Posture Based Detection of Attention in Human Computer Interaction

Patrick Heyer *, Javier Herrera-Vega, Dan-El N. Vila Rosado, Luis Enrique Sucar, and Felipe Orihuela-Espina

National Institute for Astrophysics, Optics and Electronics, Sta. Maria Tonantzintla, Puebla, Mexico
{patrickhey,vega,dnvr301080,esucar,f.orihuela-espina}@ccc.inaoep.mx
http://ccc.inaoep.mx/

Abstract. Unacted posture conveys cues about people's attentional disposition. We aim to identify robust markers of attention from posture while people carry out their duties seated in front of their computers at work. Body postures were randomly captured from 6 subjects while at work using a Kinect, and self-assessed as attentive or not attentive. Robust postural features exhibiting higher discriminative power across classification exercises with 4 well-known classifiers were identified. Average classification of attention from posture reached 76.47%±4.58% (F-measure). A total of 40 postural features were tested and those proxy of head tilt were found to be the most stable markers of attention in seated conditions based upon 3 class separability criteria. Unobtrusively monitoring posture of users while working in front of a computer can reliably be used to infer attentional disposition from the user. Human-computer interaction systems can benefit from this knowledge to customize the experience to the user changing attentional state.

Keywords: attention, human-computer interaction, posture.

1 Introduction

Human-computer interfaces can benefit from effectively interrogating the user cognitive state.

Static body postures can be mined for regulators communicating the attentional and affective state of subjects during normal human communication process [1]. In this sense, posture analysis is a plausible transparent communication channel to enhance human-computer interaction (HCI) [2–4].

In this work, we hypothesized that unacted postural features, can provide important clues regarding the user's attentional state. In other words, classification of attention is possible from postural features with reasonably high accuracy reliably across analysis methods. Moreover, we further hypothesized, that in non acted situations, some postural features can permeate through different analysis strategies consistently affording high discriminative power. Providing evidence

* Corresponding author.

R. Klette, M. Rivera, and S. Satoh (Eds.): PSIVT 2013, LNCS 8333, pp. 220–229, 2014.

regarding this latter, we perceive it as the strongest contribution of this paper. Importantly, we are not so much interested in finding the optimal classifying strategy, only a collateral goal, for which we might have then applied model selection techniques e.g. [5], but in identifying markers of attention that consistently deliver solid repeatable results regardless of the analysis approach. Indeed, intentionally, all classifiers and feature selection strategies are well known to facilitate reproduction of results as well as ensuring that we can focus our efforts on the assessment of the postural markers of attention rather than on the subtleties and implications of the analysis approach.

In order to test our hypotheses, we set up an experiment in which participants were intermittently (randomly) monitored unobtrusively, and immediately following, interrogated about their attention a few times during a normal working day. During monitoring events, skeletal features were captured employing a Kinect, to which we add a few more derived features. Extensive classification exercises using different classifiers, features selection algorithms, and parameterization allow us to explore the feature set overall classification and discriminative power of the features both individually and in association to other features. Our results suggests that both of our hypotheses are correct.

Average classification rates are above 75% despite absence of optimization attempts whilst regarding the second hypothesis we found a critical subset of 4 postural features affording sustained discriminative power regardless of the classifying analysis.

2 Related Work

Feasibility of postural analysis in HCI has been demonstrated from a range of technologies. Pressure sensing chairs are a popular choice. Using them, D'Mello and Graesser [6] investigated detection of emotions including boredom, confusion, delight, engagement and frustration. Despite the low explained variance by their logistic regression model (16%), their study had important implications for distinguishing between emotions, whether they are recognised as cognitive or affective states. Similarly Mota [7] used also a pressure sensing chair to infer three levels of attention. Their Hidden Markov Models (HMMs) achieved an overall accuracy of 87% when the user was part of the training set but only 76% in new users, questioning generalizability.

Different sensing technologies had also been explored. For instance, video analysis identified Body Lean Angle (BLA) as a successful cue in the classification of emotion [8]. This video analysis depends on color tracking, and thus needs to be calibrated to the users clothes, and would vary depending on light conditions. Grafsgaard [9] shows that seated posture can be measured in a non intrusive setting using a Kinect. Moreover, this sensing approach facilitates acquisition of a preset of metrics that can be used for postural analysis.

