Hybrid of Supervised and Unsupervised Learning Algorithms for Human Activity Recognition

Dedabrishvili Mariam, Rodonaia Irakli
International Black Sea University

Abstract

Classification process is involved in a variety of applications. This process is carried out by distinctive algorithms from supervised and unsupervised learning environment. This paper describes the selected representatives of both approaches, namely Artificial Neural Networks (ANN) and Hidden Markov Models (HMM) respectively. Algorithms are applied to the dataset derived from body mounted sensors to differentiate between different daily activities of humans. Moreover, the study incorporates the hybrid of these algorithms with the aim to improve the classification accuracy using ensembles of classifiers. Obtained results show that the combination of the algorithms outperforms the classification accuracy of separately used ANN and HMM when they are applied as classifier models. The achievement is essential while developing real-time applications in general and in particular, it is vital in elderly population's lives while dealing with Human Activity Recognition.

Keywords: Algorithm Combination**,** ANN, Classifier Accuracy, HMM, Supervised Learning, Unsupervised Learning

1. Introduction

In supervised learning, based on labeled data the algorithms are trained with predefined concepts and functions [1]. Supervised learning methods try to discover the relationship between input attributes (i.e. independent variables) and a target attribute (i.e. dependent variable). The obtained relationship represents the structure which is devoted to as a model. Discovered models generally define and explain phenomena, which are out of sight in the dataset and can be useful for prediction of the value of the dependent variable while the independent variable values are known [2].

The process of learning a set of rules from instances (examples in a training set) is known to as inductive machine learning, or in other words, building a classifier that can be utilized to generalize from new instances [3]. Figure 1 shows the process of supervised Machine Learning application to a real-world problem [4].

In unsupervised learning, algorithms have to find out interesting properties of the given a set of instances [5]. Unsupervised learning task is to find out how systems can learn to show particular input patterns in a way that replicates the statistical structure of the whole set of input patterns. Compared to Supervised Learning or Reinforcement Learning, there are no obvious target outputs nor environmental evaluations related with each input; rather the unsupervised learner holds prior biases as to what characteristics of the structure of the input ought to be captured in the output.

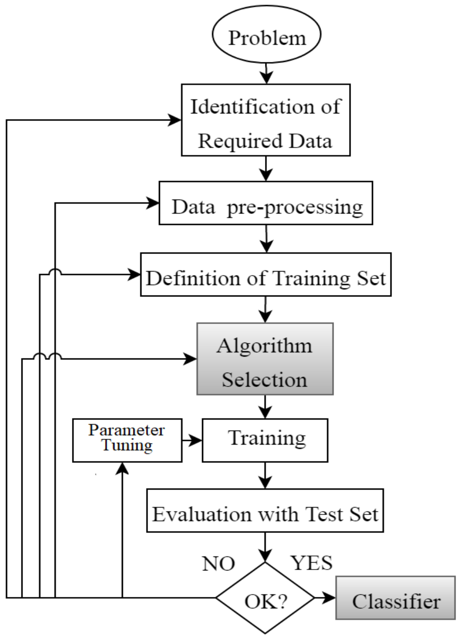

Fig.1. The process of supervised Machine Learning

Unsupervised learning is more common to human brain structure and thus it is important.

The main thing that unsupervised learning methods need to work with are the observed input patterns x_i, that are frequently thought to be independent samples from an underlying unknown probability distribution $P_I\,[x]$, and certain to what is significant. There are two classes of methods that have been proposed for unsupervised learning. The one, density estimation techniques which explicitly create statistical models (e.g. Bayesian Networks) determining how underlying causes the creation of the input. And another, feature extraction techniques which attempt to extract statistical regularities or irregularities straightforwardly from the inputs.

