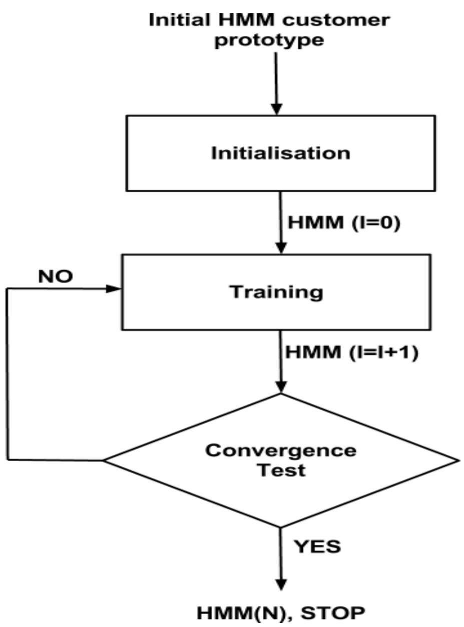


Figure 6. The training phase

D. Recognition phase

First, a signature is produced and converted to HTK format for a query customer. The Viterbi method is then used to decode it. The network, dictionary, and HMM prototypes must all be well defined. Figure 8 illustrates the whole recognition method.

It's worth mentioning that the HMM prototypes used in the recognition phase were created after n iterations of parameter re-estimation during the training stage in order to achieve system convergence.

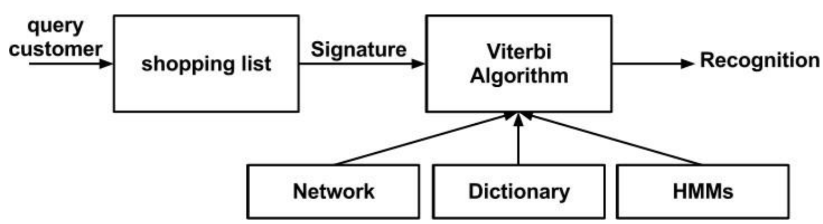


Figure 7. The Customers Recognition Phase

IV. EXPERIMENTS

The number of states is determined based on a heuristic. Table I illustrates the recognition rates obtained while utilizing various system topologies to help determine the optimal value for this parameter (different number of states). The best recognition rate is obtained when the number of states equals seven, according to the data.

**TABLE I.
RECOGNITIONRATEFORDIFFERENTCONFIGURATIONS
(DIFFERENTNUMBEROFSTATES)**

Number of States	4	5	6	7	8
Recognition Rate	89.07	69.39	87.33	93.13	-

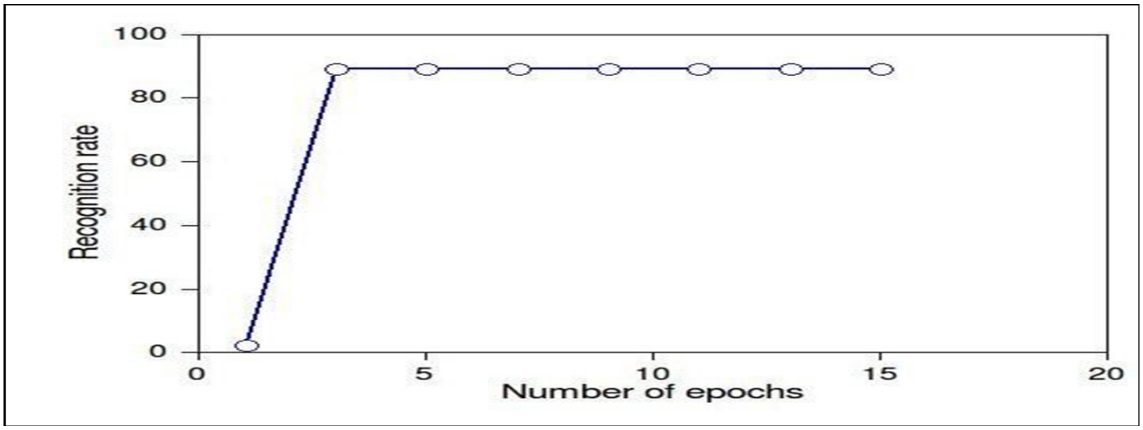


Figure 8. Recognition at E versus the Number of Training Epocs

To determine whether or not the system converges rapidly. Figure 8 depicts recognition rates at various epochs, demonstrating that the system converges near the end of the third epoch.

