Concensus of Self-features for Nonverbal Behavior Analysis

Derya Ozkan and Louis-Philippe Morency

Institute for Creative Technologies
University of Southern California
{ozkan,morency}@ict.usc.edu
http://projects.ict.usc.edu/multicomp/

Abstract. One of the key challenge in social behavior analysis is to automatically discover the subset of features relevant to a specific social signal (e.g., backchannel feedback). The way that these social signals are performed exhibit some variations among different people. In this paper, we present a feature selection approach which first looks at important behaviors for each individual, called self-features, before building a consensus. To enable this approach, we propose a new feature ranking scheme which exploits the sparsity of probabilistic models when trained on human behavior problems. We validated our self-feature concensus approach on the task of listener backchannel prediction and showed improvement over the traditional group-feature approach. Our technique gives researchers a new tool to analyze individual differences in social nonverbal communication.

Keywords: Feature selection, non-verbal behavior analysis, L_1 regularization.

1 Introduction

Nonverbal communication is a highly interactive process, in which the participants dynamically send and respond to nonverbal signals such as speech prosody, gesture, gaze, posture, and facial expression movements. These signals play a significant role in determining the nature of a social exchange. This coherence in communication plays an important role in various areas including contradict resolution [1], psychotherapeutic effectiveness [2], and improved classroom test performances [3]. One of the key challenge in social behavior analysis is to automatically discover the subset of features relevant to a specific social signal [4].

It is well known that culture, age and gender affect people's nonverbal communication [5,6]. The traditional approach for feature selection looks at the most relevant features from all observations (e.g. all human interactions in the dataset). This *group-feature* approach has the potential to select features that are not relevant to any specific individual but only to the average model. This technique is likely to miss some discriminative features which are specific to subset of the population.

A.A. Salah et al. (Eds.): HBU 2010, LNCS 6219, pp. 75–86, 2010.

In this paper, we present a feature selection approach which first looks at important behaviors for each individual, called *self-features*, before building a consensus. Figure 1 compares our self-feature concensus approach to the typical group-feature approach. To enable efficient feature selection, we propose a feature ranking scheme based on a sparse regularization method called L_1 regularization [7,8,9]. This scheme is a non-greedy ranking method where two or more features can have the same rank, meaning that these features have joint influence and they should be selected together. Our sparse feature ranking approach can be applied for both group-features and self-features.

We evaluate our approach on the task of listener feedback prediction, to predict the starting points of listener head-nods in a dyadic interaction of two people. We use a sequential probabilistic model, Conditional Random Fields, which is a recently used technique for predicting the backchannels [10]. The experiments are conducted on the RAPPORT dataset from [11] which contains 42 storytelling dyadic interactions.

The following section present related work in nonverbal behavior analysis and feature selection. In Section 3, we describe our self-feature consensus framework. Sparse ranking scheme is described in Section 4. In Section 5, we explain the dataset, features and evaluation metrics used in our experiments, and give the results on the task of listener head-nod prediction. Finally, we conclude with discussion and future work.

2 Related Work

Nonverbal behavior plays an important role in human social interactions. The ability to correctly understand and respond to social signals is considered to be the indicative of social intelligence [12] [13]. Due to it's necessity, social signal processing has became a new domain that aims to automatically sense and understand human social interactions through machine analysis [4] [14]. One of the earliest works in this domain focused on social signal detection for predicting the outcome of dyadic interactions such salary negotiations, hiring interviews, and speed-dating conversations [15]. Second focus of attention has been analysis of social interactions in multimedia recordings. There are three main tasks explored in this context: (1) analysis of interactions in small groups, (2) recognition of roles, and (3) sensing of users interest in computer characters. An extensive list of studies for each domain can be found in [4].

One of the recent approaches in dyadic interactions analysis include recognition [16] and prediction [10] of listener backchannel feedbacks. Earlier, the researchers took a unimodal approach using only either the prosodic features such as pitch and power contours [17] [18], or features like pause duration and trigram part-of-speech frequency [19]. Maatman et al. [11] presented a multimodal approach that combines the prosodic feature based method in [18] with a simple head-nod mimicking method. Later, Morency et al. [10] proposed a multimodal approach to automatically learn a predictive model of listener backchannel feedback.

