

Stress Modelling Using Transfer Learning in Presence of Scarce Data

Pablo Hernandez-Leal¹(✉), Alban Maxhuni², L. Enrique Sucar¹,
Venet Osmani², Eduardo F. Morales¹, and Oscar Mayora²

¹ Instituto Nacional de Astrofísica, Óptica y Electrónica, Sta. María
Tonantzintla, Puebla, México

{pablohl, esucar, emorales}@ccc.inaoep.mx

² CREATE-NET, Via Alla Cascata 56/D Povo, Trento, Italy

{amaxhuni, vosmani, omayora}@create-net.org

Abstract. Stress at work is a significant occupational health concern nowadays. Thus, researchers are looking to find comprehensive approaches for improving wellness interventions relevant to stress. Recent studies have been conducted for inferring stress in labour settings; they model stress behaviour based on non-obtrusive data obtained from smartphones. However, if the data for a subject is scarce, a good model cannot be obtained. We propose an approach based on transfer learning for building a model of a subject with scarce data. It is based on the comparison of decision trees to select the closest subject for knowledge transfer. We present an study carried out on 30 employees within two organisations. The results show that the in the case of identifying a “similar” subject, the classification accuracy is improved via transfer learning.

Keywords: Stress modelling · Transfer learning · Semi-supervised learning

1 Introduction

Over the last four decades there has been rising concern in many countries about the growth and consequences of work-relevant stress and burnout. Recent reports show that stress is ranked as a second most common work-related health problem across the members of the European Union [1]; the same report shows that individuals with high levels of stress were accompanied by physical and psychosocial complaints and decreased work-control for the requirements placed on them.

To date, current approaches to measuring stress rely on self-reported questionnaires [2], which can be a source of subjectivity. As such, the availability of rich set of embedded sensors in smartphones is increasingly being used to provide objective measures of behaviour phenomena.

The objective of this study was to model stress levels from different behavioural variables obtained from smartphones when the labelled data for a person is scarce.

Collected data includes information related to psychological self-assessments that are obtained from standardised validated questionnaire and the sensor data from smartphones from 30 employees in two different organisations. Our approach learns a model for each subject which is useful not only to predict but to perform comparisons among different subjects in order to obtain groups of people (clusters) that behave similarly. When a model is built for a new subject it usually contains insufficient information to have an accurate model. For this reason we use a transfer learning approach that uses data from the most similar user in order to improve the model which results in better accuracy.

We performed an experimental analysis using real data focused on predicting stress based on several smartphone sensors. The study includes 3 aspects: (i) using semi-supervised learning to complete the models for subjects with missing data, (ii) clustering the subjects based the similarity of the learned decision trees, (iii) applying transfer learning to try to improve the model of a new user with scarce data.

2 Related Work

Smartphones have already been used to detect mental health conditions, as shown in our previous work [3–6]. However, capturing what is causing an emotion change that is influenced from work-related stressors and detect the onset of stress can be quite challenging. Current methods have tried to infer stress based on physiological signals, e.g., heart-rate variability, blood pressure, body temperatures and respiration [7]. Furthermore, recent work emphasize the importance of measuring physiological signals that would help providing short-term feedback to the users in everyday activities [8]. However, these methods have as drawback that they need to be carried at all times (and in specific places in body) in order to allow accurate and continuous monitoring.

StressSense [9] proposes a method for detecting stress based on speech analysis and the variation of speech articulation. However, in real-life activities (e.g., crowded environments) this approach may lead to misinterpretation of speech and therefore of emotion.

In order to infer relation dynamics of people and behaviour changes in daily activities, smartphones have been suggested as a promising candidate to obtain user's context. Research using smartphones for long-term monitoring [10] have reported that smartphone sensing can be used to collect many types of contextual data including: physical activities, body postures and locations. In [11], the authors build a self-tracking system called MoodScope to help its users manage their mood. The system detects users mood from smartphones usage data (e-mail, call and SMS logs, application usage, web history and location changes). The authors reported their initial 66 % accuracy of 32 subjects from their daily mood and improving to 93 % after two months of training.

