

Improving classification of posture based attributed attention assessed by ranked crowd-raters

Patrick Heyer, Jesús J. Rivas, Luis Enrique Sucar, Felipe Orihuela-Espina
National Institute for Astrophysics, Optics and Electronics,
Sta. Maria Tonantzintla, Puebla, Mexico
Email: f.orihuela-espina@ccc.inaoep.mx

Abstract—Attribution of attention from observable body posture is plausible, providing additional information for affective computing applications. We previously reported a promissory 69.72 ± 10.50 ($\mu \pm \sigma$) of F-measure to use posture as a proxy for attributed attentional state with implications for affective computing applications. Here, we aim at improving that classification rate by reweighting votes of raters giving higher confidence to those raters that are representative of the raters population. An increase to 75.35 ± 11.66 in F-measure was achieved. The improvement in predictive power by the classifier is welcomed and its impact is still being assessed.

Keywords—attention, adaptation, posture, neurorehabilitation, semi-supervised learning.

March 15, 2015

I. INTRODUCTION

Attribution of cognitive state from observable body posture, and specifically attention is plausible [1], thus providing an additional communication channel for affective computing applications. The use of cognitive and emotional state information can improve human computer interaction (HCI) adapting a system to the users needs and requirements at any moment. This work builds on previous work by [10] where the focus was to classify on one of these mind states; attributed *attention*, using crowd-evaluated images as a ground truth. The previous report reached 87.59 ± 6.59 ($\mu \pm \sigma$) of accuracy and 69.72 ± 10.50 of F-measure suggesting that posture may be a feasible proxy for attributed attention. In this work we attempted to boost the previous results by ranking the evaluators based on their votes accuracy compared to the results of a previous classification. We hypothesize that ranking the raters according to their general performance, i.e. being representative of the raters population, a more representative classification model might be achieved. We are unaware of this strategy having been used before in crowd sourcing built ground truth.

II. RELATED WORK

Posture analysis for usage in human computer interaction (HCI) using different sensing methods have been proven to work in the attribution of affective state. Sensing can be achieved by ad-hoc designed sensing devices, e.g. a pressure sensing chairs [6], [7], or generic inputs e.g. video [8] or Kinect [9], [10]. The number of cognitive and affective states being inferred from observable posture continues to increase [9], [10]: frustration, involvement, endurability, engagement or attention. This list grows when some specific aspect of

posture is isolated. For instance Rivas et al [11] have attempted to decode tiredness, tension, pain and satisfaction from arm movement and gripping force. Research in psychology is also presenting cues that can guide the kind of features present in the normal communication processes [1].

III. PREVIOUS WORK: CLASSIFICATION OF POSTURE-BASED ATTRIBUTED ATTENTION

Details of the experiment have already been published elsewhere [10]. We only briefly describe the main aspects here.

A synthetic 3D model of a seated person with 6 discrete degrees of freedom (2 for head, 1 for torso, 1 for each arm and 1 for the shoulders) described table I was built and rendered in all 216 possible postures. This model was presented to a group of 4 youngsters that acted out these postures (648 images) while image and skeletal data was captured using a Kinect positioned in front of them, the combination of postural features presented where encoded in the image name. Care was taken to ensure posture was the construct being assessed e.g. face was hidden, background was neutral, etc. These pictures where uploaded to a website shown in figure 1 that allowed volunteer raters to vote the posture as attentive/not-attentive in a crowd-sourcing manner. After a two month period the 4959 votes where gathered for processing. The tabulated data from this exercise are shown in Figure 2 showing an unbalance in the classes.

TABLE I. DEGREES OF FREEDOM OF THE 3D SYNTHETIC MODEL.

Acronym	Description	Levels
Tor	Torso inclination in degrees	0: sitting straight 30: tilted forward -10: hunched back
Sho	Shoulder inclination in degrees	0: normal straight 10: driven forth
Rar	Right arm tilt in degrees	0: Aligned to the vertical axis 45: bent upward
Lar	Left arm tilt in degrees	0: Aligned to the vertical axis 45: bent upward
HeH	Horizontal head tilt in degrees	-30: tilted left 0: straight 30: tilted right
HeV	Vertical head tilt in degrees	-30: tilted left 0: straight 30: tilted right

First, images were sorted based on *vote difference* (= *Attention* - *NoAttention*), interpreted as a metric of certainty in class assignment, in Table II column 1 shows the voting difference of the images, column 2 the total number of images presenting this difference, and columns 3-4 the number of images labeled *Attention* - *NoAttention* based on this difference. Postures ambiguously evaluated i.e. with similar number

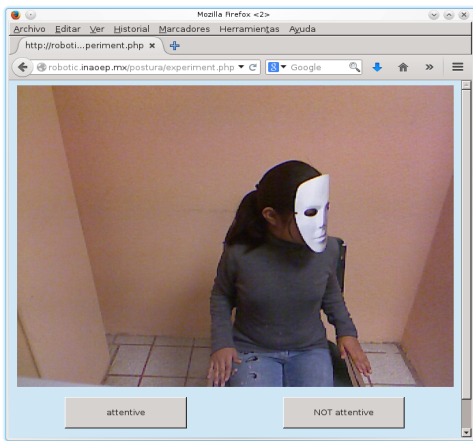


Fig. 1. Screenshot of crowd-evaluation website. (sub1Tor-0-Sho-0-Rar-0-Lar-0-HeH-30-HeV-0)