Finding and characterizing separate, low dimensional clusters is one of the common task for unsupervised learning approaches. The bigger class representing unsupervised learning approaches involves Maximum Likelihood density estimation methods. These depend on creating parameterized models $P[x; G]$ (with parameters G) of the probability distribution $P_I[x]$, where the forms of the models (and perhaps prior distributions over the parameters G) are reserved by a priori information in the form of the representative goals. These are referred as *synthetic or generative models*, since, given a specific value of G, they identify how to synthesize or generate samples x from $P[x; G]$, whose statistics ought to match $P_I[x]$. A normal model has the following structure:

$$P[x,G] = \sum_{y} P[y;G]P[x|y;G] \tag{1}$$

where, y denotes all the possible causes of the input x.

Maximum likelihood density estimation, and approximations to it, cover a wide range of the principles that have been recommended for unsupervised learning. This incorporates forms of the notion that the outputs should bear greater part of the information in the input; that they would have the capacity to reconstruct the inputs in good manner, for example, by identifying the observation status being independent or sparse; and ability to report on the underlying causes of the input [6] (Figure 2).

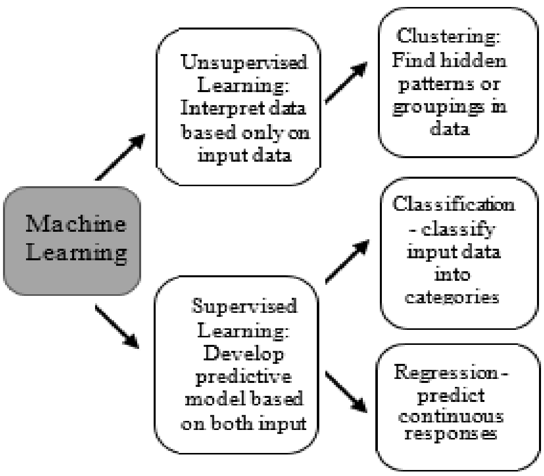

Fig.2. Machine Learning Techniques

For instance, Artificial Neural Networks [7] and Support Vector Machines (SVM) [8], represent supervised learning approaches for classification and require entirely labeled activity data. Whereas the unsupervised learning approaches, such as those based on Gaussian Mixture Models (GMMs) and Hidden Markov Models (HMMs) allow to infer automatically the labels from the data [9].

In the next sections, briefly will be explained the selected algorithms from supervised and unsupervised learning environment. Then, there will be provided combination methodology to achieve better accuracy rate while using the dataset derived from body mounted sensors to differentiate between different daily activities of humans. For that purposes Neural Networks and Hidden Markov Model will be employed. Finally, study will be concluded by experimental results and analysis.

2.1. Brief Description of NN

A typical neural network (NN) contains many simple, connected processors known as neurons, each generating a sequence of real-valued activations. Input neurons get triggered through sensors observing the environment and other neurons get triggered through weighted connections from formerly active neurons. Learning process or assignment of the credits can be explained by discovering weights that make the NN exhibit anticipated behavior. Contingent upon the problem and how the neurons are related to each other, may entail long causal chains of computational stages, where each stage transforms (frequently not in a linear way) the total activation process of the network [10].

Classification of NNs in terms of general characteristics is provided below by dividing them into several groups.

1. In the first group, there can be discovered the Feedforward Networks (FFNs), such as Multilayer Perceptron (MLP). Its fundamental feature is that their connection is only forward so they do not set up any connections between the nodes on the same layer or with previous nodes (Figure.3 [11]).

2. Through the second group, there can be found the Recurrent Networks (RCNs) that are described by the dynamism of their connectivity, so these networks save information that will be utilized later (Figure 4. [12]).