Another relevant work is from Bauer et al. [12] whose work aimed at recognizing stress from 7 students before and after the exam period. The assumption is that students are likely to be under stress during the exam sessions.

They acquired data from smartphones (location, social proximity through Bluetooth, phone calls and SMS logs) and they reported an average accuracy of 53 % during the exam session. In recent work, Bogomolov et al. [13] used call logs, SMS logs, proximity data, and self-reported surveys about personality traits. The authors reported detecting daily stress levels with a 72.28 % accuracy combining real life data from different sources. However, measuring stress in uncontrolled settings poses several difficulties since it requires the efforts of human annotators about their current perceived stress and other variables relevant to stress.

Building an accurate model to predict stress requires a considerable amount of labelled data. However, users tend not to answer the questionnaires all the time, and for a new user we will have few samples. Hence, to overcome these issues, in this study we use semi-supervised learning methods that use unlabelled instances and combine the information in the unlabelled data with the explicit classification information of labelled data for improving the classification performance [14]. Another related technique used to improve accuracy when available data is limited is transfer learning [15]. To the best of our knowledge, no study has applied semi-supervised nor transfer learning methods aiming at predicting work-related stress.

3 Data Acquisition

For this study, we collected data from 30 healthy employees¹ of two organisations located in the North-eastern part of Italy for a period of 8 weeks, further described in [3]. All subjects were given a smartphone² where the application used for this study collected data continuously as a background application. The features we extracted for each subject are categorised into two types, the first group of variables includes information of user's behaviour that was collected from the smartphone sensors during work hours, these are called *objective variables*. The second group contains *subjective* information obtained from the self-reported questionnaires, that includes mood and work-relevant stress items.

3.1 Self-reports

The self-reported questionnaire included scales that have demonstrated validity for inferring stress (the Oldenburg Burnout Inventory [16]). The survey was administered via the smartphones and the data obtained from the participants was obtained three times a day: 9:00 am, 2:00 pm, and 5:00 pm, for a period of 8 weeks.³

¹ 16 male and 14 female; married (50 %) and not married (50 %); age ranged from 26–30(16.67 %), 31–40(60 %) and above 40(23.33 %); 33.33 % of participants had an academic degree, 36.7 % had bachelor degree and 30 % had high school education.

² Samsung Galaxy S3 mini 32 GB.

³ The scales ranged from 1-to-5 and items selection has been performed considering the specificity of the Italian work-context in order to ensure their suitability.

Table 1. Overall number and percentage of Stress-responses

Variable	Level	Nr.Responses (%)	Nr.Subjects
Perceived stress	High	325 (22.18 %)	27
	<i>Moderate</i>	515 (35.15 %)	30
	<i>Low</i>	625 (42.66 %)	30
	Total responses:	1465 (100.00 %)	30

Table 2. Number of parameters and features extracted from smartphones

Category	Smartphone sensors	Attributes (Feature-Extracted)
1. Physical activity level	Accelerometer	- 3-Axis (Magnitude)
		- 3-Axis (Variance Sum [17])
2. Location	Cellular	- CellID and LACID (Number of clusters (DBSCAN) [18])
	WiFi	- Access Points (Number of clusters (DBSCAN) [18])
	Google-Maps	- Latitude and Longitude (Number of clusters (DBSCAN)[18], and distances [19])
3. Social interaction	Microphone	- Proximity interaction
		- Pitch [20], Mel-MBSES [21]
4. Social activeness	Phone calls	- Number of Incoming and Outgoing Calls
		- Duration of Incoming and Outgoing Calls
		- Most common Contact-Calls
	SMS	- Number of Incoming and Outgoing SMS
		- Duration of Incoming and Outgoing SMS
		- Most common Contact-SMS
	Calendars	- Number of Calendar-Events
App usage	- Number of used applications (Social, System)	
	- Duration of used applications (Social, System)	

In Table 1, we present the overall stress responses for the whole period (8 weeks), where we include only the questionnaires obtained from (2:00 pm and 5:00 pm). The overall number of responses from both intervals was 1465. In order to simplify the measurements of the work-related stress, we have classified the stress-level into three classes: ≤ 2 as “Low-Stress”, 3 as “Moderate-Stress”, and ≥ 4 as “High-Stress”. In Table 1 is shown that during the entire monitoring weeks, 27-subjects have perceived “High-Stress” and is about 22.18 % of the monitoring weeks.