Feature selection refers to the task of finding a subset of features that are most relevant to the model, and provides a good representation of data. It alleviates the problem of overfitting by eliminating the noisy features. With only the relevant features, a better understanding and analysis of data is facilitated. Based on the gradient-based feature selection method (grafting) in [20], Vail et. al. [21] proposed an incremental feature selection technique for Maximum Entropy Modeling. A Boosting-like method was presented in [22] that iteratively constructs feature conjunctions, which increases the conditional log-likelihood of the model when added. A well known feature selection technique based on L_1 regularization was also applied for conditional random fields in robot tag domain [9].

Although well studied in psychology and sociology [23] [5] [6], individual differences in nonverbal communication have not yet been explored through machine analysis. In this paper, we present a computational approach which enables a better analysis of individual differences in non-verbal behaviors.

3 Concensus of Self-features

Figure 1(a) shows an overview of our self-feature concensus approach. The first step of our algorithm is to find a ranked subset of the most relevant features for each person individually. We refer to this subset as self-features. Section 4 describes our feature ranking algorithm. Figure 1(b) compares our approach to the typical group-feature approach.

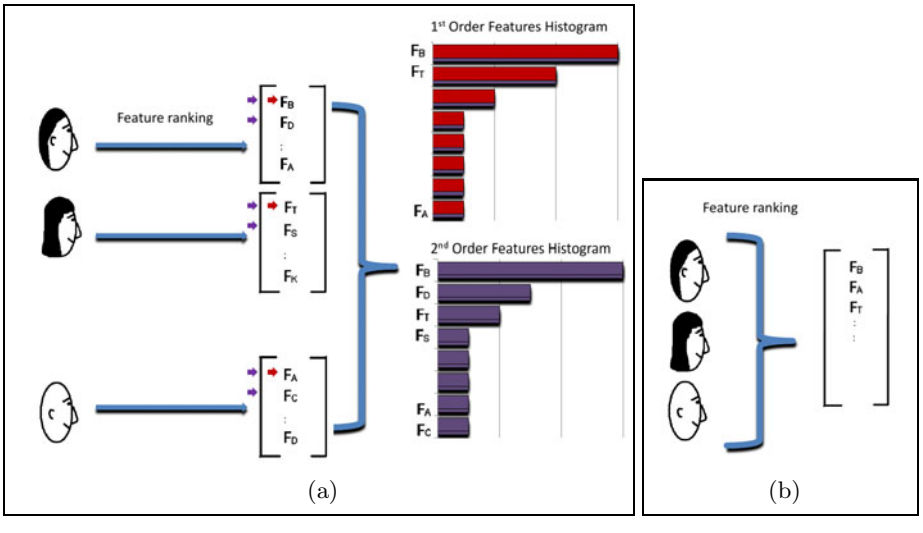

Fig. 1. (a) Self-feature concensus. Features of each person in the data is ranked first. Then, we select top n from these ranked list of self-features to construct n^{th} order histogram of feature counts. In this figure, only the 1^{st} and 2^{nd} order histograms are shown. **(b) Group-feature approach.** Features are selected by using all people's observations at once.

Once the ranked lists of self-feature are obtained, we create a consensus over self-features by using only the top n of each list. A concensus is represented by composing an n^{th} order histogram using the top n of each self-feature. This consensus provides a ranking of self-features, and we expect the relevant features to be replicated in these histograms. To remove possible outliers, we apply a threshold on the concensus features to keep only a subset of relevant features. The intuition behind this threshold is that the relevant features are expected to appear frequently in top n of many self-features corresponding to different people, whereas the outlier features would not appear that as often. The minimum required concensus threshold has been selected to be $n+1$ for an n^{th} order histogram in our experiments. Figure 1(a) shows two consensus examples: first and second order histograms.

4 Sparse Ranking

Our feature ranking scheme relies on sparse regularization that applies some constraints on model parameters during training. For a better understanding, we first describe the Conditional Random Fields model used in our experiments and then show how sparse regularization enable feature ranking in a non-greedy manner.