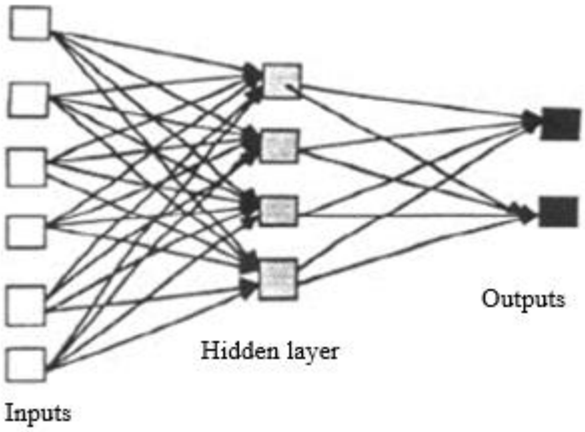

Fig. 3. Example of Feed-forward ANN

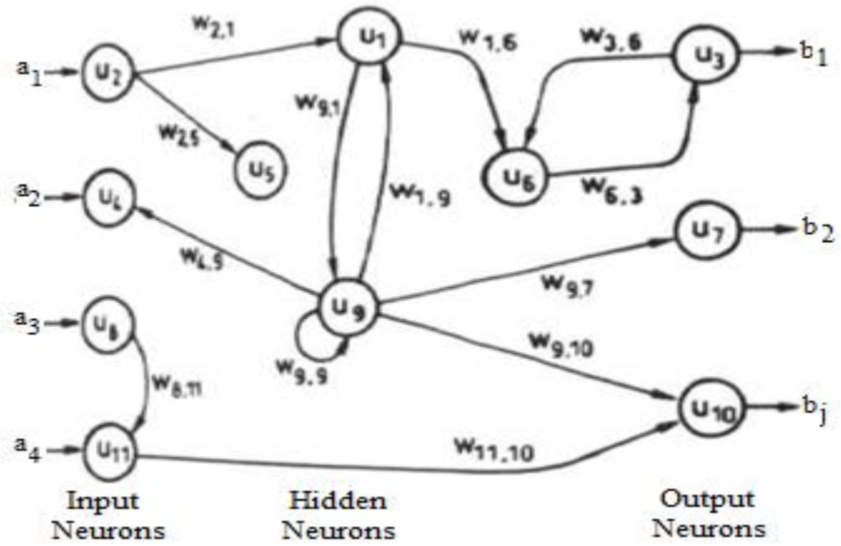

Fig.4. Example of Feedback or recurrent ANN

3. In the third group, there are the Polynomial Networks (PLNs) which usually offer proficient processing of polynomial input variables, otherwise in case of application of the sigmoidal or Gaussian functions in the training, although it would be comprehensive.

4. Modular Networks (MNs) are located in the fourth group, which involves various modules (i.e. networks) that enable solving tasks independently and then merging the answers in a logical way. The use of different network architectures is one possibility and the application of different initialization weights by leaving the same network architectures is another alternative [13].

5. in the fifth group, there can be mentioned Support Vector Machine (SVM), which represents the kernel basemodels or nucleus. The concept behind it, is to build a hyper plane as a decision surface, which maximizes the margin of separation [14].

2.2. Brief Description of MM and HMM

A Markov chain represents a discrete time stochastic process covering a finite number of states where the current state depends on the previous one [15]. In the case of human activity recognition, each activity is represented with a state. A Markov chain is well adapted to model sequential data and is often used in a more general model that is the Hidden Markov Model (HMM).

An HMM can be believed as the easiest dynamic Bayesian network. The logic behind the HMM was proposed by L. E. Baum and collaborators. It is associated to a prior work on optimal nonlinear filtering problem (i.e. stochastic processes) developed by Ruslan L. Stratonovich, who was the first to define the procedure of forward-backward algorithm [16].

As in regular Markov model, the state is visible to the observer directly, the state transition probabilities are the only parameters. In a hidden Markov model, the state is not directly visible to the observer, but output, dependent on the state, is visible. Every state has a probability distribution over the probable output tokens. Consequently, the sequence of tokens produced by an HMM provides some information about the sequence of states. The word ‘hidden’ indicates the state sequence through which the model is passes, rather than the parameters of the model.

A hidden Markov model can be assumed as a generalization of a mixture model where the hidden or latent variables, which govern the mixture component to be chosen for each observation, are associated through a Markov process instead of independent of each other.