3.2 Objective Data Acquisition and Features Extraction

In Table 2 we provide an overview of the types of smartphone data acquired for the study. In addition, we provide an overview of each category used in our study:

- **Physical Activity Level:** We measured the level of activity using accelerometer data capturing 3-axial linear acceleration continuously at a rate of 5 Hz, which was sufficient to infer physical activity levels. For extracting features, we used the method developed in the framework in [17] and we measured the magnitude and the variance sum of 26 seconds ($n = 128$ -samples) from accelerometer readings. Each segment was classified into “*high*”, “*low*”, and “*none*” activity levels.
- **Location:** Location patterns and the location changes were measured using the list of WiFi access points (AP) available with their respective BSSID address, cell tower location and Google location information (latitude, longitude). We performed clustering for WiFi by means of the received signal strength (RSS) from each access point (AP). Density-based clustering (DBSCAN) [18] was used to obtain a number of different locations (clusters) in hourly basis. Similarly, DBSCAN was used to cluster Google location and cellular tower location.
- **Social Interaction:** We have used the microphone sensor embedded on smartphones in order to capture verbal interaction within the employees when they were involved in conversation in a close proximity. We have extracted two main audio features (Pitch [20] and Mel-MultiBand Spectral Entropy Signature (Mel-MBSES) [21]) to perform speech recognition. From the classifier, we obtain the values (true, false) and we measure the distribution of the interactivity in a daily basis.
- **Social Activeness:** We have included measurements from the phone usage (number and duration of phone calls, number and length of SMS messages) and the usage of two types of apps: system (Camera, Calendar, Web-browsing, Mail) and social (WhatsApp, Facebook, Skype) that were installed on the smartphones.

4 Learning and Comparison of Models

Predicting perceived stress of the user can be modelled as a classification problem. We used decision trees [22] to model subject’s stress since this representation can be easily understood by a human, and this could help to have a better understanding of what causes stress. Also, using this representation we can compare the models for the different subjects, which is important for transfer learning. Our approach is to build a decision tree (can be seen as a model) to predict stress for each subject of the study. To learn decision trees we used the C4.5 algorithm using as attributes the objective variables presented in Sect. 3.2. The class to predict is the self-reported stress level (Sect. 3.1) (*Low*, *Mid*, *High*).

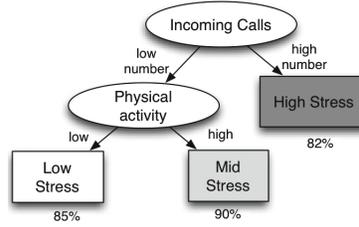


Fig. 1. Decision tree that classifies the level of Stress of one of the subjects in the study. Ovals represent decision nodes. Rectangles are leaves (terminal nodes) that give the classification value, in this case they represent low, mid or high level of stress.

In Fig. 1 we depict an example of a decision tree classifying the stress level of a subject.

Our first objective is to analyse how subjects are related to each other in terms of how similar are their models. To compare the trees, the dissimilarity measure presented in [23] is used. This measure combines the structure (the attributes of the nodes) and predictive (the predicted classes) similarities in a single value. The dissimilarity measure among two trees T_i and T_j is defined as:⁴

$$d(T_i, T_j) = \sum_{h=1}^H \alpha_h (1 - s_h) \frac{m_{h0}}{n} + \sum_{k=1}^K \alpha_k (1 - s_k) \frac{m_{0k}}{n} \quad (1)$$

where the m_{xy} values measure the predictive similarity and the α_x and s_x values measure the structural similarity. This measure can be normalized to be in the range [0–1], where 0 represents that the trees are very similar and 1 that they are totally dissimilar.