Irrespective of the sensing technology, it is perhaps more exciting the range of cognitive and affective states that might be somehow encoded in the posture. Besides machine learning literature, research from psychology can point us

towards cues that can be expected to capture the unconscious regulators present at normal communication processes [1].

3 Methods

An experiment has been conducted at the National Institute for Astrophysics, Optics and Electronics (INAOE) aiming at unveiling attentional markers encoded in the seated posture of people at work during office hours. Six volunteers, including 4 males and 2 females, were recruited from INAOE's academic staff and students (age range: 24-31 years; height mean±std=1670±94mm).

3.1 Experimental Setup

Following pilot tests, the Kinect sensor was mounted on a tripod behind, above and centered around the screen and at 120cm and with a tilt downward of -20°for optimal detection of upper body as illustrated in Figure 1a.

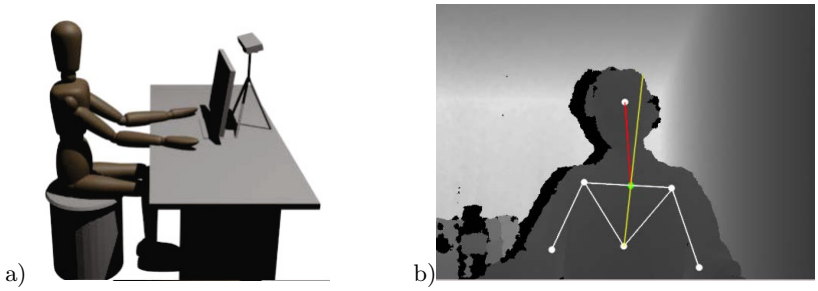

Fig. 1. a) Schematic representation of the setup of the posture capturing environment. Optimal position of the sensor was chosen following pilot over 18 different tested configurations. b) An exemplary postural sample with the feature sets represented by circles and lines. White dots corresponds to raw skeletal features (3D $< x, y, z >$ location). The connecting lines among the raw features are only illustrative as no topological relations are implied. Coloured lines represent extracted features.

For maximal ecological validity, experimental sessions were carried out at each participant regular desk. After volunteering, a suitable date would be agreed between the researcher and the subject for data collection. Data collection would proceed throughout the working day (8 hours). Instructions to the volunteer were kept to a minimum, reinforcing two critical elements: to behave as naturally as possible during the day, and to answer as honestly as possible when asked about their recent attentional state. The participants were aware of the objective of the experiment, but were blind to the times and frequency posture capture events were to occur.

3.2 Data Acquisition Protocol

To prevent anticipation as well as acted or biased postures the software was designed so that posture captures would occur at random intervals. For each posture capture event, the Kinect will take a 2 seconds video (sampled at 30 Hz) and dormant intervals lasted randomly between 40 and 60 minutes. An average of 11 posture capture events per session were acquired. The user would at all times remain unaware of the next posture capture event. Following a posture capture event, a pop up dialog in the computer screen asks the user to evaluate his preceding attitude as *ATTENTIVE* or *NOT ATTENTIVE*. The answers act as class labels conforming a self-reported ground truth.

The Kinect sensor was controlled by means of the OpenNI library [10]. Skeletal landmarks for the head, torso, shoulders and elbow joints were tracked and recorded. Wrists were ignored based on subjective appreciation. The Kinect output were saved into .oni files for off-line analysis.

3.3 Data Processing and Postural Feature Set Construction

To avoid Kinect output repetition as well as increase intra-subject subtle variability, the 60 frames long videos of each posture capture event were split into 6 equal-length chunks of 10 frames. Each chunk brought forth 1 attention observation sample. All 6 samples originating from the same posture capture event were labeled with the same attentional class. For each sample, skeletal landmarks were derived by averaging the 3D recording at each video frame. In total, 377 labeled samples were obtained. For each sample, an initial set of raw 24 postural features were retrieved directly from the skeleton expressing the average 3D $< x, y, z >$ position of 8 key locations of the upper body; *Head*, *Torso*, *Right (R)-Elbow*, *Left (L)-Elbow*, *R-Hip*, *L-Hip*, *R-Shoulder*, *L-Shoulder* as illustrated in Figure 1. In addition, this raw feature set was enriched by set of 16 derived or extracted features whether implemented from literature [9] or developed by ourselves described in Table 1. A postural feature set was therefore built for each sample incorporating a total of 40 features.