A Markov chain consists of the following components:

$Q = q_1q_2 ...q_N$	Amount of *N* **states**
$A = a_{01}a_{02} ...a_{n1} ...a_{nn}$	A **transition probability matrix** *A*, each a_{ij} demonstrating the probability of moving from state i to state j, s.t. $\sum_{j=1}^{n} a_{ij} = 1\ \forall_i$
q_0, q_F	A special **start state** and **end (final) state** that are not related to observations

While, HMM is distinctive with the subsequent components:

$Q = q1q2 ...qN$	Amount of *N* **states**
$A = a11a12 ...an1 ...ann$	A **transition probability matrix** *A*, each a_{ij} demonstrating the probability of moving from state i to state j, s.t. $\sum_{j=1}^{n} a_{ij} = 1\ \forall_i$
$O= o1o2...oT$	A sequence of *T* **observations**, each of which is drawn from a vocabulary $V = v_1, v_2, ...v_V$
$B = bi(ot)$	A sequence of **observation likelihoods**, also known as **emission probabilities**, each expressing the probability of an observation of being produced from a state *i*.
$q0, qF$	A special **start state** and **end (final) state** that are not related with observations, together with transition probabilities $a_{01}a_{02} ...a_{0n}$ out of the start state and $a_{1F}a_{2F}...a_{nF}$ into the end state.

A first-order hidden Markov model instantiates two simplifying rules. In the first place, as with a first-order Markov chain, the probability of a specific state depends only on the previous state:

$$\text{Markov Assumption: } P(q_i|q_1...q_{i-1}) = P(q_i|q_{i-1}) \tag{2}$$

Next, the probability of an output observation *oi* depends only on the state that generated the observation q_i and not on any other states or any other observations:

$$\text{Output Independence:} \quad P(oi|q1...qi,...qT,o1,...oi,...oT) = P(oi|qi) \tag{3}$$

To describe HMM algorithm, a significant tutorial can be found by Rabiner [9], who presented the idea that hidden Markov models ought to be described by three principal problems:

1: Problem (Likelihood):	Provided an HMM $\lambda = (A,B)$ and an observation sequence *O*, define the likelihood $P(O\|\lambda)$.
2: Problem (Decoding):	Provided an observation sequence *O* and an HMM $\lambda = (A,B)$, find out the best hidden state sequence *Q*.
3: Problem (Learning):	Provided an observation sequence *O* and the set of states in the HMM, learn the HMM parameters *A* and *B*.

Calculation of Likelihood: Provided an HMM $\lambda = (A, B)$ and an observation sequence *O*, define the likelihood $P\ (O|\lambda)$.

Above mentioned principles can be further described in the following way:

- Through Hidden Markov models (HMMs) a sequence of observations can be related to a sequence of hidden classes or hidden states that describe the observations.
- The process of learning the sequence of hidden states, providing the sequence of observations, is referred as decoding or inference. For decoding purposes the Viterbi algorithm is commonly utilized.
- *A* transition probability matrix and the *B* observation likelihood matrix represent the parameters of an HMM. Training of these matrices is possible using the Baum-Welch or forward-backward algorithm [17].

2.3. Supervised and Unsupervised Algorithm Combination Methodology

It is known fact that HMMs suffer from intrinsic limitations, mostly because of their arbitrary parametric assumption [18]. With this respect Artificial Neural Networks appear to be a promising alternative. The combination of the algorithms used in the study is grounded on a gradient-ascent method for global training of a hybrid ANN-HMM system, where the ANN is trained for estimating the emission probabilities of HMM states. The approach is associated to the major hybrid systems developed by Bourlard, Morgan and Bengio [19], with the goal of combining algorithm benefits within a united framework in order to overcome their constraints.

The applied method contains several functions [20]:

hmmgenerate	(Matlab function)
hmmest	(Estimates A (Transition Matrix) and PI)
	(Forward backward algorithm hmm-nn
hmmfbNN	hybrid)
	(Forward backward algorithm hmm)
hmmfbEMIS	(Find most probable path hmm-nn hybrid)
viterbiNN	(Find most probable path hmm)
viterbiEMIS	

The structure of the algorithm application is given in the Figure 5.

