Early Illness Recognition in Older Adults Using Transfer Learning

Rayan Gargees
Electrical and Computer Engineering Department
University of Missouri-Columbia
Columbia, Missouri, USA
rsgxc9@mail.missouri.edu

James Keller
Electrical and Computer Engineering Department
University of Missouri-Columbia
Columbia, Missouri, USA
KellerJ@missouri.edu

Mihail Popescu
Health Management and Informatics Department
University of Missouri-Columbia
Columbia, Missouri, USA
PopescuM@health.missouri.edu

***Abstract*—Predicting early signs of illness in older adults by utilizing a continuous, unobtrusive nursing home monitoring system has been shown to increase the quality of life and decrease the cost of care. Illness prediction is based on sensor data such as motion and bed and uses algorithms such as support vector machine (SVM) or k-nearest neighbor (kNN). One of the greatest challenges in developing prediction algorithms for sensor networks is utilizing knowledge from previous residents for predicting behavior in new ones. In this paper, we employ a transfer learning approach for addressing the cross resident training problem. We validate our method by conducting a retrospective study on three residents from TigerPlace, a retirement community in Columbia, MO, where apartments are fitted with wireless networks of motion and bed sensors. The ground truth, the daily presence or absence of the illness, was manually evaluated using nursing visit reports from an home-grown electronic medical record (EMR) system. In this study, transfer learning SVM approach outperformed other three methods, regular SVM, one class SVM, and one class kNN, resulting in average areas under the curve (AUCs) of 0.74, 0.52, 0.60 and 0.62 for three residents, respectively.**

Keywords— Transfer learning, early illness detection, sensor networks, SVM, one class classifier.

I. Introduction

As US population is aging, there has been a significant increase in development of health monitoring technology. Also, the likelihood of developing a health problem increases as people age. In the same time, many older adults in the US prefer to live independently for as long as they are able, despite the starting of conditions such as frailty and dementia and risk of falling that require increased attention and monitoring [1,2]. Sensor networks installed in the home have emerged as a viable solution for aging in place.

Our living lab is called TigerPlace [3, 4], an aging in place facility in Columbia, Missouri. 60 apartments at TigerPlace have been fitted with sensor networks that include motion sensors, bed sensors, and depth camera since 2005. In this work, we only use data from bed and motion sensors (PIR) deployed in various apartment location such as bathroom, bedroom, kitchen and living room. The motion sensors provide information related to the resident activity around the apartment. The bed sensor provides information on bed motion (low, medium, high), heart rate (low/high) and respiration (low, high). The data provided by these sensors is securely sent to an off-site level 4 database. The sensor information may be visualized over a web interface by TigerPlace clinical personnel.

The sensor data capture behavior and physiological information about the resident that is then linked to existent medical records from the TigerPlace EMR. The days in which an EMR nursing report described some resident complaint were named "negative" or ‘‘abnormal’’ (i.e. they reported some health problem that could be linked to an abnormal behavior). The days for which no EMR report was found, were denoted as "positive" or ‘‘normal.’’ While this is somewhat a rough estimate of the abnormal behavior, it was the only one available at that time.

The main goal of this work is to detect early illness by predicting abnormal (negative) behavior that may be associated with it. To this end, we train several classifiers using sensor data to predict abnormal behavior. While training/testing a classifier with data from only one resident (apartment) provides reasonable results [5], it is not a realistic approach since when we deploy a new sensor network we need the algorithm to provide reasonable results based on previous training data. Our training approach has two challenges: first, the dataset is strongly imbalanced (typically not many abnormal event labels available) and, second, the distribution of the training data may be quite different among the various residents due to their behavior and their diseases. To overcome these problems, we investigate a solution based on transfer learning. We compare our new solution based on transfer learning method with three other methods, a regular support vector machine (SVM)[6], and two one class classifiers: a support vector domain data description (SVDD), and k-nearest neighbor data description (KNNDD) [7].

