The Hierarchical Classification Model using Support Vector Machine with Multiple Kernels in Human Behavioral Pattern Recognition

Sorin Soviany[1], Virginia Săndulescu[1], Sorin Puşcoci[1]
[1] Communications Terminals and Telematics Dept., I.N.S.C.C.,
Bucharest, Romania, sorin.soviany@inscc.ro, s_virg@yahoo.com, sorin.puscoci@inscc.ro

***Abstract*— The paper proposes a classification model for human behavioral patterns recognition in which the decisions are provided based on several Support Vector Machines classifiers within a multi-level decision structure. SVMs are suitable for applications in which the input data feature spaces are very large, involving many features. The human behavior recognition is a relevant example of such application. On the other hand, the model proposes several kernels to be applied for SVMs in order to achieve improved performances for the real applications in which the human behavioral patterns recognition is required, as in the case of tele-assistance services.**

Keywords—hierarchical classification; human behavioral pattern.

I. Introduction

Telemonitoring systems for medical applications require various hardware and software components for processing input data. This is especially true for tele-assistance applications in which the subjects should be at their homes and there is a significant amount of data that should be acquired from the living environment. These applications concern the personal well-being and safety of the monitored individuals.

A significant functional component of these systems is the human behavior recognition module, providing support for the most suitable decisions in very challenging applications such as those based on ADL (Activities of Daily Living) identification of elderly people within their homes [1]. A human behavioral pattern recognition system includes all the technical means (hardware, software, decision algorithms) for the accurate recognition of the most relevant abnormal features or outliers in the human behavior [1][2].

In order to provide the optimal accuracy for human behavior recognition, typically a large amount of input data should be processed. This is important in order to generate the suitable features for the classification stage, with an approach based on machine learning algorithms.

The current solutions are typically based on supervised learning and context-based classification with Hidden Markov Models (HMM), especially when there are relationships among the behavioral classes [1]. There is still a strong interest for solutions based on various machine learning algorithms, because they have a significant optimization potential according to the personal safety applications requirements.

In this paper we define a hierarchical classification method in which the behavioral patterns recognition is performed using SVM (Support Vector Machines) with several kernels within a multi-level decision structure. SVMs are suitable for applications with large feature spaces. The human behavioral recognition is a relevant example of such application. The proposed model uses several kernels in order to find out the optimal performance.

The remainder of this paper has the following structure: Section II contains a definition of behavioral analysis-based telemonitoring systems and some related works including recent developments in this area using machine learning approaches; Section III presents the proposed classification model for behavioral recognition together with some application scenarios; Section IV concludes the paper showing that it is actually an ongoing research with working in progress.

II. TELEMONITORING USING BEHAVIORAL ANALYSIS

The telemonitoring process with behavioral analysis includes all activities for data acquisition, pre-processing and processing that provide the suitable parameters related to the subject, also with their transmission to a central application server for making the appropriate decisions.

Behavioral analysis is the process in which the most relevant information is extracted and recognized from the raw data in order to provide the suitable final decision according to the targeted application requirements. As concerning the human behavior recognition, this process requires a careful data selection in order to generate the most informative features for the further processing steps and to the final decision sub-subsystem. Typically, the data analysis process uses a history-based approach that generates the subject's actions profile or an ADL pattern with some statistics related to his or her actions during a certain period.

The typical implementations of the existing behavioral patterns recognition systems are based on supervised classification or context-based classification (using HMMs).

In this paper we only consider the supervised learning based approach.

In this case, the behavioral pattern for a telemonitored subject results from a set of input variables that are measured during a certain time. These variables describe some dynamic attributes of the telemonitoring subject, such as his/her gestures, actions or postures [1]. Besides of this primary data, more relevant features could be derived from them using various statistic amounts that are computed for certain periods.

On the other hand, the relevant data (feature sets) labeling is done in such a way to ensure a meaningful association with daily human activities. The behavioral classes are defined according to ADLs of the telemonitored subjects [1].

There are many human behavior recognition systems based on various classification algorithms which provide performance levels in accordance to their applications requirements.

For example, a classification approach with Gaussian Mixture Models and Mahalanobis distance for ADL patterns of elderly people identification is presented in [3]. In this case movement data from accelerometers are used.

A different approach for of human behavioral data classification based on Least Square-Support Vector Machine is described in [4], with application for smartphone sensors data.

Another SVM-based classification method for human behavioral patterns recognition is proposed in [5]. In this example an optimization for multi-class discrimination is applied in order to reduce the computational complexity.

These are only a few relevant examples of supervised learning systems with application on human behavioral recognition. The focus of all these developments is to design high-performance behavioral patterns recognition systems for applications with various requirements [1]. The supervised learning-based approaches are still subject for further improvements, but the achievements will be strongly dependent on the available input data and the generated behavioral feature sets. A performance vs. complexity trade-off will be required in order to provide the optimal results by finding the best operating point of the classifier.