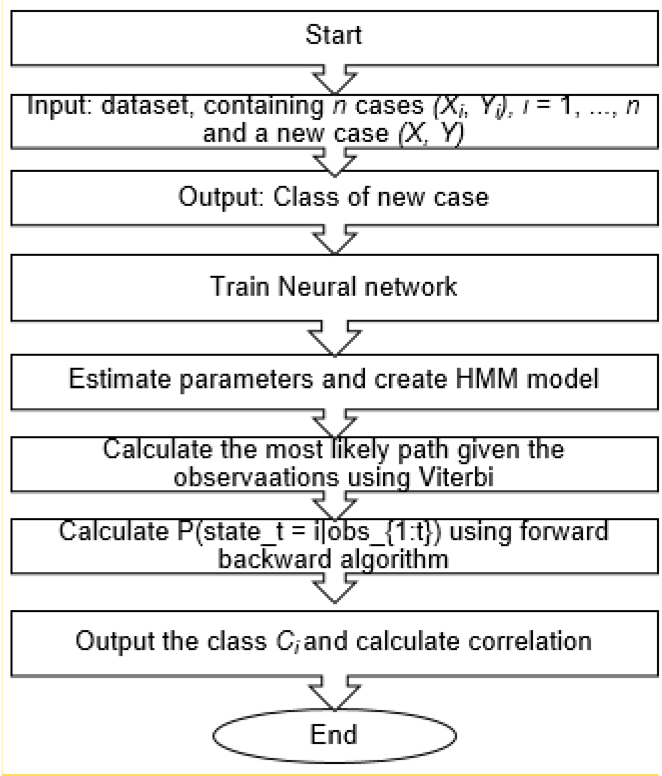

Fig. 5. The pseudo-code of the NN-HMM Hybrid Algorithm

3. Conclusion

3.1. Experimental Results and Analysis

Here experimental results of the combination of supervised and unsupervised algorithms are presented. The results are achieved through the application of ANN and HMM algorithms to the HAR dataset.

During the learning process of the dataset, new cases were classified according to the following process:

(i) First of all, by training the Neural Networks

(ii) Then, by estimating the parameters (Transition Matrix, Emission Matrix and PI) and by creating HMM model

(iii) After that, by calculating the most likely path given the observations using Viterbi algorithm

(iv) And finally, by Calculating the probability, P(state_t = i|obs_{1:t}) using Forward Backward algorithm.

Through the learning process of the training data by ANN, the training error is noticeably decreased by 2.6%, see Figure 6.

The results of the experiments is given in Table 1. Employed data in this particular experiment is row data without any calculated features. Details about the employed data can be found in [5, 21].

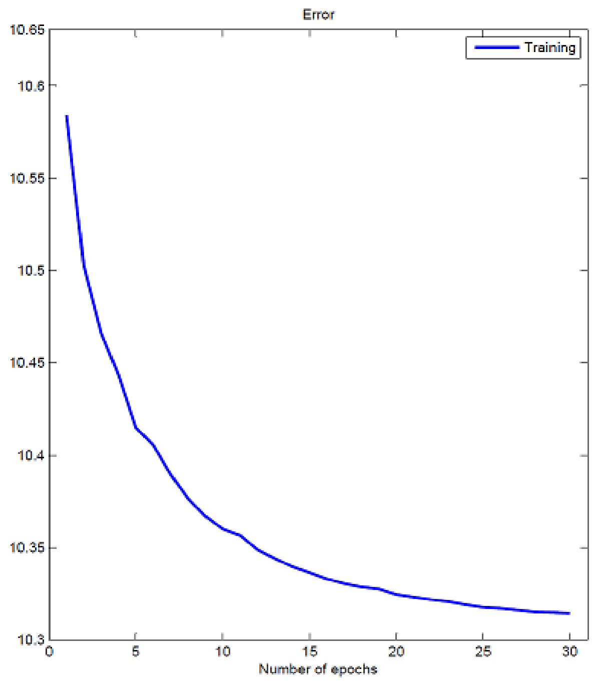

Fig.6. Decrease of error rate through the

Performances of the algorithms separately and in hybrid manner (NN-HMM) using row data

Tab.1

Class	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	Average %
HMM	75.334	96.325	69.458	87.312	73.819	91.836	69.754	94.387	73.275	89.686	62.195	80.034	80.28458
NN	88.017	93.738	75.014	89.577	85.817	86.959	77.055	92.123	87.601	88.677	92.038	94.599	87.60125
HMMNN	91.023	95.028	85.012	93.031	87.702	88.021	86.055	90.327	85.706	86.857	90.038	92.001	**89.23342**

Tables 1-4 provide global confusion matrices and demonstrate the predicted classes and prediction errors while performing classification.