3.4 Classification Analysis

The 377 labeled feature vectors conform our dataset for classification. A total of 5940 ($= 3 \times 22 \times 3 \times 30\times$) classification exercises were carried out in Weka [11] arising for the use of 3 different feature selection strategies, 4 different classifiers (with 22 different parameterizations summarized in Table 2), 3 different dataset partitions for training-test purposes and 30 fold repetitions.

Three feature selection strategies were employed:

No Attribute Selection. Feature selection is absent. The full set of features is used for classification.

Correlation-based Feature Selection (CFS). A subset of features with high correlation with the class but uncorrelated with each other avoiding attribute redundancy is selected [12].

Table 1. Description of derived postural feature subset. Raw skeletal features can be appreciated in Figure 1.

Feature	**Description**
Elbow distance	Euclidean distance between elbows
Shoulder distance	Euclidean distance between shoulders
Collar	Midpoint R-Shoulder/L-Shoulder
Torso-Collar	Vector formed by Torso-Collar
Collar-Head	Vector formed by Collar-Head
PPunto	Cross product of Torso-Collar Collar-Head vectors
Angle	Angle on the YZ axis between Torso-Collar and Collar-Head
Quartile head	Quartile occupied by the head position of the observation with respect to the head position throughout the subject's session
Quartile torso	Quartile occupied by the torso position of the observation with respect to the torso position throughout the subject's session
Quartile collar	Quartile occupied by the collar position of the observation with respect to the collar position throughout the subject's session

Consistency-based Feature Selection. A feature set is chosen such that it is minimal on number of features and complies with a consistency criterion[13].

In addition, four well-known classifiers were tested:

Naive Bayes. A probabilistic classifier where features are assumed to be conditionally independent given the class. A two level graph (class-features) is built and class is predicted by evidence propagation of feature values [14].

Tree Augmented Naive Bayes (TAN). An extension of a Naive Bayes (NB) classifiers where the structure is augmented allowing edges among attributes, thus dispensing with its strong assumptions about independence [15].

Decision Tree. The classifier is given by a tree where attributes are tested along the tree in successive nodes and when a leaf is reached the instances is classified with the class assigned to the leaf [16].

Support Vector Machines (SVM). A classifier is obtained by constructing the hyperplane in the n-dimensional space that split optimally instances in its respective class, where the optimal criteria is defined in terms of the widest possible margin between the hyperplane and the instances (support vectors) closer to it [17].

Finally, 3 random dataset partitions were explored (train/test); 70/30%, 75/25% and 80/20%. Each classification exercise is a combination of selecting one combination of feature selection strategy, classifier and parameterization, and dataset partition at a time.

Table 2. Parameterizations for the 4 classifiers. BS: Binary Splits; CF: Confidence Factor; mNO: minimum number of objects; nF: Internal number of folds; REP: Reduce error pruning; STR: Subtree raising; UP: Unpruned; UL: Use Laplace; UKE: Use Kernel Estimator; USD: Use Supervised Discretization; RBF: Radial Basis Functions.

Classifier	Parameterization
Decision Tree (DT)	•BS=False; CF=0.25; mNO=2; nF=3; REP=False; STR=T; UP=False; UL=False •BS=True; CF=0.25; mNO=2; nF=3; REP=False; STR=T; UP=False; UL=False •BS=False; CF=0.25; mNO=2; nF=3; REP=True; STR=T; UP=False; UL=False •BS=True; CF=0.25; mNO=2; nF=3; REP=True; STR=T; UP=False; UL=False •BS=True; CF=0.25; mNO=2; nF=3; REP=False; STR=T; UP=True; UL=False •BS=False; CF=0.25; mNO=2; nF=3; REP=False; STR=T; UP=True; UL=False •BS=False; CF=0.25; mNO=2; nF=3; REP=False; STR=T; UP=True; UL=True •BS=False; CF=0.25; mNO=2; nF=3; REP=False; STR=False; UP=False; UL=True •BS=False; CF=0.25; mNO=2; nF=3; REP=False; STR=False; UP=False; UL=False
Naive Bayes (NB)	•UKE =False; USD=False •UKE =True; USD=False •UKE =False; USD=True
Support Vector Machine (SVM)	•Polynomial Kernel (p=1) •Polynomial Kernel (p=2) •Polynomial Kernel (p=3) •Polynomial Kernel (p=4) •Polynomial Kernel (p=5) •RBF Kernel (G=0.01) •RBF Kernel (G=0.05)
Tree Augmented Naive Bayes	•Estimator = simple; α=0.25 •Estimator = simple; α=0.5 •Estimator = simple; α=0.75