III. THE BEHAVIORAL RECOGNITION METHOD

The main contribution in this work is the hierarchical classification method that it is proposed for human behavioral patterns recognition. The overall framework for the proposed method definition includes:

- The system functional architecture;
- The behavioral features;
- The hierarchical classification sub-system based on Support Vector Machines with various kernels;
- The case study: day and night scenario applications.

A. The system functional architecture

The overall functional system architecture for the human behavioral recognition is depicted in fig.1.

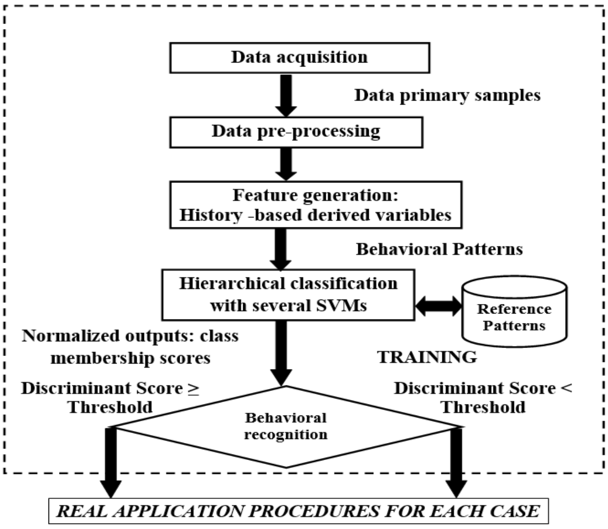

Fig. 1: The functional system architecture

The focus of our research concerns the classification module of the overall system. In order to design a suitable solution, the following elements should be specified:

- The behavioral features that are generated from the primary data. In this case a history-based approach is applied for the useful variables derivation;
- The classification design, in this case with a hierarchical approach based on several SVMs with different kernels. The classifiers outputs are normalized to provide class membership scores between 0 and 1.

The pre-processing stage includes all operations that remove the noise from the input data samples ensuring the appropriate feature sets for the further classification stage. A suitable approach is to design an automatic tool for noise detection and removal, for instance using a cluster analysis technique; in this case, we apply a combination of k-means and hierarchical clustering algorithms in order to separate noise from the useful and most informative data. The reason is that hierarchical clustering provides the initialization data for the k-means algorithm (the initial cluster centroids selection step) [6], [7].

B. The behavioral features

The behavioral features are those data elements that result from the pre-processing stage of the raw data and provide the relevant information about the individual behavior under certain conditions, for instance in case of the application scenarios based on home ADL identification. ADL concerns some typical activities of the subject. The ADL-related features are obtained based on a statistical approach, in which we count for different actions of the individual during some fixed periods. In this way we derive history-based variables, with various statistics (for instance means, standard deviations) about movements and other actions that are typical for daily living in a home environment.

This statistic approach provides a first subset of behavioral features. The second subset results from the methodology defined in [3], as used in our previous work [1]. We complete the data features generated with some additional statistics in order to provide a better association with the activities of daily living, for some specified telemonitoring scenarios (day and night). The acceleration data-based components of the behavioral feature sets provided a feature vector with 6 components according the referred methodology [3]. The additional statistic features increase the dimensionality to more than 20 (actually we limit it to 30) so that for the classification stage a more complex model is necessary; this is because the simplest LDC (linear discriminant classifier), used in [1], could be efficiently trained with small datasets only for low-dimensional feature spaces. A small-sized feature space could be covered by a small number of training examples, according to [7], and a classifier with lower complexity was suitable in that case.

C. The hierarchical classification based on Support Vector Machines with various kernels

In the supervised learning approach, the human behavior recognition is defined as a multi-class problem, in which every behavioral class is defined for a specific application scenario (in this case, a home telemonitoring application for elderly people). For the considered scenarios we define the following 5 behavioral classes (that are related to some typical home activities or movements, according to ADL framework):

- Behavioral class C1: 'normal'. There are no deviations on statistical features that are computed for the fixed periods as concerning the acceleration data and other amounts that are related to the daily activities (ADL);
- Behavioral class C2: 'small-deviation, low-level abnormal'. There are minor deviations from the normal values on the computed statistical features, within a range of +/-10%;
- Behavioral class C3: 'medium-deviation, medium-level abnormal'. There is an increased deviation from the normal values of the computed features, ranging to +/-10 to +/-20% from the normal values;
- Behavioral class C4: 'high-deviation, high-level abnormal'. The deviation of behavioral features becomes significant for the fixed periods, ranging between +/-20 to +/-30% from the normal values ;
- Behavioral class C5: 'very high-deviation, critical abnormal'. Here the deviation exceeds 30% from the normal values for all of the behavioral features. In this case, if the noise was previously removed using a clustering-based automatic tool, this class shows a very critical health condition of the telemonitoring subject.