Traditional supervised machine learning techniques depend on the assumption that the training data and test data are drawn from a similar probability distribution and that the classification task is the same for both datasets [8]. However, in practice, it is often necessary to relax this assumption and consider the test information to have an alternate probability distribution or to permit the classification task to change. In this situation, traditional machine learning techniques often fail to classify the test data. Based on motion and bed sensor data from a particular resident of TigerPlace, a model can be learned to predict the present activity occurring in his/her home. However, the model may then be tested with a different resident, in a different home, or with different activity labels. If the model is not modified to the new situations, the prediction accuracy will drop significantly. Transfer learning techniques have been proposed to handle these types of situations. The motivation of transfer learning is that the knowledge learned previously could be applied to solve a new problem with better or faster solution. In the field of machine learning, transfer learning has many benefits such as saving time to learning new tasks, requiring less information from the experts, and making the learned model more robust [8].

II. METHODOLOGY

The idea of transfer learning has many definitions, but the shared component of all of them is that the learning procedure is enhanced by using extra information other than the source dataset. In each transfer learning algorithm, a source task is related to source domain, and target task is linked to the target domain. The procedure for transfer learning system needs two steps: first, learning the source task, second, transfer and use the knowledge from the first step to improve the learning of target task [9,10].

In regular linear SVM, the label of a data vector x is determined by the sign of a linear decision function, i.e., $\hat{y} = sgn\big(f(x)\big) = sgn(x^T w)$, where $w = \{w_i\}_{i=1}^{M}$ are the model parameters. Training a linear SVM classifier involves the following optimization problem:

$$min_{\omega,b}\ \|w\|^2 + \mathrm{C}\sum_{i=1}^{N} \max(0,1 - y_i(w^T x_i + b)) \tag{1}$$

where w is the perpendicular to the separating hyperplane and the scoring function which predicts the label of the test sample z_i is $f(z_i) = sign(w^T Z_i)$. The $\|w\|^2$ term encourages margin maximization and the other term, hinge loss $\sum_{i=1}^{N} \max(0,1 - y_i(w^T x_i + b))$ determines the location of the separating hyperplane by minimizing the misclassification error. The hyperparameter C controls the tradeoff between margin maximization and hinge loss.

Transfer learning support vector machine is a model based on transfer regularization formulations with adaptive SVM [11,12] which was introduced for adapting SVM classifiers for new domains. The key idea is to learn from the source model $w_s$by regularizing the distance between the learned model w and w^s. The classifier is linear and is indicated by a format vector w, with a scoring function $w^T x$, where x is the feature vector. At that point the task is to learn w for the target group using a few training instances x_i, and the source group detector w^s. The formulation is:

$$L_A = min_{w,b}\|w - \Gamma w^s\|^2 + C\sum_{i=1}^{N} max\big(0,1 - y_i(w^T x_i + b)\big), \tag{2}$$

where Γ controls the measure of transfer regularization, C controls the weight of the loss function, and N the number of samples. The main difference in this formulation is the use of squared loss instead of hinge loss [13,14]. Even though squared loss is more sensitive to outliers, it also provides an analytic least square solution for the objective. Furthermore, it enables efficient leave-one-out cross validation which is used for optimization of the hyperparameter Γ, the amount of transfer. Intuitively transfer regularization for an SVM is like a spring between Γw^s and w, and is equivalent to providing training samples from the source class. The transfer can also be understood by expanding the regularization term. Assume that w^s is l_2 normalized to 1 then

$$\|w - \Gamma w^s\|^2 = \|w\|^2 - 2\Gamma\|w\| \cos\vartheta + \Gamma^2, \tag{3}$$

where $\|w\|^2$ provides the "normal" SVM margin maximization and $-2\Gamma\|w\|\ cos\,\vartheta$ induces the transfer by maximizing $cos\,\vartheta$, i.e. by minimizing the angle ϑ between the w^s and w [15]. However, the term $-2\Gamma\|w\|\ cos\,\vartheta$ also encourages $\|w\|$to be larger (as this reduces the cost) and this prevents margin maximization. Thus Γ, which should define the amount of transfer regularization, becomes a tradeoff parameter between margin maximization and knowledge transfer.