3.5 Evaluation

We have evaluated the performance of every pair *feature selection algorithm-classifier* based on standard metrics:

- Sensitivity=TP/(TP+FN)
- Specificity=TN/(TN+FP)
- Precision=TP/(TP+FP)
- Recall=TP/(TP+FN)
- F-Measure=2(Precision $\times$ Recall)/(Precision+ Recall)

where TP represents true positives, TN are true negatives, FP are false positives and FN are false negatives. Output of the classification exercises were stored in a database for further statistical processing.

Statistical analysis was carried out in package R [18]. Effect of the classifier and feature selection on the overall classification accuracy as measured by Precision, Recall and F-measure was assessed using an ANOVA model (significance threshold 5%). Tukey post-hoc analysis was used to conduct pairwise comparison when the ANOVA model was found significant. The point biserial correlation coefficient between measured features and categorical independent attentional class was calculated in MATLAB (Mathworks, UK).

Table 3. Summary of classification results by classifier and feature selection strategy expressed in percentage. Results indicate mean±std grand averaged across the 3 different dataset partitions and 30 fold.

Precision			
Classifier \Selection	**NoAttSelection**	**CfsSubsetEval**	**ConsistencySubsetEval**
Decision Tree	77.58±6.37	77.93±7.52	78.04±7.15
Naive Bayes	78.35±6.62	78.58±5.69	78.90±6.50
Tree augmented NB	75.55±6.05	76.43±6.57	76.43±6.52
SVM	70.46±5.94	69.01±4.94	69.42±5.27
Recall			
Classifier \Selection	**NoAttSelection**	**CfsSubsetEval**	**ConsistencySubsetEval**
Decision Tree	75.90±7.23	77.77±0.50	78.51±9.24
Naive Bayes	69.09±6.77	72.40±6.43	72.14±6.59
Tree augmented NB	75.93±6.66	77.35±6.49	78.40±7.02
SVM	84.32±5.93	86.36±5.30	86.91±5.22
F-Measure			
Classifier \Selection	**NoAttSelection**	**CfsSubsetEval**	**ConsistencySubsetEval**
Decision Tree	76.39±4.60	77.02±5.32	77.63±4.72
Naive Bayes	73.09±4.68	75.10±4.29	75.04±4.37
Tree augmented NB	75.43±4.11	76.63±4.79	77.07±4.59
SVM	76.49±3.84	76.55±3.69	77.02±4.01

4 Results

4.1 Classification of Attention

Following visual inspection all 377 samples were accepted for further analysis. Sensitivity reached $79.22 \pm 9.02\%$ (max. 98.52%) and specificity reached $66.62 \pm 13.26\%$ (max. 97.22%). Average classification reached an F-Measure of 76.47% and the best classification achieved 88.55%. Classification rates by classifier and feature selection strategy are presented in Table 3.

There was a statistically significant difference between classifiers as determined by one-way ANOVA ($F(3,5936) = 69.91$, $p < .000$). Tukey post-hoc test revealed that the F-Measure was statistically significantly different for Naive Bayes. There was a statistically significant difference between feature selection strategies as determined by one-way ANOVA ($F(2,5937) = 3.17$, $p < .000$). Tukey post-hoc test revealed that there was significant differences between not carrying out feature selection and other strategies, as well as between feature selection strategies.

4.2 Markers of Attention

In order to identify stable markers of attention different criteria might be considered; (a) those features that are more consistently selected[1], (b) the correlation

[1] This obviously excludes the no feature selection strategy.