4.1 Initial Dataset

From the set of 30 subjects, we initially removed those that had a significant number of missing values (mainly in the questionnaires for self-evaluation of their stress level). Thus, having a remaining set of 18 subjects.

A decision tree was learned for each subject and using the distance in Eq. (1) we compared all pairs of models to obtain a similarity matrix. From that matrix we performed hierarchical clustering using the unweighted pair group method with arithmetic mean algorithm which yields the dendrogram depicted in Fig. 2, where a box with a colour indicates the average self-reported stress for that subject. From the figure we can observe 3 clusters with 7, 6 and 4 subjects. The largest cluster (with 7 subjects) roughly corresponds to subjects which reported low levels of stress in average (denoted by the blue boxes). The second major cluster (with 6 subjects) corresponds to subjects who reported a mid level of

⁴ Let $1, \dots, H$ as the leaves of T_i , and $1, \dots, K$ as the leaves of T_j , m_{hk} is the number of instances which belong to both the h th leaf of T_i and to the k th leaf of T_j .

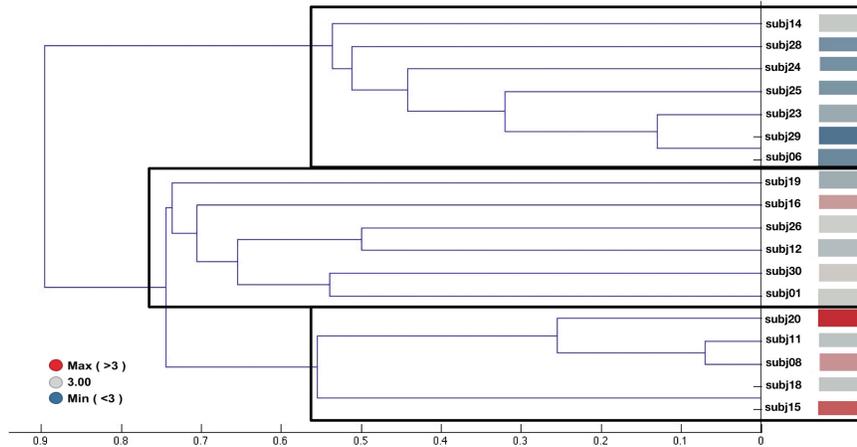


Fig. 2. Dendrogram obtained by computing similarities between models of each subject (using only 18 subjects). Three major clusters can be noted, colour boxes correspond to average stress for different subjects (Color figure online).

stress (grey boxes). The third cluster with only 4 subjects shows subjects with high and mid level of stress.

Since the initial data had a large portion of missing values (20.03 % of overall dataset), semi-supervised learning was used to fill those. In this study, we use a single classifier method called self-training [14] to classify the unlabelled data. These new data is added to the training set, the classifier is re-trained and the procedure repeated.

After applying the semi-supervised learning phase, there is enough data to compute comparisons with the 30 users of the study. A dendrogram of the similarity of all users is depicted in Fig. 3. In this case we can observe 4 main clusters. One cluster (bottom, 9 subjects) is formed by those subjects whose stress level was consistently the same for the entire study. Note, that in this case subjects were clustered by the rule of behaviour (always the same stress level) not by the prediction itself. Another cluster is formed by subjects that on average showed low and mid level of stress (6 subjects). A third cluster (with 4 subjects) contains subjects with high and mid levels of stress. The largest cluster is formed with 10 subjects which do not show a consistent level of stress.

