

# Applying Grammar Inference To Identify Generalized Patterns of Facial Expressions of Pain

**Ute Schmid**

**Michael Siebers**

**Dominik Seuß**

*Faculty Information Systems and Applied Computer Science, University of Bamberg*

UTE.SCHMID@UNI-BAMBERG.DE

MICHAEL.SIEBERS@UNI-BAMBERG.DE

DOMINIK.SEUSS@STUD.UNI-BAMBERG.DE

**Miriam Kunz**

**Stefan Lautenbacher**

*Faculty Humanities, Institute of Psychology, University of Bamberg*

MIRIAM.KUNZ@UNI-BAMBERG.DE

STEFAN.LAUTENBACHER@UNI-BAMBERG.DE

**Editors:** Jeffrey Heinz, Colin de la Higuera and Tim Oates

## Abstract

We present an application of grammar inference in the domain of facial expression analysis. Typically, only *sets* of AUs which occur in a given time frame are used for expression analysis, the *sequence* in which these AUs occur is ignored. We wanted to explore whether the structural patterns of AU appearances contain diagnostically relevant information. We applied alignment-based learning (ABL) to data of facial expressions of pain collected in a psychological study. To evaluate the quality of the induced grammars we applied cross-validation to estimate the generalization error. We can show that the learned grammars have reasonably high coverages for unseen pain sequences. However, the number of rules of the learned grammars cannot be reduced substantially without loss of generalization.

## 1. Introduction

We want to explore an application of grammatical inference (GI) for structural pattern recognition in the domain of facial expression analysis (Fasel and Luetttin, 2003). To our knowledge, this is the first time that the applicability of GI in this domain is explored. One approach to expression analysis is based on the Facial Action Coding System (FACS, Ekman and Friesen, 1978) – an anatomically based system defined over 43 different facial movements, the so called action units (AUs). For manual as well as for automated classification the usual procedure is to identify which AUs occur in which intensities in a given time frame, e.g. 5 seconds. That is, the current practice is to identify the *set* of AUs presently active in a face to assign the underlying mental state. However, there might be diagnostically relevant information in the *sequence* in which the AUs appear. Knowledge about empirically valid sequences of AUs might also contribute for more realistic emotion generation in avatars and humanoid robots.

To explore whether AUs appear in specific sequences, we analyzed data which were obtained in a psychological study of facial expression of pain in patients with dementia. Subjects were 42 demented patients (22 female, 20 male), 54 age-matched healthy controls (43 female, 11 male), and 28 young subjects (16 female, 12 male) which were exposed to

Table 1: Number of subjects ( $\#sub$ ), of sequences ( $\#seq$ ), of occurring AUs and AU compounds ( $\#AUs$ ), mean, minimum and maximum sequence lengths ( $\bar{l}$ ,  $min_l$ ,  $max_l$ ), mean, minimum and maximum number of different terminals per sequence ( $\bar{t}$ ,  $min_t$ ,  $max_t$ ) for pain episodes with high pressure stimulation.

Group	$\#sub$	$\#seq$	$\#AUs$	$l$	$min_l$	$max_l$	$\bar{t}$	$min_t$	$max_t$
Overall	86	347	76	4.03	1.00	17.00	3.54	1.00	13.00
Demented	31	142	58	4.70	1.00	16.00	4.19	1.00	13.00
Healthy	55	205	53	3.57	1.00	17.00	3.09	1.00	13.00
Old	31	123	44	3.76	1.00	14.00	3.21	1.00	13.00
Young	24	82	33	3.28	1.00	17.00	2.91	1.00	11.00

mechanically induced pain of various intensities. The details of the experimental setting are reported in [Kunz et al. \(2007\)](#).

The facial responses were analysed by a trained FACS coder and the identified AUs were entered in a database with code, onset-time, offset-time and intensity rating. For a first exploration, we only considered the onset time of each AU, that is, we did not consider time intervals. AUs which appeared together in the same frame of the video recordings were treated as self-contained symbols. Overall there occurred 76 different AUs and AU compounds which constitute the alphabet of the to be induced grammar. Facial responses have a relatively high threshold and mostly only occur at moderate to high pain intensities. Therefore, we only considered sequences which occurred during high pressure. The numbers of sequences for the different groups together with average lengths and numbers of different terminals are given in Table 1. Some additional details of the data are given in appendix A.