4.1 Conditional Random Fields

Conditional Random Field (CRF) [24] is a probabilistic discriminative model for sequential data labeling. It is an undirected graphical model that defines a single log-linear distribution over label sequences given a particular observation sequence. CRF learns a mapping between a sequence of multimodal observations $\mathbf{x} = \{x_1, x_2, ..., x_m\}$ and a sequence of labels $\mathbf{y} = \{y_1, y_2, ..., y_m\}$. Each y_j is a class label for the $j^{\texttt{th}}$ frame of a video sequence and is a member of a set $\mathcal{Y}$ of possible class labels, for example, $\mathcal{Y} = \{\texttt{head-nod}, \texttt{other-gesture}\}$. Each frame observation x_j is represented by a feature vector $\phi(x_j) \in \mathbf{R}^d$, for example, the prosodic features at each sample.

Given the above definitions, the conditional probability of y is defined as follows:

$$P(\mathbf{y} \mid \mathbf{x}, \theta) = \frac{1}{Z(x)} exp(\sum_{\alpha} \theta_\alpha F_\alpha(\mathbf{y}, \mathbf{x})) \tag{1}$$

where θ is a vector of linear weights, and $Z(x)$ is a normalization factor over all possible states of $\mathbf{y}$. Feature function F_α is either a state function $s_k(y_j, \mathbf{x}, j)$ or a transition function $t_k(y_{j-1}, y_j, \mathbf{x}, j)$. State function s_k depends on the correlation between label at position j and the observation sequence; while transition function t_k depends on the entire observation sequence and the labels at positions i and i-1 in the label sequence.

Given a training set consisting of m labeled sequences $(\mathbf{x_i}, \mathbf{y_i})$ for $i = 1...m$, training of conditional random fields involves finding the optimum parameter set, θ, that maximizes the following objective function:

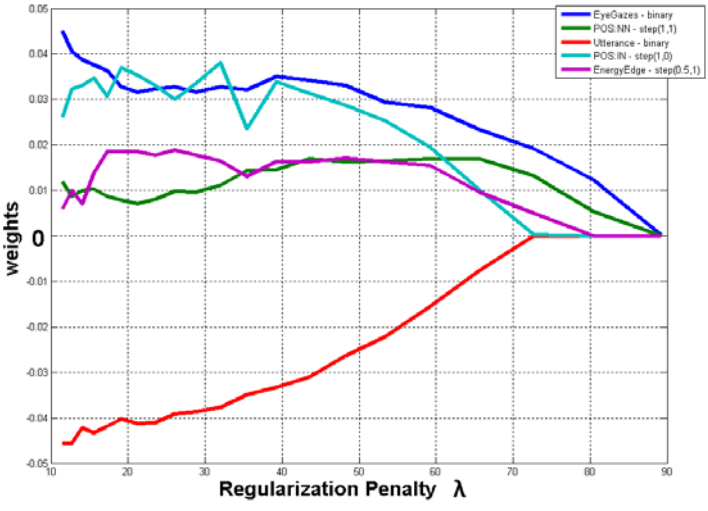

Fig. 2. Sparse ranking using regularization path. As λ goes from higher to lower values, feature weights start to become non-zero based on their relevance to the model.

$$L(\theta) = \sum_{i=1}^{m} \log P(\mathbf{y}_i \mid \mathbf{x}_i, \theta) \tag{2}$$

which is the conditional likelihood of the observation sequence.

4.2 The Method

Our method exploits regularization technique which provides smoothing when the number of learned parameters is very high compared to size of available data. Using a regularization term in the optimization function during training can be seen as assuming a prior distribution over the model parameter. The two most commonly used priors are Gaussian(L_2 regularizer) and Exponential (L_1 regularizer) priors. A Gaussian prior assumes that each model parameter is drawn independently from a Gaussian distribution and penalizes according to the weighted square of the model parameters. An Exponential prior penalizes according to the weighted L_1 norm of the parameters and is defined as follows:

$$R(\theta) = \lambda \parallel \theta \parallel_1 = \lambda \sum_i \mid \theta_i \mid \tag{3}$$

where θ is the model parameters and $\lambda > 0$. In training of conditional random fields, this regularization term is added as a penalty in the log-likelihood function that is optimized. Therefore, Equation 2 becomes:

$$L(\theta) = \sum_{i=1}^{m} \log P(\mathbf{y}_i \mid \mathbf{x}_i, \theta) - R(\theta) \tag{4}$$

L_1 regularization results in sparse parameters vector in which many of the parameters are exactly zero [25]. Therefore, it has been widely used in different domains for the purpose of feature selection [22] [9]. The λ in Equation 3 determines how much penalty should be applied by the regularization term. Larger values indicate larger penalty, thus produces sparser vector parameters.