4.2 Similarity Matrices

The process described in previous section was used to obtain a similarity matrix, depicted in Fig. 4(a), where the more similar a subject is to another the more darker that square is (subjects are ordered by clusters). Then we decided to remove 50 % of the data from all subjects (this will be useful for our transfer learning approach in the next section). The similarity matrix using this reduced

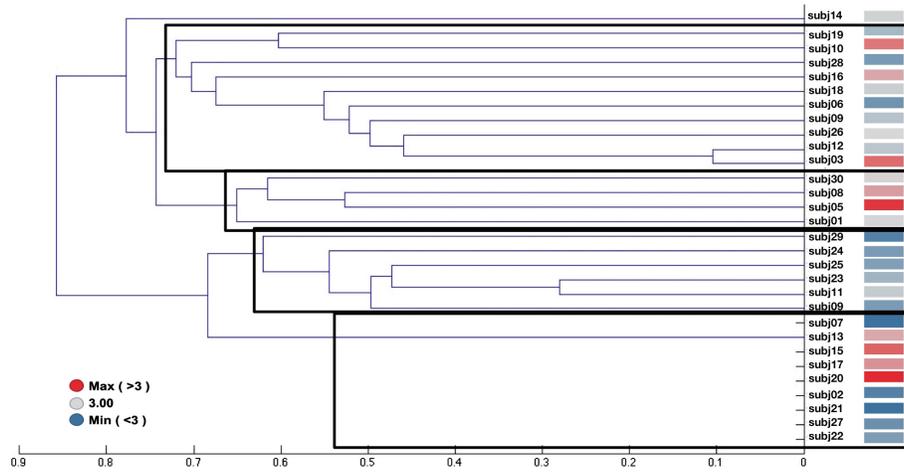


Fig. 3. Dendrogram obtained by computing similarities between models of each subject (for the 30 subjects) after using semi-supervised learning to fill missing data. Colour boxes correspond to average stress for each subject (Color figure online).

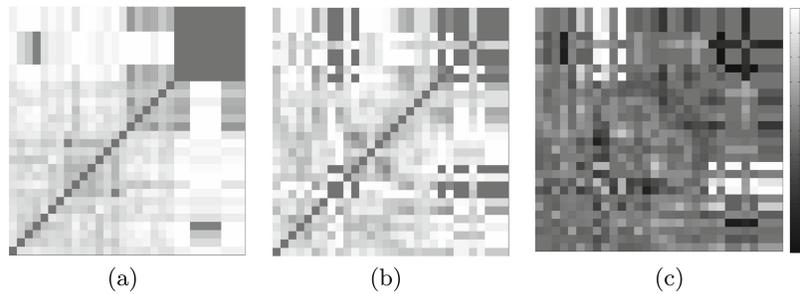


Fig. 4. Similarity matrices of 30 users using (a) all data (after semi-supervised learning) and (b) with 50 % of instances removed –darker cells indicate high similarity. (c) depicts the difference between (a) and (b); a white cell indicates a + difference, black a – difference, and grey no difference.

data is depicted Fig. 4(b), and in Fig. 4(c) we depict the matrix resulting from the difference of (a) and (b) where a grey value means no difference.

In summary, we have three similarity matrices with different quantity of data. For each matrix we computed the average values, the initial data with 18 subjects showed a more disperse set of distances with an average of 0.65 ± 0.18 (higher value, means subjects are more different to each other). After the semi-supervised algorithm was applied the average distance was 0.55 ± 0.16 even when number of subjects increased (30 subjects). Finally, when the data was reduced the average distance decreased to 0.49 ± 0.15 , which may not happen in all cases (note that removed data was randomly selected).

4.3 Missing Data and Semi-supervised Learning

Now we analyse the effect of adding/removing data in terms of the comparisons among models, i.e. similarity matrices. Thus, we computed the absolute difference of each entry between an *original* and a *modified* matrix:

$$\Delta_{i,j}(\textit{original}, \textit{modified}) = |e_{i,j}^{\textit{original}} - e_{i,j}^{\textit{modified}}| \quad (2)$$

Table 3. Percentage of entries whose value increased more than $\epsilon = 0.1, \dots, 0.5$ from matrix *original* to matrix *modified*.