These behavioral classes definition is based on the statistical ADL-related features and acceleration data. The thresholding is fixed based on the history of the recorded events and also on some specific application criteria.

The applied classifier is a hierarchical one, with the decision structure depicted in fig. 2. Each decision level is based on a binary classifier with target vs. non-target decisions. This classifier is named detector because it is applied to detect a target class.

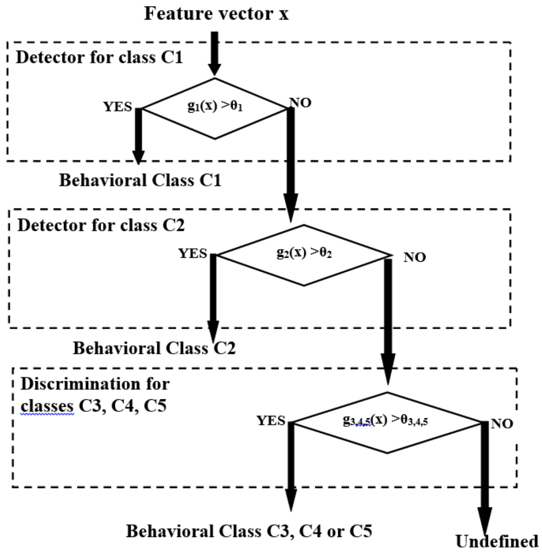

Fig.2: The hierarchical decision model for the behavioral patterns classification

The 1st level decisions are for the 'normal' class C1. The next levels concern the classes with increased deviations of the behavioral features. Here the behavioral classes C3, C4 and C5 are grouped into a single class.

Given the complexity of the feature space for the human behavioral recognition, we apply a classification model based on SVM (Support Vector Machine) for each decision level. SVM classifier has suitable performances on high dimensional feature spaces. On the other hand, the typical non-linearity of the behavioral feature space requires some kernels for SVM, in order to evaluate the discriminant function within a transformed linear space, but with the cost of a higher dimensionality.

In this case, the kernel-based SVM discriminant function g_i has the following model that is applied for the decision level i:

$$g_i(x) = \mathrm{sgn}(\sum_{j=1}^{N} \alpha_j y^j \cdot K_i(x^j, x) + w_0), i = \overline{1,3} \quad (1)$$

where [8]:

N is the number of training samples per class (considering balanced classes here);

x is the current test feature vector, x^j is the training feature vector;

α_j are the Lagrange multipliers for the optimization problem of finding the maximum margin hyperplane;

w_0 is the maximum margin hyperplane offset parameter;

y^j is the behavioral class label, $y^j \in \{-1,1\}$. The labeling concerns the following classes: C1 (target) or not-C1 for level 1, C2 (target) or not-C2 for level 2, C3,4,5 (target) or not-C3,4,5 for level 3.

K_i is the kernel mapping that should be applied. We consider a polynomial kernel for each of the decision level, with various degrees, according to [8]

$$K_i(x, y) = (x \cdot y + a_i)^{b_i}, i = \overline{1,3} \quad (2)$$

in which the coefficients a_i and b_i are constants for the decision level i. Actually b_i is the main parameter with significant impact on the classification accuracy for this kernel. The degree assignment is as following: $b_1 = 1$, $b_2 = 2$ and $b_3 = 3$. The Gaussian kernel seems to be not very appropriate for this example of behavioral patterns classification, despite of its very large utilization in many applications requiring classification with kernel SVMs.

The training is done with 90 samples per behavioral class, to prevent reaching the peaking point for a certain ratio training set size vs. feature space size.

D. Case study: day scenario, night scenario

The previously defined classification model of behavioral patterns could be applied for 2 scenarios: day scenario and night scenario. These scenarios are differentiated by the frequency of some actions of the individuals, their movements and typical activities within their homes according to the time of the day. The performances for these 2 application scenarios are evaluated on ROC (Receiver of Operating Characteristic) curves with the following target behavioral classes: C1, C2, C3,4,5, respectively (the behavioral classes C3, C4 and C5 are grouped together). Fig. 3 depicts TPR (True Positive Rate or detection rate) vs. FPR (False Positive Rate or false alarms rate) for day scenario. Fig. 4 depicts the ROC curves for the nigh scenario application of the behavioral recognition system.