## 2. Application of ABL to Sequences of Facial Expressions

ABL ([van Zaanen, 2002](#)) realises unsupervised learning of context-free grammars by aligning sequences. Since ABL learns grammars which cover all presented sequences, learning might result in grammars with a high number of rules, many of which might cover (sub-)sequences which occur only once in a sample. Therefore, we defined a *heuristics for rule elimination* which should significantly reduce the number of grammar rules while obtaining high precision. We enriched the learned grammars with relative expansion frequencies. Then, for each sequence the relative expansion frequency can be calculated by multiplying the relative expansion frequencies of all involved rewrite rules.

While in standard machine learning there is an established approach for evaluating the quality of learned hypotheses, this is not true for unsupervised empirical grammar inference. One method proposed by [van Zaanen and Geertzen \(2008\)](#) is to measure the ability of the learned grammar to classify sequences based on language membership. The main drawback of this method is that results depend on how sequences are selected which are assumed to be negative examples.

In the context of supervised learning, precision and recall of a learned hypothesis can be estimated by  $k$ -fold cross-validation. We propose to apply *cross-validation* to evaluate the quality of pain grammars learned with ABL. Because there are only positive instances,

we will use coverage (precision) as a measure for quality. Furthermore, we propose to apply cross-validation in two different ways: First, the standard method can be applied to estimate precision by obtaining the average coverage of induced grammars over  $k$  folds. Second, we want to investigate whether a learned grammar can be reduced using the heuristic rule elimination method. This can be realized by learning  $k$  grammars for  $k$  sub-sets of training data and reducing the rule set of each grammar by heuristic rule elimination using successively higher thresholds for relative frequencies. In contrast to standard cross-validation, the coverage of the reduced grammars is assessed for the *complete set* of sequences (in contrast to the  $k$ -th subset).

To assess coverage in the standard cross-validation approach, it is necessary to generate acceptor automata for the learned grammars. Given an acceptor, we can count the number of sequences in the test set covered by the grammar. To assess coverage in the rule reduction approach, the learned grammars are used to generate all sequences for different threshold values and obtain the amount of sequences in the test set included in the generated sequences. Coverage is calculated as  $\frac{|S_i \cap S_T|}{|S_T|}$  with  $S_i$  ( $i = 1, \dots, k$ ) as sequences generated by grammar  $i$  and  $S_T$  as sequences in the general test set, that is, the number of all sequences.

### 3. Results

For cross-validation, the set of sequences was randomly divided into 5 batches of approximately equal size. We controlled for equal distribution with respect to membership to subgroups (i.e. young, old, healthy etc.) when constructing the batches. For the resulting batches, the number of different AUs and the lengths of sequences have comparable distributions (see appendix C). To estimate precision, we evaluated how good the grammar learned from four batches is able to classify sequences from the remaining test batch. Since we learn from positive data, classification is operationalized as the number of unseen sequences accepted by a learned grammar. We constructed an acceptor for the learned grammars using DParser<sup>1</sup>. Results show that the learned grammars generalize reasonably well with an average precision of 0.65 (see also appendix C).

Application of ABL to the empirically found sequences of facial expressions of pain resulted in quite complex grammars: The grammar characterizing all sequences consists of 1314 rules and gives 225 alternatives in the starting rule (see appendix B for an excerpt). If a coverage of about 70% is acceptable then 1233 rules (93.8%) are left for the complete sample of subjects. The coverage rate is initially dropping faster than the number of rules with increasing threshold. The rules decrease faster when the coverage is around 50%. This means that, unfortunately, for our set of data it is not possible to reduce the grammar without loss of precision.

A follow-up hypothesis was that ideosyncraties are specific for sub-groups of subjects, such as young/old or healthy/demented. Because the number of sequences in the subgroups is small, we did not apply cross-validation. For the sub-groups, the grammars consist of smaller numbers of rules (see Tab. 2) – mainly due to the smaller number of sequences ( $r = .981$  between number of sequences and number of grammar rules). We applied heuristic rule reduction to test whether sub-groups show more coherent behavior than the overall

---

1. <http://dparser.sourceforge.net/>

group. However, the results given in Table 2 show that it is not possible to remove rules without loss of precision.

Table 2: Overview of induced grammars with total number of rules ( $\#total$ ), number of expansions for the starting symbol ( $\#top$ ), and number of rules ( $\#rules$ ) and coverage rate in percent (cov.%) for different thresholds.