Figure 2 shows the effect of regularization on feature weights. This regularization path was created by starting with a high regularization penalty λ where all the features are zero and then gradually reduce the regularization until all the features have non-zero values. In this path, if a feature becomes non-zero in earlier stages (i.e., large λ), this signifies that it is an important feature. Our ranking scheme is based on this observation. We rank the features in the order of it's becoming non-zero in the regularization path. The pseudo code for our algorithm is as follows:

$ranked\_features = empty$
for $\lambda = \infty$ down to 0 **do**
  train a CRF with current λ
  for all nonzero feature params θ_i **do**
    if θ_i is NOT in selected_features **then**
      $ranked\_features = selected\_features + \theta_i$
    end if
  end for
end for
return $ranked\_features$

5 Experiments

We test the validity of our approach on the multimodal task of predicting listener nonverbal backchannel (i.e., listener head-nods). Backchannel feedback prediction has received considerable interest due to its pervasiveness across languages and conversational contexts [11] [10].

5.1 The Data

We are use the RAPPORT dataset [11] that contains 42 dyadic interactions between a speaker and a listener. Data is drawn from a study of face-to-face narrative discourse ('quasi-monologic' storytelling). In this dataset, participants in groups of two were told they were participating in a study to evaluate a communicative technology. Subjects were randomly assigned the role of speaker and listener. The speaker viewed a short segment of a video clip taken from the Edge Training Systems, Inc. Sexual Harassment Awareness video. After the speaker finished viewing the video, the listener was led back into the computer room, where the speaker was instructed to retell the stories portrayed in the clips to the listener. The listener was asked to not talk during the story retelling. Elicited stories were approximately two minutes in length on average. Participants sat approximately 8 feet apart.

5.2 Multimodal Features and Encodings

We use four different type of multimodal features in our models: prosodic, lexical, part-of-speech, and visual gesture features. **Prosody** refers to the rhythm, pitch and intonation of speech. Several studies have demonstrated that listener feedback is correlated with a speaker's prosody [17]. Listener feedback often follows speaker pauses or filled pauses such as "um" (see [19]). We encode the following prosodic features, including standard linguistic annotations and the prosodic features suggested by Ward and Tsukhara [18]:

– Downslopes in pitch continuing for at least 40ms; regions of pitch lower than the 26th percentile continuing for at least 110ms (i.e., lowness); drop or rise in energy of speech (i.e., energy edge); fast drop or rise in energy of speech (i.e., energy fast edge), vowel volume (i.e., vowels are usually spoken softer), pause in speech (i.e., no speech).

Gestures performed by the speaker are often correlated with listener feedback [26]. Eye gaze, in particular, has often been implicated as eliciting listener feedback. Thus, we encode speaker looking at the listener as our **visual gesture** feature.

Some studies have suggested an association between **lexical** features and listener feedback [19]. Therefore, we include top 100 individual words (i.e., unigrams) that are selected based on their frequency in the data.

Finally, we attempt to capture syntactic information that may provide relevant cues by extracting four types of features from a syntactic dependency structure corresponding to the utterance. Using a part-of-speech tagger [27], we extract the part-of-speech tags for each word (e.g. noun, verb, etc.) as our **Part-of-speech(POS)** features.

We encode our features using 13 different encoding templates as introduced by [10]. The purpose of this encoding dictionary is to capture different relationships between speaker features and listener backchannels. For instance, listener backchannels sometimes happen later after speaker features, or when the speaker features are present for certain amounts of time and its influence may not be constant over time. To automatically obtain these relations, we use three encoding templates in our experiments: **binary encoding** that is designed for speaker features which influence on listener backchannel is constraint to the duration of the speaker feature, **step function** that is a version of binary encoding with two additional parameters: width of the encoded feature and delay between the start of the feature and its encoded version. and **ramp function** that linearly decreases for a set period of time (width parameter). Step and ramp functions are used with 6 different parameters(width and delay): (0.5 0.0), (1.0 0.0), (0.5 0.5), (1.0 0.5), (0.5 1.0), (1.0 1.0) for step, and (0.5 1.0), (1.0 1.0), (2.0 1.0), (0.5 0), (1.0 0), (2.0 0) for ramp.