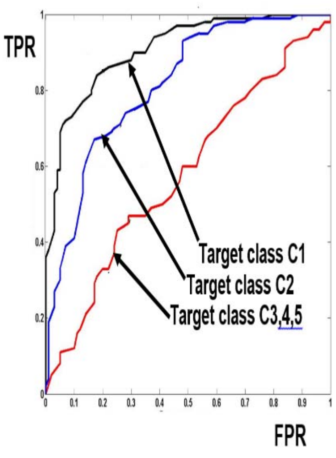

Fig. 3: The system performance for day application scenario

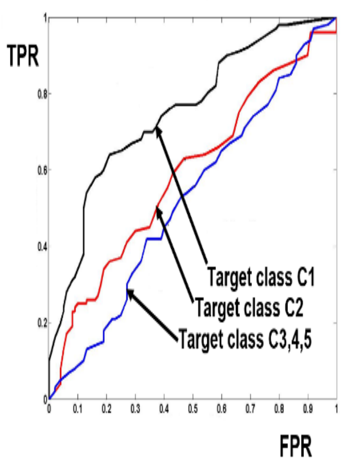

Fig. 4: The system performance for night application scenario

For the night application scenario the performance achievements are lower than for the day scenario, despite of using the same SVM classifiers with polynomial kernels. The data quality is also very important to find out an optimal operating point of the designed system. In both cases, the optimization process finds out the best operating point of the classifier, in respect to some thresholds that are specific to the recognition application. The goal is to find the maximum detection rate on the target behavioral class (TPR) for a certain false alarm rate on the same target (FPR); the last amount should be as low as possible. This optimization could be achieved with an appropriate adjustment of the Training set size vs. Feature space dimensionality ratio in order to prevent the peaking phenomenon and to meet the behavioral recognition application requirements.

IV. CONCLUSIONS

In this paper we defined a hierarchical classification model for behavioral recognition with application for personal safety of the individuals (in telemonitoring of elderly people within their homes). It is an ongoing research with assessment of several classifiers for behavioral data, in order to see what performances could be achieved by varying the models structure and for different application scenarios. These models should support the design of an optimal decision system that should be able to accurately recognize various abnormal situations that could occcure within the persons homes.

The system performance is strongly dependent on the input data quality, and the optimization process should consider not only the algorithms, but also the behavioral features. Further improvements of the proposed model should include a more careful behavioral feature selection and an improved design of the SVM classifiers for each decision level, to achieve optimal TPR vs. FPR rations, depending on the target behavioral class that is relevant for the application. A detection rate (TPR) between 85% and 95% vs. a false alarm rate (FPR) less than 25% represents a significant improvement for the proposed method versus the other actual solutions based on SVMs. These target performances could be achieved with a suitable parameterization of the kernels. The model should be evaluated for several application scenarios and several behavioral classes; a multi-class optimization could be appropriate for the further performances improvements if the application requires for several behavioral classes recognition. The depth of the classification hierarchy should be also considered within the concerns about the computational complexity and execution time.

REFERENCES

[1] S. Soviany, S. Puşcoci: “An Optimized Classification Method for Human Behavioral Patterns Recognition”, The 5th IEEE International Conference on E-Health and Bioengineering - EHB 2015 Grigore T. Popa University of Medicine and Pharmacy, Iaşi, Romania, November 19-21, 2015.

[2] S. Soviany, S.Puşcoci:., “A Hierarchical Decision System for Human Behavioral Recognition”, ECAI 2015- International Conference -7th Edition, Electronics, Computer and Artificial Intellingence, Bucharest. România, 25-27 June 2015.

[3] B.Bruno, F.Mastrogiovanni, A.Sgorbissa, T.Vernazza, R.Zaccaria: “Analysis of Human Behavior Recognition Algorithms based on Acceleration Data”, 2013 IEEE International Conference on Robotics and Automation (ICRA), Karlsruhe, Germany, May 6-10, 2013.

[4] L.Pei, R.Guinness, R.Chen, J.Liu, H.Kuusniemi, Y.Chen, L.Chen, J.Kaistinen: “Human Behavior Cognition Using Smartphone Sensors”, Sensors 2013, 13, 1402-1424.

[5] D. Anguita, A. Ghio, L. Oneto, X.Parra and J.L.Reyes-Ortiz: “Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine”, 6th International Conference on Ubiquitous Computing and Ambient Intelligence &4th International Workshop on Ambient Assisted Living-UCAmI&IWAAL 2012, December 3-5, 2012, Vitoria-Gasteiz, Spain

[6] Pang-Ning Tan, Michael Steinbach, Vipin Kumar: “Introduction to Data Mining”, Addison-Wesley, 2005. ISBN : 0321321367

[7] S.Theodoridis, K.Koutroumbas: “Pattern Recognition”, 4th edition, Academic Press, Elsevier, 2009

[8] T.Fletcher: “Support Vector Machines Explained”, 2009, https://www.iitk.ac.in/eeold/archive/courses/2013/intel-info/d4pdf2-2.pdf