Patterns, Prototypes, Performance

A. Batliner[1], D. Seppi[2], B. Schuller[3], S. Steidl[1], T. Vogt[4], J. Wagner[4], L. Devillers[5], L. Vidrascu[5], N. Amir[6], and V. Aharonson[7]

[1] Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany
[2] Fondazione Bruno Kessler (FBK) – irst, Trento, Italy
[3] Institute for Human-Machine Communication, Technische Universität München (TUM), Germany
[4] Multimedia Concepts and their Applications, University of Augsburg (UA), Germany
[5] Spoken Language Processing Group (LIMSI-CNRS), Orsay Cedex, France
[6] Dep. of Communication Disorders, Sackler Faculty of Medicine, Tel Aviv University (TAU), Israel
[7] Tel Aviv academic college of engineering (AFEKA), Tel Aviv, Israel

Abstract. We address the impact of more or less clear, i.e. prototypical, cases on the classification performance in emotion recognition.

We report on classification results for emotional user states (German database of children interacting with a pet robot, recorded at our institute). We are modelling the following four classes: motherese, neutral, emphatic, angry. Starting with five emotion labels per word, we obtain chunks, i.e. meaningful units such as clauses or phrases, with different degrees of prototypicality: the more labels in a chunk belong to the class the chunk is attributed to, the higher its prototypicality. Six sites computed acoustic and linguistic features independently from each other, following in part different strategies. The initiative to co-operate was taken by us within the European Network of Excellence HUMAINE under the name CEICES (Combining Efforts for Improving automatic Classification of Emotional user States). A total of 4232 features were pooled together and grouped into 10 low level descriptor types: duration, energy, pitch, spectrum, cepstrum, voice quality, and wavelets as acoustic, and bag-of-words, part-of-speech, and semantic classes as linguistic feature types.

We apply Support Vector Machines using 150 features each with the highest individual Information Gain Ratio. Figure 1 shows the classification performance for chunks with different prototypicality, for acoustic and for linguistic features separately, and for both acoustic and linguistic features taken together. Note that above a threshold $thr = 80$ (80 percent or more of the labels belong to the class the whole chunk is attributed to) our 4-class problem is mutilated because of sparse data, thus realistic estimates can only be obtained for $thr \leq 80$.

Our data and our results can be seen as being typical for realistic databases: a tidy, balanced set of classes is not given, and can even less be maintained when going over to more prototypical constellations. However, we can demonstrate that the degree of prototypicality chosen clearly amounts to a marked difference in classification performance, e.g. for thr w/o (baseline) vs. 70 to 16.8 percent points (26.7% relative improvement). This difference is higher than the one normally obtained by optimizing feature sets or classifiers.

From: HSS-Cooperation Seminar 2008, Wildbad Kreuth

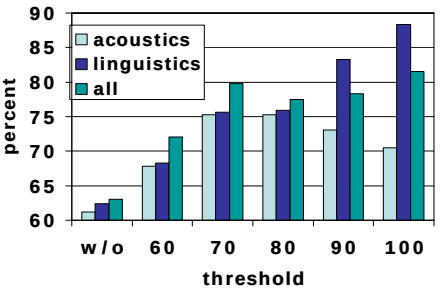

Fig. 1. Classification performance in percent on the y-axis for chunks with different prototypicality; threshold given on x-axis

If we are using only prototypes of different degrees of prototypicality for training, and are testing on the baseline thr w/o, classification performance goes down systematically. Thus prototypes cannot model variability in the data and, used for training, yield sub-optimal results. Even if our prototypes cannot simply be put on the same level as acted data, this result makes it less probable that using acted data for training is the solution for the sparse data problem.

Details and more references are given in [1].

References

1. Seppi, D., Batliner, A., Schuller, B., Steidl, S., Vogt, T., Wagner, J., Devillers, L., Vidrascu, L., Amir, N., and Aharonson, V.: Patterns, Prototypes, Performance: Classifying Emotional User States. Proceedings of Interspeech, Brisbane. 2008. to appear