5.3 Methodology

We performed hold-out testing by randomly selecting a subset of 10 interactions (out of 42) for the test set. The training set contains the remaining 32 dyadic

interactions. All models evaluated in this paper were trained with the same training and test sets. The test set does not contain individuals from the training set. Validation of model parameters was performed using a 3-fold strategy on the training set. For L_1 regularization, λ ranged $1000 * 0.95^k, k = [20, 22..170]$. For L_2 regularization, the validated range was $10^k, k = [-3..3]$. The training of CRF models was done using the hCRF library [28].

The performance is measured by using the F-measure, which is the weighted harmonic mean of precision and recall. Precision is the probability that predicted backchannels correspond to actual listener behavior. Recall is the probability that a backchannel produced by a listener in our test set was predicted by the model. We use the same weight for both precision and recall, so called F_1. During validation we find all the peaks (i.e., local maxima) from the marginal probabilities. These backchannel hypotheses are filtered using the optimal threshold from the validation set. A backchannel (i.e., head-nod) is predicted correctly if a peak happens during an actual listener backchannel with high enough probability.

5.4 Results

We ran four experiments: (1) group-feature approach with sparse ranking, (2) effect of the order parameter on self-feature concensus, (3) analysis of selected self-features and (4) comparison of self-feature concensus to group-feature approach.

For the first experiment, we apply our sparse ranking scheme using all the training people in a group-feature manner. To show the effect of sparse ranking, we train a separate CRF for each subset of group-features. For comparison, we trained one CRF using all features (1833 encoded features). All CRFs were retrained using L_2 regularization following previous work on CRF-based backchannel prediction [10]. (L_1 was still used during the sparse ranking step).

Table 1. Group-features with sparse ranking. We incrementally add features as they appear in the regularization path and use for retraining. Each row shows the features added at that stage, therefore the model at this stage is retrained with these new features plus the features above it. Final row shows values for using all the features instead of feature selection.

Features	**Precision**	**Recall**	F_1
EyeGazes-binary	0.16469	0.14164	0.1523
... + POS:NN-step(1,.5) **... + VowelVolume**-step(.5,1)	0.15281	0.25903	0.19222
... + Pause-step(1,0) **... + Lowness**-step(1,.5)	0.19818	0.37516	**0.25935**
... + POS:NN-step(1,1)	0.2002	0.1918	0.19591
... + Lowness-step(1,0) **... + VowelVolume**-step(.5,.5)	0.20512	0.1943	0.19956
Baseline: All features *No feature selection*	0.1643	0.6079	**0.2587**

Table 2. Selected features with self-feature concensus using histograms of different orders (after outlier rejection)

$1^{st}Order$	$2^{nd}Order$	$3^{rd}Order$
POS:NN-step(1,1)	**POS:NN**-step(1,1)	**POS:NN**-step(1,1)
Utterence-binary	**POS:NN**-step(1,.5)	**POS:NN**-step(1,.5)
EyeGaze-binary	**Utterence**-binary	**Utterence**-binary
Pause-binary	**EyeGaze**-binary	**EyeGaze**-binary
POS:DT-step(1,.5)	**EyeGaze**-step(1,1)	**Pause**-step(1,0)
Lowness-step(1,0)	**Pause**-binary	**POS:DT**-step(1,.5)
	Pause-step(1,0)	**Lowness**-step(1,0)
	POS:DT-step(1,.5)	**Lowness**-step(1,.5)
	Lowness-step(1,0)	

Precision, recall and F_1 values are given in Table 1. In each row, features are added as they appear in the L_1 regularization path of our sparse ranking scheme. The best performance happens in the third step with five selected features and F_1 value of 0.25935. The last row of Table 1 represents the performance when no feature selection is applied (all features are used). This result shows that sparse ranking can find a subset relevant of features, with performance similar to the baseline model that contain all features.

For the same listener backchannel prediction task, Morency et al. [10] used a greedy-forward feature selection method on the RAPPORT dataset. Although, the experimental set up was slightly different (i.e. different test and train sets were used), the best precision, recall and F_1 values archived with this method were 0.1862, 0.4106, 0.2236, respectively.