Data		$\Delta(\textit{Original}, \textit{Modified}) > \epsilon$				
Original	Modified	$\epsilon = 0.1$	$\epsilon = 0.2$	$\epsilon = 0.3$	$\epsilon = 0.4$	$\epsilon = 0.5$
Initial	Semi-supervised added data	0.39	0.22	0.14	0.07	0.05
Semi-supervised	Removed random 50 % data	0.55	0.29	0.19	0.13	0.11

Table 3 shows the percentage of entries where $\Delta_{i,j} > \epsilon$ with $\epsilon = 0.1, \dots, 0.5$ between two matrices. After applying the semi-supervised approach, 39% of the entries in the similarity matrix changed more than 0.1, which is a small change in similarity for a large portion of entries. In contrast only 5% of entries changed more than 0.5. Results show that after the semi-supervised approach the similarity matrices were slightly altered with an average value of 0.12 ± 0.14 , meaning there were no drastic changes in similarities. When we reduced the data by 50% the difference between matrices increased to 0.19 ± 0.20 which is expected since the data was reduced significantly. These results are important to show that the similarity used is robust to missing values. This will be useful in the next section since we start with the reduced data but using transfer learning we can improve their accuracy.

5 Transfer Learning

Our approach assumes a set of previously learned models along with their respective data. Then, a new subject appears; however, its associated data is scarce which results in having a model with poor predictive accuracy. Information from the other subjects could be useful to improve the model. First we learn a model t_i for the new subject i and compare with the rest T using Eq. 1. The model $k = \operatorname{argmin}_{t_j \in T} d(t_i, t_j)$ which is the most similar to t_i is selected and its data is transferred to i . A new model is learned using the original and the transferred data.

We applied the described approach on the data which has a percentage of data removed. The average distance to the nearest subject is 0.28 and there are 18 subjects that have only one nearest subject. These subjects were selected for the proposed transfer learning approach.

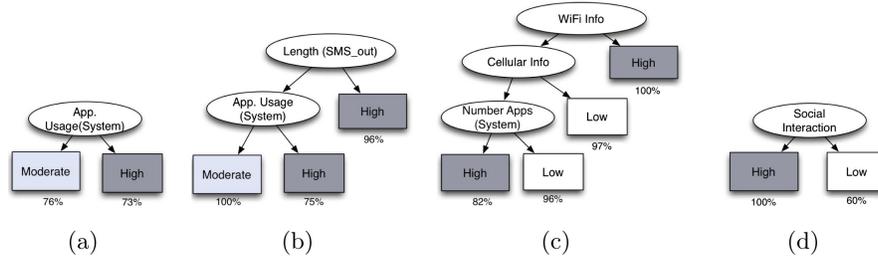


Fig. 5. Learned models of different subjects: *Subj30* (a) and its most similar *Subj17* (b). *Subj29* (c) and its most similar model *Subj05* (d).

Table 4. Classification accuracy, Δ transfer shows the difference between no transfer and transfer columns. All data shows the accuracy using all original data (upper bound). The number of initial and transferred instances is shown. The top part of the table shows the results when the distance to the closest subject is small (<0.37), while the bottom when it is large (>0.37).