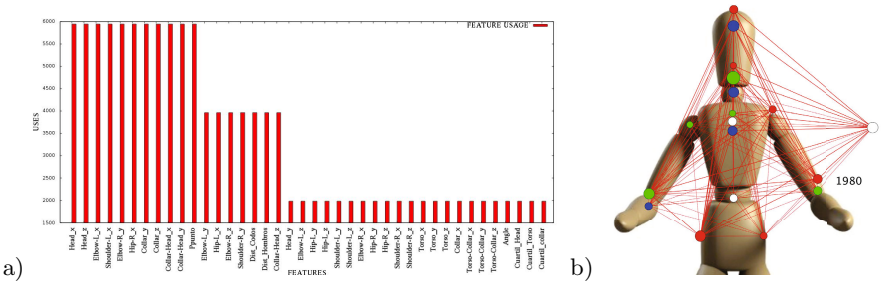

Fig. 2. a) Histogram of feature usage across feature selection strategies. The consequence of the 3 selection strategies can be easily appreciated; b) Feature co-occurrence for the consistency based feature selection. Coloured circles correspond to raw features; x - red, y - green, z - blue. White circles correspond to derived features. The one rightmost is the *PPunto*. Size of the circle is proportional to the point biserial correlation coefficient between the feature and the class descriptor. The number to the right indicates the number of cases or simulations represented. Features not selected under this scheme have been ghosted for presentation purposes. An analogous graph can be constructed for the CFS.

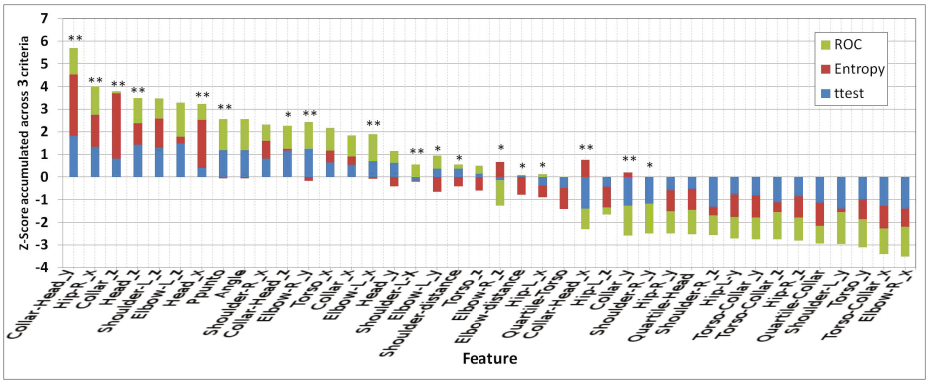

Fig. 3. Ranking of features by class separability criteria, proxy of individual discriminative power. The stacked bars corresponds to the three different criteria used, namely; t-test, entropy and receiver operating characteristic (ROC) curve. For each criterion, the absolute value of the criterion used were normalised to a z-score (within criterion output distribution). The higher the bar, the more individual discriminative power by the feature. Combined discriminative power of feature sets is the result of the feature selection strategies. The asterisks above each bar indicates when the feature has been filtered by the feature selection strategies; one asterisk indicate that it has been selected only by one strategy, and the double asterisk indicate that the feature has been selected by both feature selection strategies.

coefficient between the feature and the class, (c) degree of co-occurrence of features, and (d) feature discriminative power.

Frequency of feature usage across classification exercises is presented in Figure 2a. Degree of co-occurrence among features as well as the point biserial

correlation coefficient between the features and the attentive state is presented in Figure 2b. Finally, feature discriminative power as estimated from average classification across simulations filtering out the individual feature is illustrated in Figure 3.

5 Discussion and Conclusions

Previous work [9, 6] had already demonstrated the feasibility of estimating attention by unobtrusively monitoring the user's posture. This study further elaborates on this topic by aiming to identify robust postural features that can maintain discriminative power across different classification approaches.Our results suggest that unacted posture is a reasonable proxy of attention with average classification accuracy despite no attempt of optimizing the classification approach.

Our results suggest that persons paying attention to the computer have a tendency to tilt their heads. Specifically, we have identified the set of features conformed by Collar-Head-y, R-Hip-x, Collar-z and Head-z to maintain high discriminative power across different analysis but also being consistently selected and exhibiting high correlation with attentional state. The take home message is that the given adequate postural descriptors, classification of attention can be resilient to changes in the classification approach. However, since both the classifiers and feature selection attributes are possible confounders we cannot claim that the salient set of features found here will remain so under different configurations. Consequently, we believe it is key to focus research, not so much in the analysis but in identifying those postural markers of attention.

The major limitation of this feasibility study is the assumption that attention must be directed to the computer system. Other situations during work may obviously require the user to be attentive but not facing the screen. Currently, we lack expert evaluation of the video streams, thus there is the implicit assumption that the self-assessed attentional class labels are correct. This prevented comparison with D'Mello et al's [4]. Comparison against [9] is also difficult since they used Pearson coefficient which is inappropriate for categorical variables.