Our second experiment studies the effect of the order parameter on self-feature concensus. We constructed feature histograms with orders 1, 2, and 3 by looking at the top 1^{st}, 2^{nd}, and 3^{rd} features in each list. Then, we applied a threshold of 2, 3, and 4 respectively on the histograms for outlier rejection. The list of features for each order is listed in Table 2. This result is really interesting since the same features appear in all three consensus.

For our third experiment, we analyze the features selected for our task of head-nod prediction. It is interesting that some features are selected by both self-feature concensus and group-feature approach, such as *Pause, EyeGaze, Lowness, POS:NN. Utterance* and *POS:DT* are the two features selected by self-feature concensus approach that do not appear in the top 20 features from the group-feature approach. *POS:DT* refers to determiners in language, such as *the, this, that. Utterance* refers to the beginning of an utterance. Mixed together, these two features represent moments where the speaker starts an utterance with a determiner. To show the relative importance of the *Utterance* and *POS:DT* features, we added these two features to the list of features obtained by group-feature approach and trained a new CRF model. Precision, recall and F_1 values are 0.21685, 0.38653, 0.27783, respectively. We see an improvement over group-feature approach, showing the importance of self-feature concensus.

Table 3. Precision, recall and F_1 values of retrained CRFs with group-feature approach and self-feature concensus

Method	**Precision**	**Recall**	F_1
self-feature concensus			
Order 1	0.2192	0.4939	0.3037
Order 2	0.23802	0.48628	**0.3196**
Order 3	0.24449	0.28211	0.26196
group-feature approach	0.19818	0.37516	0.25935
Baseline: all features	0.1643	0.6079	0.2587

Our last experiment compares our self-feature concensus approach to the typical group-feature approach. Using the selected self-features from Table 2, we retrained a L_2 regularized CRFs over all training instances. Precision and recall values for these retrained CRFs of self-feature concensus and group-feature approach (best result from first experiment) are given in Table 3. The best F_1 value achieved with 2^{nd} order histogram is 0.3196. Also, all three self-feature concensus models perform better F_1 than the group-feature approach and the CRF trained with all features (i.e., no feature selection). This results show that using self-features improves listener backchannel prediction.

6 Conclusion

Nonverbal behaviors play an important role in human social interactions and a key challenge is to build computational models for understanding and analyzing this communication dynamic. In this paper, we proposed a framework for finding the important features involved in human nonverbal communication. Our self-feature concensus approach first looks at important behaviors for each individual before building a consensus. It avoids the problem with the group-feature approach which focused on the average model and oversees the inherent behavioral differences among people. We proposed a feature ranking scheme exploiting from L_1 regularization technique. This scheme relies on the fact that adding more penalty on the model parameters will result in sparser results in which only the important features will be promoted.

Our framework was tested on the task of listener head-nod prediction in dyadic interactions. We used the RAPPORT dataset that contains 42 dyadic communications between a speaker and a listener. The results are promising and provides improvement over traditional group-feature approach. In our future work, we plan to use this framework for different prediction tasks, such as gaze aversion and turn-taking prediction.

References

1. Drolet, A.L., Morris, M.W.: Rapport in conflict resolution: Accounting for how face-to-face contact fosters mutual cooperation in mixed-motive conflicts. Journal of Experimental Social Psychology 36, 26–50 (2000)