Subject	No transfer	Transfer	All data	Δ Transfer	Initial I	Transf. I	$d(\text{nearest})$
<i>Subj28</i>	57.35	63.23	77.94	5.88	35	26	0.18
<i>Subj10</i>	44.89	51.02	71.42	6.13	26	31	0.27
<i>Subj12</i>	55.93	54.23	62.71	-1.69	18	31	0.32
<i>Subj18</i>	70.27	62.16	75.67	-8.10	31	18	0.32
<i>Subj24</i>	67.14	67.14	71.42	0.00	36	31	0.36
<i>Subj05</i>	70.68	65.51	86.20	-5.17	29	37	0.36
<i>Subj30</i>	42.85	53.57	78.57	10.71	37	29	0.36
<i>Subj09</i>	57.69	73.07	76.92	15.38	18	35	0.36
Average	58.35	61.24	75.10	2.89	28.75	29.75	0.31
<i>Subj25</i>	85.71	83.67	89.79	-2.04	24	31	0.39
<i>Subj04</i>	81.25	71.88	84.37	-9.38	32	31	0.42
<i>Subj08</i>	57.41	50.00	55.55	-7.41	27	35	0.46
<i>Subj16</i>	61.11	62.96	74.07	1.85	30	29	0.48
<i>Subj14</i>	51.56	48.44	82.81	-3.13	32	31	0.49
<i>Subj23</i>	53.33	50.00	58.33	-3.33	32	35	0.53
<i>Subj19</i>	60.00	53.33	90.00	-6.67	33	26	0.54
<i>Subj01</i>	72.86	61.43	78.57	-11.43	41	32	0.58
<i>Subj29</i>	62.07	44.83	79.31	-17.24	29	33	0.60
<i>Subj11</i>	65.45	74.55	72.72	9.09	30	29	0.62
Average	65.08	60.11	76.55	-4.97	31	31.2	0.51

Table 5. Classification accuracies with different threshold for applying transfer learning and percentage of instances with transfer approach.

Transfer threshold	% Transferred instances	Average accuracy
0.00	0.00	62.09
0.30	0.11	62.76
0.37	0.44	63.37
0.50	0.72	62.26
1.00	1.00	60.61

We compare the results using the reduced data with and without the transfer learning approach to classify the same data. Accuracy is obtained by learning a classifier using either: the reduced data or the reduced plus transferred data, then testing that model on the data without removed instances. The results showed that using transfer did not improve the accuracy for all subjects. This happens because there is information we are ignoring of when transfer will be more useful: the distance to the nearest subject. The idea is to use transfer when the distance is small (when the model is close to another) and not when distance increases. For example, in Fig. 5(a) and (b) we depict trees that of subjects 30 and 17 which have a $d = 0.36$. In this case trees are similar in their decision nodes. In contrast we show models of subjects 29 and 5 (in Fig. 5(c) and (d)) which have $d = 0.60$, here these trees show different decision nodes.

Using transfer only when distance was less than a *transfer threshold* (in this case < 0.37) improves the accuracy from 58.35 to 61.24 (Table 4), in contrast when $d \geq 0.37$ is better not use transfer learning since the models are far from each other and this causes a negative transfer. We performed experiments varying the threshold with values 0.0, 0.3, 0.37, 0.5, 1.0, average accuracy for all subjects is shown in Table 5. From the results we can see that the trivial approaches of not using transfer or using transfer on all subjects do not obtain the best results. However, selecting the appropriate threshold of transfer increases the accuracy.

Finally, we also performed experiments removing only 20% of the data and repeated the transfer learning process. However, the transfer learning approach did not improve the accuracy. Preliminary analysis show that transferring all the data from the closest user is not always the best choice, a selection of the data must be applied and we leave that as future work.

6 Conclusion

Using smartphones to predict the affective state of a person, such as stress, requires a considerable amount of data to build a user-specific model. However, having enough labelled data is difficult. In this paper we have proposed an approach that combines semi-supervised and transfer learning to deal with this issue. An experimental evaluation was conducted with 30 subjects to predict stress at work. The results show that transfer from similar subjects can

improve the classification accuracy, but using transfer from dissimilar persons could be detrimental. Future research ideas are to select instances of data to transfer from several similar models.

Acknowledgements. The work on this paper was partially funded by EC Marie Curie IRSES Project UBIHEALTH - 316337.