Support Vector domain Description algorithm by [16] is also used in this paper to be compared with transfer learning SVM. SVDD can be used for outlier recognition. A spherically shaped decision boundary around an arrangement of objects is built by a set of support vectors describing the sphere boundary. It has a likelihood of changing the information to new feature spaces without much extra computational cost. Of a data set containing N data objects, $\{x_i, i = 1,\ldots,N\}$, a description is required. We try to find a sphere with minimum volume, containing all (or most of) the data objects. This is extremely sensitive to the most outlying object in the target data set. When one or a couple of remote objects are in the training set, a huge sphere is obtained which will not characterize the data very well. Therefore, we allow for some data points outside the sphere and introduce slack variables ξ (analogous to [16]).

The formulation below is to maximizing concerning α_i:

$$L = \sum_i \alpha_i(x_i.x_i) - \sum_{i,j} \alpha_i\alpha_i(x_i x_i), \tag{4}$$

with constraints $0 \le \alpha_i \le C, \sum_i \alpha_i = 1$,where α_i are the Lagrange multipliers. The inner products of objects $(x_i,\ x_j)$ can be substituted by a kernel function $K(x_i, x_j)$, when this kernel $K(x_i,\ x_j)$ satisfes Mercer's theorem. This implicitly maps the objects x_i into some feature space and when a suitable feature space is chosen, a better, tighter description can be acquired. No explicit mapping is required, the problem is expressed totaly in terms of $K(x_i, x_j)$. Therefore, all inner products (x_i, x_j) is substituted by a proper $K(x_i, x_j)$. and the problem of finding a data domain description is now given by

$$L = \sum_i \alpha_i K(x_i.x_i) - \sum_{i,j} \alpha_i\alpha_i K\big(x_i x_j\big), \tag{5}$$

with constraints $0 \le \alpha_i \le C, \sum_i \alpha_i = 1$. A test object z is accepted when

$$K(z,z) - 2\sum_i \alpha_i K(z.x_i) + \sum_{i,j} \alpha_i\alpha_j K\big(x_i x_j\big) \le R^2, \tag{6}$$

Different kernel functions K result in different description boundaries in the original input space. In this paper radial basis function kernel, RBF is used.

The third method that we used in this paper for comparison to our transfer learning SVM approach is the k-nearest neighbor data description, KNNDD [7], which is a one class classifier version of kNN. In its simple version, just the distance to the k-th nearest neighbor is utilized. In the one-class classifier

KNNDD, a test object z is accepted when its local density is larger or equal to the local density of its (first) nearest neighbor in the training set $NN^{tr}(z) = NN_1^{tr}(z)$. For the local density estimation, k = 1 is used [7]:

$$f_{NN^{tr}}(z) = NN^{tr}(z) = I\left(\frac{\|z-NN^{tr}(z)\|}{V\|NN^{tr}(z)-NN^{tr}(NN^{tr}(z))\|} \leq 1\right). \quad (7)$$

This means that the distance from object z to its nearest neighbor in the training set $NN^{tr}(z)$ is compared to the distance from this nearest neighbor $NN^{tr}(z)$ to its nearest neighbor.

III. Datasets

A. Synthetic data set

For the synthetic data set, we chose the case where the target data set lies mostly in a different distribution than a source data set from which we want to enable knowledge transfer. To simulate this scenario, we generated two imbalanced data sets, a source, or training, data set, and a target, or testing, data set. To generate data we randomly sampleD 500 and 200 source points, each from simple 2D Gaussian distributions. The first class with a mean (2,3), the second with a mean (6,3) and both with covariance matrix ((1,1.5),(1.5,3)). For the target data set we generated more imbalanced data with 500 and 50 points for the first and second class respectively. The first class with mean (-10,0), the second with mean (-9,0) and both with covariance matrix ((0.05 ,0.1),(0.1 ,8)) as shown in Fig. 1.

B. Real world data set

Sensor data of three residents from TigerPlace is considered in this paper as shown in the Table I. We mention that in addition to the sensor data we have available all the clinical records (medication, nursing visits, hospitalizations, etc.) for these residents. The labeling of each day ("good" vs. "bad"), which represents the ground truth, was performed manually by the authors based on the nursing visit reports and other clinical records. Figure 2 shows three feature (3D) visualization of Resident1 data.