Global Confusion matrix obtained with NN Classifier using row data Tab.2

True Classes	Predicted Classes: A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}	A_{12}
A_1	88.02	3.08	0	0	0	0	0	0	0	8.08	4.48	2.12
A_2	1.08	93.74	5.12	0	0	0	0	0	0	0.32	1.19	0.06
A_3	0.00	0	75.01	3.25	4.55	0	0	0	0	0	0	0
A_4	0.00	0	2.11	89.58	6.18	0	0	0	0	0	0	0
A_5	0.00	0	5.04	7.17	85.82	2.50	0	0	1.57	0.96	0	0
A_6	0.00	0	0	0	1.21	86.96	8.48	4.78	0.05	0	0	0.94
A_7	0.00	0	12.72	0	0	0.37	77.06	2.04	6.00	0	0	0
A_8	0.00	0	0	0	0	0	5.15	92.12	4.06	0	0	0
A_9	0.00	0	0	0	0	0.08	9.31	1.06	87.60	0	0	0
A_{10}	0.17	3.18	0	0	2.24	10.09	0	0	0	88.68	0.27	0
A_{11}	7.51	0	0	0	0	0	0	0	0.72	0	92.04	2.28
A_{12}	3.22	0	0	0	0	0	0	0	0	1.96	2.02	94.60

Global confusion matrix obtained with HMM using raw data Tab.3

True Classes	Predicted Classes: A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}	A_{12}
A_1	55.33	1.70	1.08	0	0.62	0	0	0	0	3.19	23.52	14.57
A_2	2.83	86.22	0.47	0	0	0	0	0	0	1.50	6.97	2.01
A_3	0.12	0	39.86	32.82	12.53	0	0	0	0	10.62	0.24	3.82
A_4	0.10	0	9.58	87.21	3.11	0	0	0	0	0	0	0
A_5	0.67	0	7.20	0.29	73.61	0.10	1.06	0	1.44	15.55	0	0.10
A_6	0	0	0	0	3.15	91.63	0.88	0	2.18	2.16	0	0
A_7	0	0	0	0	2.24	0.50	29.74	35.33	27.95	4.25	0	0
A_8	0	0	0	0	0	0	13.14	81.38	5.48		0	0
A_9	0	0	0	0	2.13	0	37.03	16.70	33.75	10.39	0	0
A_{10}	0	0	0	0	9.20	0	0	0	1.15	89.66	0	0
A_{11}	19.59	1.38	2.53	0	0	0	0	0	0	2.38	56.95	17.17
A_{12}	16.65	0	3.72	0	2.44	0	0	0	0	5.75	11.10	60.34

Global Confusion matrix obtained with NN-HMM Classifier using row data **Tab.4**

		Predicted Classes											
		A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}	A_{12}
True Classes	A_1	91.02	2.03	0	0	0	0	0	0	0	5.53	4.35	4.09
	A_2	0.11	95.03	4.80	0	0	0	0	0	0	3.97	1.32	0.69
	A_3	0.00	0	85.01	2.87	5.42	0	0	1.96	0	0	0	0
	A_4	0.00	0	1.77	93.03	5.67	0	0	0	0	0	0	0
	A_5	0.00	0	2.24	4.10	87.70	2.50	0	0	0	2.96	0	0
	A_6	0.00	1.76	0	0	1.21	88.02	3.84	2.57	0.08	0	0	0.14
	A_7	0.00	0	6.18	0	0	0.37	86.06	0.08	7.46	0	0	0
	A_8	0.00	0	0	0	0	0	3.06	90.33	6.74	0	0	0
	A_9	0.00	0	0	0	0	0.88	7.04	5.06	85.71	0	0	0
	A_{10}	0.17	1.18	0	0	0	8.23	0	0	0	86.86	2.39	0
	A_{11}	5.51	0	0	0	0	0	0	0	0.01	0	90.04	3.08
	A_{12}	3.19	0	0	0	0	0	0	0	0	0.68	1.90	92.00