2. Tsui, P., Schultz, G.: Failure of rapport: Why psychotheraputic engagement fails in the treatment of asian clients. American Journal of Orthopsychiatry 55, 561–569 (1985)
3. Fuchs, D.: Examiner familiarity effects on test performance: Implications for training and practice. Topics in Early Childhood Special Education 7, 90–104 (1987)
4. Vinciarelli, A., Pantic, M., Bourlard, H.: Social signal processing: Survey of an emerging domain. Image and Vision Computing Journal 26, 1743–1759 (2009)
5. Matsumoto, D.: Culture and Nonverbal Behavior. In: The Sage Handbook of Nonverbal Communication. Sage Publications Inc., Thousand Oaks (2006)
6. Linda, L., Carli, S.J.L., Loeber, C.C.: Nonverbal behavior, gender, and influence. Journal of Personality and Social Psychology 68, 1030–1041 (1995)
7. Ng, A.Y.: Feature selection, l-1 vs. l-2 regularization, and rotational invariance. In: International Conference on Machine Learning (2004)
8. Smith, A., Osborne, M.: Regularisation techniques for conditional random fields: Parameterised versus parameter-free. In: Dale, R., Wong, K.-F., Su, J., Kwong, O.Y. (eds.) IJCNLP 2005. LNCS (LNAI), vol. 3651, pp. 896–907. Springer, Heidelberg (2005)
9. Vail, D.L.: Feature selection in conditional random fields for activity recognition. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (2007)
10. Morency, L.P., de Kok, I., Gratch, J.: Predicting listener backchannels: A probabilistic multimodal approach. In: Conference on Intelligent Virutal Agents, pp. 243–255 (2008)
11. Maatman, M., Gratch, J., Marsella, S.: Natural behavior of a listening agent. In: Panayiotopoulos, T., Gratch, J., Aylett, R.S., Ballin, D., Olivier, P., Rist, T. (eds.) IVA 2005. LNCS (LNAI), vol. 3661, pp. 25–36. Springer, Heidelberg (2005)
12. Albrecht, K.: Social intelligence: The new science of success. John Wiley and Sons Ltd., Chichester (2005)
13. Thorndike, E.L.: Intelligence and its use. Harpers Magazine 140, 227–235 (1920)
14. Vinciarelli, A., Pantic, M., Bourlard, H., Pentland, A.: Social signal processing: State-of-the-art and future perspectives of an emerging domain. In: 16th ACM International Conference on Multimedia (2008)
15. Curhan, J.R., Pentland, A.: Thin slices of negotiation: Predicting outcomes from conversational dynamics within the first five minutes (2007)
16. Morency, L.P., de Kok, I., Gratch, J.: Context-based recognition during human interactions: Automatic feature selection and encoding dictionary. In: 10th International Conference on Multimodal Interfaces (2008)
17. Nishimura, R., Kitaoka, N., Nakagawa, S.: A spoken dialog system for chat-like conversations considering response timing. In: Matoušek, V., Mautner, P. (eds.) TSD 2007. LNCS (LNAI), vol. 4629, pp. 599–606. Springer, Heidelberg (2007)
18. Ward, N., Tsukahara, W.: Prosodic features which cue back-channel responses in english and japanese. Journal of Pragmatics 23, 1177–1207 (2000)
19. Cathcart, N., Carletta, J., Klein, E.: A shallow model of backchannel continuers in spoken dialogue. In: European Chapter of the Association for Computational Linguistics, pp. 51–58 (2003)
20. Perkins, S., Lacker, K., Theiler, J., Guyon, I., Elisseeff, A.: Grafting: Fast, incremental feature selection by gradient descent in function space. Journal of Machine Learning Research 3, 1333–1356 (2003)
21. Riezler, S., Vasserman, A.: Incremental feature selection and l1 regularization for relaxed maximum-entropy modeling. In: Conference on Empirical Methods on Natural Language Processing, pp. 174–181 (2004)

22. McCallum, A.R.C.: Efficiently inducing features of conditional random fields. In: 19th Conference on Uncertainty in Artificial Intelligence (2003)
23. Gallaher, P.E.: Individual differences in nonverbal behavior: Dimensions of style. Journal of Personality and Social Psychology 63, 133–145 (1992)
24. Lafferty, J., McCallum, A., Pereira, F.: Conditional random fields: Probabilistic models for segmenting and labelling sequence data. In: International Conference on Machine Learning (2001)
25. Tibshirani, R.: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B 58, 267–288 (1994)
26. Burgoon, J.K., Stern, L.A., Dillman, L.: Interpersonal Adaptation: Dyadic Interaction Patterns. Cambridge University Press, Cambridge (1995)
27. Sagae, K., Tsujii, J.: Dependency parsing and domain adaptation with LR models and parser ensembles. In: Proceedings of the CoNLL Shared Task Session of EMNLP-CoNLL, pp. 1044–1050. Association for Computational Linguistics (2007)
28. hCRF library (2007), `http://sourceforge.net/projects/hcrf/`