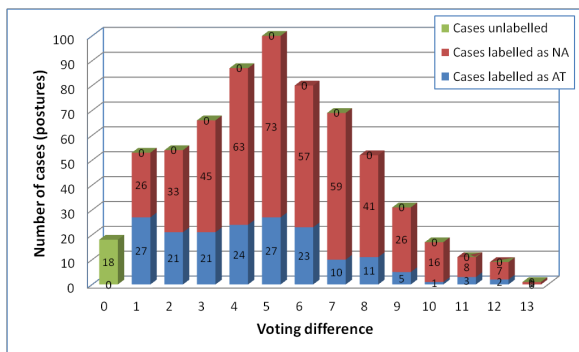


Fig. 2. Distribution of the voting according to class labeling confidence. Reproduced from [10].

of Attention/NoAttention labels, were discarded for training; and only those postures with a clear agreement among raters ($vote\ difference \geq 6$) were selected, leaving a subset of 270 images, 55 labeled as Attention and 215 as NoAttention. These were used for training.

TABLE II. SAMPLES SORTED BY VOTE DIFFERENCE.

Vote difference	Number of images	Attention	NoAttention
13	1	0	1
12	9	2	7
11	11	3	8
10	17	1	16
9	31	5	26
8	52	11	41
7	69	10	59
6	80	23	57
5	100	27	73
4	87	24	63
3	66	21	45
2	54	21	33
1	53	27	26
0	18	0	0
Totals	648	175	473

A Naïve Bayes [3] with structural improvement classifier model was learned from the labeled posture dataset using the encoded postural labels as features and a majority voting scheme for labeling. The structure learning algorithm proposed in [12] [4] is used to decide whether a structure learning algorithm could improve the accuracy. The best model was

TABLE III. PREVIOUSLY REPORTED RESULTS FROM TEST SET AT [10]. RESULTS ARE INDICATED AS MEAN \pm STD.

Set	Naïve Bayes only	Structurally improved NB.
Accuracy	84.53 \pm 2.48	87.59 \pm 6.59
Sensitivity (Recall)	50.63 \pm 20.32	56.37 \pm 22.32
Specificity	93.19 \pm 2.71	95.87 \pm 3.70
Precision	65.93 \pm 4.79	80.46 \pm 15.43
F-Measure	60.45 \pm 7.90	69.72 \pm 10.50

chosen to be enhanced with structural improvement. Classical 2×5 k-fold cross validation mechanism [2] was used for internal validity.

A. Summary of previous results

The best Naïve Bayes model (with structural improvement) reached a maximum classification accuracy of 96.29% (mean \pm std: 87.59 \pm 6.59) and F-measure reached 82.35 ($\mu \pm \sigma = 69.72 \pm 10.50$). Table III summarized these results of the structurally enhanced Naïve Bayes which are the starting position for the classification presented here.

IV. METHODS: IMPROVING CLASSIFICATION BASED ON VOTER RANKING

One of the observations we made in the previous work was that the results might have been affected by a strong unbalance in the class representation [10] where had been asymmetrically assigned. Here, to alleviate the unbalanced classification of the crowd-evaluation, a ranking system was proposed. This method ranks the voters according to their accuracy compared with the results of a classification exercise not using this ranking system. First the samples with $vote\ difference > 6$ are classified using a structurally improved Naïve Bayes classifier based on the feature set encoded in the image name and synthetically relabeled using the the classification results for the new labels, these synthetic labels were then used to rank voters based on their voting accuracy using Eq. 1:

$$voter\ rank = \frac{votes\ equal\ to\ classifier}{total\ votes\ from\ voter}. \quad (1)$$

The samples were then reassigned to the class determined by the higher ranked voters, assuming these are representative of the raters population. A new set of labels was constructed as summarized in Table IV. These newly ranked samples were used to train a completely new structurally improved Naïve Bayes classifier using samples with $vote\ difference > 7$. Finally, internal validity was again established using a 2×5 cross-fold mechanism.

Apart of helping in balancing the classes, the proposed ranking system filters noise, produced by outliers in the voting process. As any noise filter there is always a possibility of over-filtering eliminating any benefit from using crowd-evaluation to build a ground-truth.

V. RESULTS

The final classification using the retrained model achieved an average accuracy of 87.80 \pm 6.56 and a F-measure of 75.35 \pm 11.66, and reaching a peak 98% classification, as summarized in Table V. Accuracy was only marginally improved, but overall classification benefited from higher sensitivity. A

TABLE IV. SAMPLES SORTED AFTER VOTER RANKING AND LABEL REASSIGNMENT.

Vote difference	Number of images	Attention	NoAttention
12	1	0	1
11	7	2	5
10	10	1	9
9	31	5	26
8	36	5	31
7	76	9	67
6	95	23	72
5	103	26	77
4	94	19	75
3	75	29	46
2	59	27	32
1	45	21	24
0	16	0	0
Totals	648	167	481

Mann-Whitney U test (significance level set at $\alpha = 5\%$) was applied to the F-measure results, and the difference was found to be non-significant ($p = 0.06$).