To conclude, this paper presents the study of the human activity recognition dataset using hybrid of the algorithms which combines Neural Network and Hidden Markov Model based on a gradient-ascent technique for global training of a hybrid ANN/HMM system. In this particular circumstance, ANN is trained for estimating the emission probabilities of the states of the HMM. The accuracy rate for recognition using hybrid classifier is higher than in separately applied algorithms, which can be explained by approximately 9% better classification in case of HMM and 2% enhancement while applying ANN to the dataset.

Based on the study achievements, different combination of several classifiers can constitute a promising approach as many classifiers applied to the same dataset have the potential to generate different decision boundaries, which are able to display different patterns, provide complementary decisions and advance the accuracy level.

ლიტერატურა-References-Литература:

1. Zoila Ruiz, J. S.-R. (2017). A Survey of Machine Learning Methods for Big Data. *Springer, Cham.* Springer International Publishing AG. (In English)
2. Maimon, O., & Rokach, L. (2005). *Introduction to Supervised Methods.* (In English)
3. Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. (2007). Machine learning: a review of classification and combining techniques. Springer. (In English)
4. Kotsiantis, S.B. (2007). *Supervised Machine Learning: A Review of Classification Techniques.* (In English)
5. Attal, F., Mohammed, S., Dedabrishvili, M., & Chamroukhi, F. (2015). Physical Human Activity Recognition Using Wearable Sensors. *Sensors,* 31314-31338. (In English)
6. Dayan, P. (1999). *Unsupervised Learning.* (In English)
7. Webb, A. (2003). *Statistical Pattern Recognition.* John Wiley & Sons: Malden, MA, USA. (In English)

8. Vapnik, V. (2000). *The Nature of Statistical Learning Theory.* Springer Verlag. (In English)

9. Rabiner, L. R. (1989). *A tutorial on hidden Markov models.* (In English)

10. Schmidhuber, J. (2014). *Deep Learning in Neural Networks: An Overview.* (In English)

11. Artificial Intelligence https://www.iusb.edu/ ~danav/teach/c463/forward_nn.gif (In English)

12. Artificial Intelligence https://www.iusb.edu/ ~danav/teach/c463/rec_nn.gif (In English)

13. Zhang, G. P., & Berardi, V. L. (2001). Time series forecasting with neural network ensembles: An application for exchange rate prediction. *Journal of the Operational Research Society 52(6)*, 652–664. (In English)

14. Carpinteiro, O., Leite, J., Pinheiro, C., & Lima, I. (2011). Forecasting Models for Prediction in Time Series. *Artificial Intelligence Review.* Springer. (In English)

15. Chamroukhi, F., Mohammed, S., Trabelsi, D., Oukhellou, L., & Amirat, Y. (2013). Joint segmentation of multivariate time series with hidden process regression for human activity recognition. . *Neurocomputing, 120*, 633–644. (In English)

16. Bonomi, A., Goris, A., Yin, B., & Westerterp, K. (2009). Detection of type, duration, and intensity of physical activity using an accelerometer. *Med. Sci. Sports Exerc. 41*, pp. 1770–1777. (In English)

17. Jurafsky, D., & Martin, H. J. (2017). Chapter 9: Hidden Markov Models. (In English)

18. Trentin, E., & Gori, M. (2003). Robust Combination of Neural Networks and Hidden Markov Models for Speech Recognition. *IEEE Transactions On Neural Networks, Vol. 14, No. 6.*
(In English)