We also varied the size of the testing and training corpus to show the efficacy of the suggested technique. The recognition rates for various setups are shown in Table II.

**TABLE II.
RATE OF RECOGNITION IN VARIOUS SETTINGS (DIFFERENT SIZES OF TRAINING AND TEST CORPUS)**

Training	Test	Recognition rate
1280(80%)	320(20%)	93.12%
960 (60 %)	640(40%)	93.59%
800 (50 %)	800(50%)	92%

The above Table summarises that the recognition rate reaches 93.12 percent when 80 percent of the corpus is used for training and 20% for recognition testing. It stays high even when half of the data is used as a training corpus and half as a test corpus (92 percent).

V. CONCLUSION

The suggested method allows market managers to assess a query customer's class, which helps with the selection of advertising and bargains delivered to the customer's location. HTK is a very strong technique that may be utilized when dealing with pattern recognition difficulties, according to results obtained utilizing a list of 1600 randomly generated consumers belonging to sixteen different classifications. The overall recognition rate is 93.59 percent.

Injecting actual data into the system will enhance this article. It can also be enhanced by using online recognition software.

REFERENCES

[1] W.G.Wee, "A survey of pattern recognition, “Adaptive Processes, Volume:7, pp.25,1968.

[2] W.HachichaandA.Ghorbel,"A survey of control-chart pattern-recognition literature (1991 2010) based on a new conceptual classification scheme," Computers & Industrial Engineering, Volume 63, Issue1,pp. 204-222 ,August 2012.

[3] C.J.D.M.Verhagen,"Some general remarks about pattern recognition; its definition; its relation with other disciplines; a literature survey," Pattern Recognition, Volume 7, Issue 3, pp. 109-116, September 1975.

[4] J. Mantas, "Methodologies in pattern recognition and image analysis—A brief survey, "Pattern Recognition, Volume 20, Issue1, pp.1-6,1987.

[5] D.S.ThomasandA.Mitiche,"Asymptotic optimality of pattern recognition by regression analysis Original Research Article," Neural Networks, Volume7,Issue 2, pp.313-320,1994.

[6] A. Ekbal and S. Saha, "Simulated annealing based classifier ensemble techniques: Application to part of speech tagging," Information Fusion, In Press, Corrected Proof, July 2012.

[7] Th.W.M. Vissers, J. Chwilla, H.J. Kolk, "The interplay of heuristics and parsing routines in sentence comprehension: Evidence from ERPs and reaction times," Biological Psychology, Volume 75, Issue 1, pp. 8-18, April 2007.

[8] J. Misraand I.Saha, "Artificial neural networks in hardware: A survey of two decades of progress," Neuro-computing, Volume 74, Issues 1–3, pp.239-255, December 2010.

[9] M.Y. Cheng and F.V. Roy, "Evolutionary fuzzy decision model for cash flow prediction using time- dependent support vector machines, "International Journal of Project Management, Volume 29, Issue 1, pp.56- 65, January2011.

[10] K.F. Lam and J.W. Moy, "A piecewise linear programming approach to the two-group discriminant problem – an adaptation to Fisher’s linear discriminant function model, “European Journal of Operational Research, Volume 145, Issue2, pp.471-481, March2003.

[11] Y. MAO, Y.Q. ZHOU, R.F. LI, X.J. WANG and Y.X. ZHONG, "Semi-supervised learning via manifold regularization," The Journal of China Universities of Posts and Telecommunications, Volume 19, Issue 6, pp.79-88,December2012.

[12] J.Triantafilis,I.O.A.Odeh,B.MinasnyandA.B.McBratney,"Elucidation of physiographic and hydrogeological features of the lower Namoi valley using fuzzy k-means classification of EM34data,"Environmental Modelling & Software, Volume 18, Issue 7, pp. 667-680,September2003.

[13] A.E. Mercer and M.B. Richman, "Assessing Atmospheric Variability using Kernel Principal Component Analysis, "Procedia Computer Science,Volume12,pp.288-293,2012.

[14] A.Maurer, "Unsupervised slow subspace-learning from stationary processes Original Research Article," Theoretical Computer Science, Volume 405, Issue3, pp.237-255, October 2008.

[15] S. Young, HTK Book, 2006.