References

1. Agence européenne pour la sécurité et la santé au travail, Malgorzata Milczarek, Eusebio Rial-González, and Elke Schneider. Occupational safety and health in figures: stress at work-facts and figures. Office for Official Publications of the European Communities (2009)
2. Näätänen, P., Kiuru, V.: Bergen burnout indicator 15. Edita (2003)
3. Ceja, E., Osmani, V., Mayora, O.: Automatic stress detection in working environments from smartphones' accelerometer data: a first step. *IEEE J. Biomed. Health Inform.* **PP**(99), 1 (2015)
4. Osmani, V.: Smartphones in mental health: detecting depressive and manic episodes. *IEEE Pervasive Comput.* **14**(3), 10–13 (2015)
5. Ferdous, R., Osmani, V., Mayora, O.: Smartphone app usage as a predictor of perceived stress levels at workplace. *IEEE* **8**, 20–23 (2015)
6. Grunerbl, A., Muaremi, A., Osmani, V., Bahle, G., Ohler, S., Troster, G., Mayora, O., Haring, C., Lukowicz, P.: Smartphone-based recognition of states and state changes in bipolar disorder patients. *IEEE J. Biomed. Health Inf.* **19**(1), 140–148 (2015)
7. Bakker, J., Pechenizkiy, M., Sidorova, N.: What's your current stress level? detection of stress patterns from gsr sensor data. In: 2011 IEEE 11th International Conference on Data Mining Workshops, pp. 573–580. IEEE (2011)
8. Liu, K.K-L.: A personal, mobile system for understanding stress and interruptions. Master's thesis, MIT Media Arts and Science (2004)
9. Lu, H., Frauendorfer, D., Rabbi, M., Mast, M.S., Chittaranjan, G.T., Campbell, A.T., Gatica-Perez, D., Choudhury, T.: Stresssense: detecting stress in unconstrained acoustic environments using smartphones. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing, pp. 351–360 (2012)
10. Kocielnik, R., Sidorova, N., Maggi, F.M., Ouwerkerk, M., Westerink, J.H.D.M.: Smart technologies for long-term stress monitoring at work. In: 2013 IEEE 26th International Symposium on Computer-Based Medical Systems, pp. 53–58 (2013)
11. Likamwa, R., Liu, Y., Lane, N.D., Zhong, L.: Moodscope: building a mood sensor from smartphone usage patterns. In Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services, pp. 389–402 (2013)
12. Bauer, G., Lukowicz, P.: Can smartphones detect stress-related changes in the behaviour of individuals? In: 2012 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), pp. 423–426 (2012)
13. Bogomolov, A., Lepri, B., Ferron, M., Pianesi, F., Pentland, A.S.: Daily stress recognition from mobile phone data, weather conditions and individual traits. In: Proceedings of the ACM International Conference on Multimedia, pp. 477–486 (2014)

14. Zhu, X.: Semi-supervised learning. In: Sammut, C., Webb, G.I. (eds.) *Encyclopedia of Machine Learning*, pp. 892–897. Springer, Heidelberg (2010)
15. Pan, S.J., Yang, Q.: A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **22**(10), 1345–1359 (2010)
16. Demerouti, E., Bakker, A.B.: *The oldenburg burnout inventory: a good alternative to measure burnout and engagement. Handbook of stress and burnout in health care*. Hauppauge, NY: Nova Science (2008)
17. FUNF - Open Sensing Framework (2014). <http://funf.org/about.html>
18. Birant, D., Kut, A.: St-dbscan: an algorithm for clustering spatial-temporal data. *Data Knowl. Eng.* **60**(1), 208–221 (2007)
19. Robusto, C.C.: The cosine-haversine formula. *Am. Math. Mon.* **64**(1), 38–40 (1957)
20. Hedelin, P., Huber, D.: Pitch period determination of aperiodic speech signals. In: *1990 International Conference on Acoustics, Speech, and Signal Processing, 1990, ICASSP-90*, pp. 361–364. IEEE (1990)
21. Harris, F.J.: On the use of windows for harmonic analysis with the discrete fourier transform. *Proc. IEEE* **66**(1), 51–83 (1978)
22. Quinlan, J.R.: *C4.5 Programs for Machine Learning*. Morgan Kaufmann, San Mateo (1993)
23. Miglio, R., Soffritti, G.: The comparison between classification trees through proximity measures. *Comput. Stat. Data Anal.* **45**(3), 577–593 (2004)