TABLE I. DATA FOR THE THREE RESIDENTS IN TIGER PLACE

Resident NO	Total records	Positive (feel good) days	Negative (feel bad) days
Resdident1	441	360	81
Resdident2	744	709	35
Resdident3	499	164	335

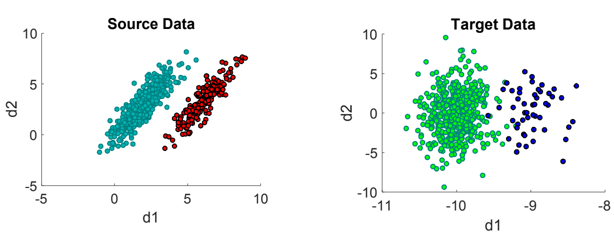

Fig. 1. Synthetic data sets: source (left), target (right).

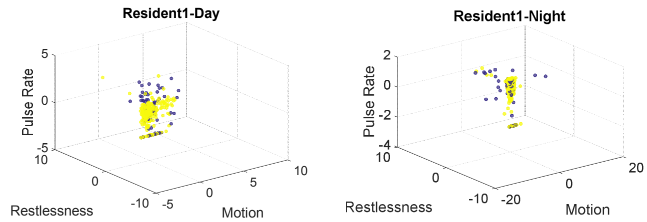

Fig. 2. Normalized data of Resident1 for the day (left) and night (right).

IV. System Architecture and Implementation

TigerPlace [3, 4] is an independent living facility for seniors planned and created because of a joint effort between Sinclair School of Nursing, University of Missouri and Americare Systems Inc. of Sikeston, Missouri. An essential objective of TigerPlace is to help the inhabitants not only deal with their sicknesses but also remain as healthy and independent as possible. Each resident included in the study has a data logger in his or her apartment that collects data from wireless sensors. The data logger date time stamps the data, and logs them into a document that is sent to a database on a protected server via a wired network connection. The sensor network comprises of a few sorts of sensors mounted in various places throughout the residents' apartments, counting motion sensors, bed sensors, and stove temperature sensor. The motion sensors are set in various spots, such as bedroom, bathroom, kitchen, living room, etc. and some of the residents have this sort of sensor installed on the door of the refrigerator, kitchen cabinets, and even drawers. They capture resident movement over his/her apartment by producing a signal as often as there is movement around them. The bed sensors are in certainty sets of sensors, made out of a pneumatic sensor strip over the bed and a motion sensor attached to the bed headboard.

The sensor strip and motion sensor joined to the bed are connected, and they work likewise to the motion sensors mentioned before: they fire as long as they identify action. Unlike the motion sensors, the bed sensor strip catches three types of activities, which are organized on three or four levels of severity. Early illness recognition depends on the intuition that if the resident does not feel well, his/her sleep and motion patterns are altered. In this study, we used four features to represent the resident behavior: the total number of firings for motion, bed restlessness, pulse rate and breathing sensors, respectively, for every hour of the day preceding the nursing report (considered at 12 pm). The number of features was doubled by splitting them in day and night.

Data processing was as follows. First, we aggregated the sensor data; features 1-4 were the sum of sensor data for the night hours (7pm-7am); features 5-8 where the sum of the sensor data for the day hours (7am-7pm) and represent the sensor activity prior a nursing visit. Then, the data is normalized for the three residents, after that, the data is passed through a transfer learning SVM algorithm described in section II. In this case, the user is prompted to control the amount of knowledge want to be transfer and the weight of the loss function by changing Γ and C values respectively. The other three algorithms which are also described in section II are applied to be compared with our main algorithm.