19. Bourlard, H., and Morgan, N. (1998), "Hybrid HMM/ANN Systems for Speech Recognition: Overview and New Research Directions," in *Adaptive Processing of Sequences and Data Structures*, C.L. Giles and M. Gori (Eds.), Lecture Notes in Artificial Intelligence (1387), Springer Verlag (ISBN 3-540-64341-9), pp. 389-417. (In English)

20. Palm, R. B. (2012). *Prediction as a candidate for learning deep hierarchical models of data.* Lyngby, Denmark: Technical University of Denmark, DTU Informatics, E-mail: reception@imm.dtu.dk. (In English)

21. Dedabrishvili, M. (2017). Study of Activity Recognition Dataset Using Combined Probabilistic and Instance Based Algorithms. *Journal of Technical Science & Technologies; Volume 6, Issue 2*, 27-33. (In English)

კონტროლირებადი და არაკონტროლირებადი დასწავლის ალგორითმების ჰიბრიდი ადამიანის აქტივობის ამოცნობისათვის

მარიამ დედაბრიშვილი, ირაკლი როდონაია
შავი ზღვის საერთაშორისო უნივერსიტეტი

რეზიუმე

კლასიფიკაციის პროცესი გამოიყენება მრავალ აპლიკაციაში. ეს პროცესი ხორციელდება სხვადასხვა ალგორითმების მეშვეობით კონტროლირებადი და არაკონტროლირებადი სწავლების პირობებში. წინამდებარე სტატია აღწერს ორივე მიდგომის შერჩეულ წარმომადგენლებს, კერძოდ, ხელოვნურ ნეირონულ ქსელებს (ANN) და დამალულ მარკოვის მოდელებს (HMM). ალგორითმები გამოიყენება სხეულზე დამაგრებული სენსორებიდან წამოსული მონაცემების შესასწავლად, ადამიანის ყოველდღიური აქტივობების დიფერენცირებისთვის. უფრო მეტიც, კვლევაში წარმოდგენილია ამ ალგორითმების ჰიბრიდი, რომელიც მიზანმიმართულია კლასიფიკატორთა ანსამბლის გამოყენებით კლასიფიკაციის სიზუსტის გასაუმჯობესებლად. მიღებული შედეგები აჩვენებს, რომ ალგორითმების კომბინაცია აღემატება ცალ-ცალკე გამოყენებული ANN და HMM კლასიფიკაციის სიზუსტეს, როდესაც ისინი კლასიფიკატორის მოდელების სახით გამოიყენება. მიღწევა მნიშვნელოვანია, ზოგადად, რეალურ დროში სხვადასხვა პროგრამა/აპლიკაციის შემუშავების საქმეში და ამასთან, განსაკუთრებული მნიშვნელობის მქონეა ხანდაზმული მოსახლეობისთვის, მათი ქმედების ამოცნობის თვალსაზრისით.

ГИБРИД КОНТРОЛИРУЕМЫХ И НЕКОНТРОЛИРУЕМЫХ АЛГОРИТМОВ ОБУЧЕНИЯ ДЛЯ РАСПОЗНАВАНИЯ ЧЕЛОВЕЧЕСКОЙ ДЕЯТЕЛЬНОСТИ

Дедабришвили М., Родонаиа И.
Международный Черноморский Университет

Резюме

Процесс классификации связан с различными приложениями. Этот процесс осуществляется с помощью специальных алгоритмов из контролируемой и неконтролируемой учебной среды. В настоящем документе описываются избранные представители обоих подходов, а именно искусственные нейронные сети (ANN) и скрытые марковские модели (HMM) соответственно. Алгоритмы применяются к набору данных, полученному из датчиков, установленных на теле, чтобы различать различные повседневные действия людей. Кроме того, исследование включает гибрид этих алгоритмов с целью повышения точности классификации с использованием ансамблей классификаторов. Полученные результаты показывают, что комбинация алгоритмов превосходит точность классификации отдельно используемых ANN и HMM, когда они применяются в качестве классификационных моделей. Достижение важно при разработке приложений в режиме реального времени в целом и, в частности, жизненно важно для жизни пожилых людей, когда речь идет о распознавании человеческой деятельности.