TABLE V. CLASSIFICATION RESULTS (MEAN \pm STD) ON TEST SET FOLLOWING CLASS REASSIGNMENT BASED ON RANKED VOTERS.

Set	Naïve Bayes Only	Structurally improved NB.
Accuracy	85.03 \pm 3.35	87.80 \pm 6.56
Sensitivity (Recall)	51.76 \pm 22.33	73.69 \pm 15.13
Specificity	92.96 \pm 4.17	92.99 \pm 5.40
Precision	65.00 \pm 6.40	80.43 \pm 12.33
F-Measure	58.40 \pm 14.81	75.35 \pm 11.66

VI. CONCLUSION

An overall 6 point increase was observed in F-measure with the proposed strategy. This increment in the classifier overall performance suggests; (i) that our hypothesis that higher classification rates can be achieved by reweighting votes based on voter representativeness of the raters' population is feasible, and (ii) that class unbalanced was indeed strongly affecting our results. These are however not definitive conclusions; statistics failed to found significant differences, by reassigning class labels we may be affecting data topology (i.e. new results are not necessarily consequence of our hypothesis being correct), and voters bias may actually be favoring model overfitting. Nevertheless, the proposed strategy is, on an initial evaluation, a promissory solution for improving predictive power of the model.

One of the benefits of improving classification of attention in HCI is in the area of automatic tutoring where an automatic adjustment to the affective state of the student has shown an increase in the learning rate. Other areas that can directly benefit from affective state inference are psychological and medical monitoring and diagnosis of different disorders related to stress. On the other hand using posture as a proxy of affective state is useful in the sense that posture can be monitored in a non-intrusive manner using cameras and other similar sensors.

ACKNOWLEDGMENTS

Microsoft Latin American and Caribbean Research (LAC-CIR) Federation (R1211LAC001), Project SIREEX from the Mexican CONACYT (218709), Red UbiSalud from the Mexican CONACYT (U0003-2015-1-253669) and Scholarship No. 339981 from the Mexican CONACYT.

REFERENCES

- [1] P. Ekman and W. V. Friesen. The repertoire of nonverbal behavior: categories, origins, usage, and coding. *Semiotica*, 1:49–98, 1969.
- [2] Donald Michie, D. J. Spiegelhalter, C. C. Taylor, and John Campbell, editors. *Machine Learning, Neural and Statistical Classification*. Ellis Horwood, Upper Saddle River, NJ, USA, 1994.
- [3] Thomas M. Mitchell. *Machine Learning*. McGraw-Hill, Inc., New York, NY, USA, 1 edition, 1997.
- [4] Michael J. Pazzani. Searching for dependencies in bayesian classifiers. In *Preliminary Papers of Intelligence and Statistics*, pages 424–429. Springer-Verlag, 1996.
- [5] Mota, S., Picard, R. W. Automated Posture Analysis for Detecting Learner's Interest Level. In: Computer Vision and Pattern Recognition Workshop, (2003).
- [6] D'Mello, S. K., Graesser, A. C.: Mining Bodily Patterns of Affective Experience during Learning. The 3rd International Conference on Educational Data Mining (EDM), de Baker, R. S. J.; Merceron, A. Jr., P. I. P. (Eds.) 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. Proceedings of the 6th International Conference on Human Robot Interaction (HRI), Billard, A.; Jr., P. H. K., Adams, J. A., Trafton, J. G. (Eds.) Lausanne, Switzerland, March 6-9, 305–312 (2011)
- [9] Grafsgaard, J. F., Boyer, K. E., Wiebe, E. N., Lester, J.C.: Analyzing Posture and Affect in Task-Oriented Tutoring. Proceedings of the Twenty-Fifth International Florida Artificial Intelligence Research Society Conference (FLAIRS), Marco Island, Florida. May 23-25, Youngblood, G. M., McCarthy, P. M. (Eds.) AAAI Press, (2012)
- [10] Jesús Joel Rivas, Patrick Heyer, Felipe Orihuela-Espina, and Luis Enrique Sucar. Towards incorporating affective computing to virtual rehabilitation; surrogating attributed attention from posture for boosting therapy adaptation. In *10th International Symposium on Medical Information Processing and Analysis (SIPAIM 2014)*, Cartagena de India, Colombia, OCT 2014.
- [11] Jesús Joel Rivas, Felipe Orihuela-Espina, Luis Enrique Sucar, Lorena Palafox, Jorge Hernández Franco, and Nadia Berthouze-Bianchi Detecting affective states in virtual rehabilitation In *1st Workshop on Adaptive Treatment and Therapies (WATT 2015) in 9th International Conference on Pervasive Health 2015*, Istanbul, Turkey, MAY 2015.
- [12] M. Martinez, L.E. Sucar, Learning an optimal naive Bayes classifier International Conference on Pattern Recognition (ICPR), August 2006, Vol. 3, pp. 1236-1239, 2006.