V. Experimental Results

A. *Experiments on the synthetic data set*

The main goal of this work was to determine the practicability of using transfer learning SVM for early illness recognition based on the sensor data by using training data from another resident. First, we experiment transfer learning SVM algorithm with our synthetic data which is imbalanced and has a different distribution of the source and target data. The ROC obtained by classifying each target data set is shown in Figure 3. From Figure 3 we see that the transfer learning SVM approach outscored the other three classifiers. The AUC was 0.998 for Transfer learning SVM while it was 0.915, 0.988 and 0.994 for the Regular SVM with RBF kernel, SVDD, and KNNDD respectively.

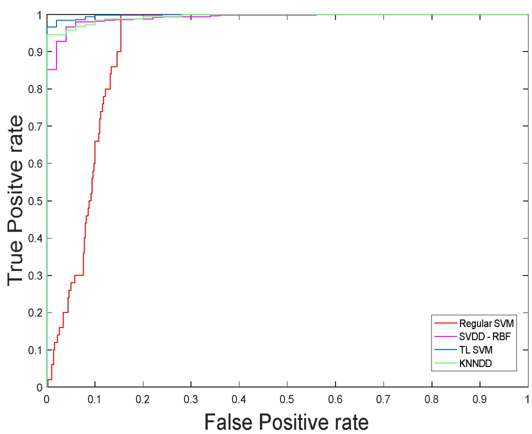

Fig. 3. ROCs curves for transfer learning SVM comparing with Regular SVM with RBF kernel, SVDD, and KNNDD classifiers run on synthetic data.

B. *Experiments on the real data set*

For the real world data set, we run a set of experiments to determine the best percentage of data that we need to transfer from the target to the source. Table II shows that the AUC increases as a number of knowledge increases then it stop increasing after reaching 10% of bad day data. The ROC obtained by classifying resident3 after training on resident1 using transfer learning SVM algorithm is given in Figure 4. For comparison, we also show the ROCs curves obtained for the same data Regular SVM with RBF kernel, SVDD, and KNNDD classifiers. It is clear that transfer learning is outperforming the other three methods. The AUC for transfer learning SVM was 0.765, while it was 0.52, 0.44 and 0.55 for Regular SVM with RBF kernel, SVDD, and KNNDD classifiers, respectively.

Table III illustrates the rest of combinations by considering specific resident as training data and testing on another one. Then the average over all train/test combinations is shown for all methods in the last line of Table III. The best average was obtained for the transfer learning SVM algorithm.

TABLE II. Amount of transfer knowledge versus AUC

Bad Day knowledge transfer	**AUC**
1 (3%)	0.7648
2 (6%)	0.7646
3 (10%)	0.7647
7 (20%)	0.7646
11 (30%)	0.7647
14 (40%)	0.7647
17 (50%)	0.7647
21 (60%)	0.7647

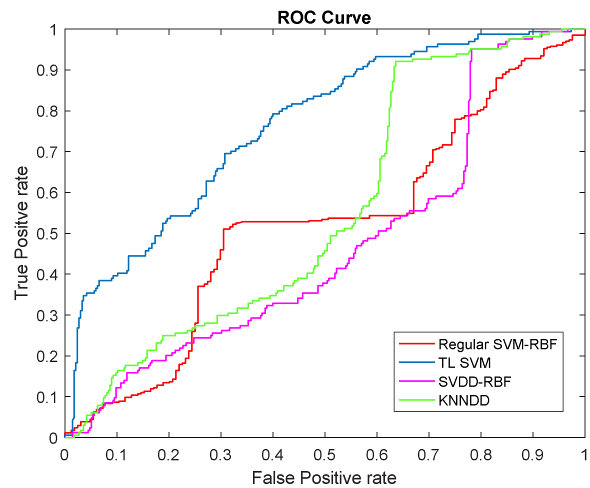

Fig. 4. ROCs curves for transfer learning SVM comparing with Regular SVM with RBF kernel, SVDD, and KNNDD classifiers run on sensor data from resident 3 after training at resident 1 in TigePlace.