We continue to collect more data to enlarge our dataset to eliminate limitations associated to small cohort size, and plan to keep developing more aggressive features. A leave-one-out validation will explore generalization to subjects outside the cohort. This research may impact office-based HCI systems which might tailor the experience to the user changing attentional state.

Acknowledgments. This project has been funded by Microsoft Latin American and Caribbean Research (LACCIR) Federation (R1211LAC001).

References

1. Ekman, P., Friesen, W.V.: The repertoire of nonverbal behavior: Categories, origins, usage, and coding. Semiotica 1, 49–98 (1969)
2. Castellano, G., Villalba, S.D., Camurri, A.: Recognising Human Emotions from Body Movement and Gesture Dynamics. In: Paiva, A.C.R., Prada, R., Picard, R.W. (eds.) ACII 2007. LNCS, vol. 4738, pp. 71–82. Springer, Heidelberg (2007)

3. Kapoor, A., Mota, S., Picard, R.W.: Towards a learning companion that recognizes affect. In: AAAI Fall Symposium (2001)
4. D'Mello, S., Jackson, T., Craig, S., Morgan, B., Chipman, P., White, H., Person, N., Kort, B., El-Kaliouby, R., Picard, R., Graesser, A.: AutoTutor Detects and Responds to Learners Affective and Cognitive States. In: Proceedings of the Workshop on Emotional and Cognitive Issues in ITS in Conjunction with the 9th International Conference on Intelligent Tutoring Systems, p. 13 (2008)
5. Escalante, H.J., Montes, M., Sucar, L.E.: Journal of Machine Learning Research 10, 405–440 (2009)
6. D'Mello, S.K., Graesser, A.C.: Mining Bodily Patterns of Affective Experience during Learning. In: de Baker, R.S.J., Merceron, A., Pavlik, P.I. (eds.) The 3rd International Conference on Educational Data Mining (EDM), Pittsburgh, PA, USA, June 11-13 (2010)
7. Mota, S., Picard, R.W.: Automated Posture Analysis for Detecting Learner's Interest Level. In: Computer Vision and Pattern Recognition Workshop (2003)
8. Sanghvi, J., Castellano, G., Leite, I., Pereira, A., McOwan, P.W., Paiva, A.: Automatic analysis of affective postures and body motion to detect engagement with a game companion. In: Billard, A., Jr., P. H. K., Adams, J. A., Trafton, J. G (Eds.) Proceedings of the 6th International Conference on Human Robot Interaction (HRI) Lausanne, Switzerland, March 6-9, pp. 305–312 (2011)
9. Grafsgaard, J.F., Boyer, K.E., Wiebe, E.N., Lester, J.C.: Analyzing Posture and Affect in Task-Oriented Tutoring. In: Youngblood, G.M., McCarthy, P.M. (eds.) Proceedings of the Twenty-Fifth International Florida Artificial Intelligence Research Society Conference (FLAIRS), Marco Island, Florida, May 23-25. AAAI Press (2012)
10. OpenNI, `http://www.openni.org/`
11. Weka Machine Learning Project Weka University of Waikato, `http://www.cs.waikato.ac.nz/~ml/weka`
12. Hall, M.A.: Correlation-based Feature Subset Selection for Machine Learning. Hamilton, New Zealand (1998)
13. Liu, H., Setiono, R.: A probabilistic approach to feature selection - A filter solution. In: 13th International Conference on Machine Learning, pp. 319–327 (1996)
14. John, G.H., Langley, P.: Estimating Continuous Distributions in Bayesian Classifiers. In: Eleventh Conference on Uncertainty in Artificial Intelligence, San Mateo, pp. 338–345 (1995)
15. Friedman, N., Geiger, D., Goldszmidt, M.: Bayesian network classifiers. Machine Learning 29(2-3), 131–163 (1997)
16. Witten, I.H., Frank, E., Hall, M.A.: Data Mining: Practical Machine Learning Tools and Techniques, 3rd edn. Morgan Kaufmann, Burlington (2011)
17. Cortes, C., Vapnik, V.N.: Support-Vector Networks. Machine Learning 20 (1995)
18. R statistical package, `http://www.r-project.org/`