TABLE III. AUC results for the all combination residents

Resident Train>>Test	**AUC Transfer Learning SVM**	**AUC Regular SVM**	**AUC SVDD**	**AUC KNNDD**
Res1 >>Res3	**0.765**	0.52	0.44	0.55
Res2>>Res3	**0.799**	0.52	0.43	0.48
Res1>>Res2	**0.708**	0.64	0.70	0.71
Res3 >>Res1	**0.653**	0.39	0.71	0.72
Res3>>Res2	**0.838**	0.45	0.71	0.63
Res2>>Res1	**0.66**	0.61	0.64	0.65
Average	***0.737***	*0.52*	*0.60*	*0.62*

VI. Conclusion

With the help of three resident case studies, we showed that the presented early illness recognition method based on transfer learning performs much better compared to individual resident learning, even with very few target instances are available. The performance of transfer approach increases when additional target instances are added, but it will stop increasing after a certain limit. Since data labeling is difficult in real elder care setting such as TigerPlace, transfer learning could help train illness recognition classifiers across residents. Using transfer learning SVM, we could detect early signs of illness in elderly residents of TigerPlace based on unobtrusive monitoring sensors. The recognition of early signs of illness might help nursing staff provide interventions that hat may anticipate grave clinical events such as heart attacks or strokes.

ACKNOWLEDGMENT

The authors would like to thank National Library of Medicine grant #R01LM012221 for the financial support.

REFERENCES

[1] R Shahriyar, F Bari, K Gourab, SI Ahamed and M Akbar, "Intelligent Mobile Health Monitoring System (IMHMS)," International Journal of Control and Automation, Vol.2, No.3, September 2009.

[2] J Rowan, ED Mynatt, "Digital Family Portrait field trial: support for Aging in Place," Proc. of the SIGCHI Conf. on Human Factors in Computing Systems, pp. 521-530), New York: ACM Press, 2005.

[3] MJ Rantz, Marek KD, Aud M.A., Johnson R.A., Otto D., Porter R. "TigerPlace: A New Future for Older Adults. J. of Nursing Care Quality," vol. 20, no. 1, pp. 1-4, 2005.

[4] M Skubic, Alexander G., Popescu M., Rantz M., Keller J., "A Smart Home Application to Eldercare: Current Status and Lessons Learned," Technology and Health Care 17(3):183-201, 2009.

[5] M Popescu, A Mahnot, "Early Illness Recognition in Older Adults Using In-Home Monitoring Sensors and Multiple Instance Learning," Methods of Informatics in Medicine, 2012, 51(4):359-67.

[6] Vapnik, V., "The Nature of Statistical Learning Theory," Springer, New York, 1995.

[7] DMJ Tax. "One-class classification," Ph.D. thesis, Delft University of Technology, June 2001.

[8] KD Feuz and DJ Cook, "Transfer Learning across Feature-Rich Heterogeneous Feature Spaces via Feature-Space Remapping (FSR)," ACM Transactions on Intelligent Systems and Technology (TIST), Volume 6 Issue 1, April 2015

[9] A Jain, J. M. Keller and J. C. Bezdek, "Quantitative and qualitative comparison of periodic sensor data," 2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), Las Vegas, NV, 2016, pp. 37-40.

[10] SJ Pan, Q Yang, "A Survey on Transfer Learning," in IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345-1359, Oct. 2010.

[11] B. Bócsi, L. Csató, and J. Peters, "Alignment-based transfer learning for robot models," The 2013 International Joint Conference on Neural Networks (IJCNN), Dallas, TX, 2013, pp. 1-7.

[12] X. Li, "Regularized Adaptation: Theory, Algorithms, and Applications," Ph.D. thesis, University of Washington, USA, 2007.

[13] J. Yang, R. Yan, and A. Hauptmann. "Adapting SVM classifiers to data with shifted distributions," In ICDM Workshops 2007, 2007.

[14] T. Tommasi and B. Caputo. "The more you know, the less you learn: from knowledge transfer to one-shot learning of object categories," In Proceedings of the British Machine Vision Conference, 2009.

[15] Y. Aytar and A. Zisserman, "Tabula rasa: Model transfer for object category detection," 2011 International Conference on Computer Vision, Barcelona, 2011, pp. 2252-2259.

[16] DMJ Tax and R.P.W. Duin, "Support vector domain description," Pattern Recognition Letters, vol. 20, pp. 1191–1199, 1999