Group	$\#total$	$\#top$	$10^{-7}$		$10^{-5}$		$10^{-4}$	
			$\#rules$	cov.%	$\#rules$	cov.%	$\#rules$	cov.%
Overall	1314	225	1308	79.54	909	55.91	387	43.80
Demented	611	116	611	91.55	564	62.68	359	43.66
Healthy	681	138	681	94.15	622	70.73	353	53.65
Old	418	98	418	97.56	414	80.49	314	59.35
Young	232	51	232	100.00	232	93.90	219	74.39

#### 4. Conclusion

Our initial exploration whether sequences of expressions of pain convey diagnostically relevant information provided promising results. That is, using ABL as an out of the box grammar inference system, we could show that grammars which describe the general structure of pain sequences do generalize reasonably well to unseen pain sequences. To validate our findings we currently prepare an empirical study with a face avatar which will be animated with grammatical and agrammatical sequences of action units for pain. We are hoping that human subjects will show shorter performance times and smaller performance errors when classifying grammatical in contrast to agrammatical AU sequences. Furthermore, using ABL is only one possible approach to learn generalized sequences. Currently we are working on a swarm-algorithm based approach to grammar induction. Other fruitful approaches might be Hidden Markov Models or approaches to learning hybrid grammars might be tested.

#### Acknowledgments

This study was supported by the Dr. Werner Jackstädt-Stiftung. Thanks to Jirka Lewandowski who conducted some initial analyses. Special thanks to Menno van Zaanen who willingly answered our questions about ABL. Thanks also to the anonymous reviewers for very helpful suggestions.

#### References

- Ekman, P. and Friesen, W. V. (1978). *The Facial Action Encoding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, Palo Alto, CA.
- Fasel, B. and Luetttin, J. (2003). Automatic facial expression analysis: a survey. *Pattern Recognition*, 36(1):259–275.
- Kunz, M., Scharmann, S., Hemmeter, U., Schepelmann, K., and Lautenbacher, S. (2007). The facial expression of pain in patients with dementia. *Pain*, 133:221–228.

van Zaanen, M. (2002). *Bootstrapping Structure into Language: Alignment-Based Learning*. PhD thesis, University of Leeds, UK.

van Zaanen, M. and Geertzen, J. (2008). Problems with evaluation of unsupervised empirical grammatical inference systems. In ICGI 2008, volume 5278 of *LNCS*, pages 301–303. Springer.

## Appendix A. Results of the Psychological Study and Data Description

Kunz et al. (2007) identified which AUs occurred in at least 5% of the exposures to pain (see Tab. 3). A main result of the study was that the set of pain relevant AUs was comparable for demented and healthy subjects.

Table 3: AUs with a critical occurrence of more than 5% in denominated pain segments in healthy controls and demented patients (abbreviated from Table 1 in Kunz et al., 2007).

AU	Description	AU	Description
AU1, AU2	brow raiser	AU4	brow lower
AU6, AU7	orbit tightening	AU9, AU10	levator contraction
AU17	chin raise	AU25, AU26, AU27	mouth opening
AU45	eye blink		

For sequence generalization learning, we kept all AUs appearing in a sequence, including such AUs which occurred less than 5% over all pain episodes. Overall there occurred 76 different AUs and AU compounds which constitute the alphabet of the to be induced grammar (sorted by frequency of occurrence):

$$\Sigma = \{\text{au7, au4, au25, au6, au26, au9, au18, au10, au17, au14, au12, au43, au1, au25-26, au6-7, au2, au1-2, au20, au24, au32, au16, au4-7, au5, au23, au6-7-9, au15, au38, au19, au6-9, au4-9, au19-25-26, au7-10, au6-7-10, au17-24, au4-7-9, au4-6-7, au30, au28, au6-12, au6-10, au7-9, au4-6-7-10, au9-25, au4-43, au31, au45, au4-6-7-9}\} \cup 29 \text{ further AUs and compounds occurring only once over all pain episodes.}$$

Some example sequences are given in Table 4.

Table 4: Examples for AU sequences over alphabet  $\Sigma$ .

au7
au6-7
au7 au4 au7 au9
au12 au19-25-26 au19-25-26 au28 au17 au4 au19-25-26
au6-7-9 au43 au4 au1-2 au7 au12 au43 au4 au9 au7 au1-2

## Appendix B. Extract from an Induced Pain Grammar

A → CI DB .012 | au4-7-9 .003 | au6 FAB au25 BJC GGB BF .003 | au25 .009 |  
 au4 au4 ED .003 | au4 F .003 | au2 EJ .003 | au6 EB .003 |  
 au7 GEB GGB DFB .003 | ... 215 more

CI → au25 .071 | au1-2 .036 | au14 .036 | au7 .393 | au12 .107 | au17 .036 | D DI .107 |  
 J I .036 | au4 .036 | au4-7 .036 | au7 I .036 | EFC I .036 | HJ I .036

EJ → IGB F .125 | au5 .125 | au7 .25 | au4-7-10 .125 | BFD DB .125 |  
 au7 DB .125 | DDC DCC .125

DB → au25 0.091 | JEB F .018 | au1-2 .036 | au14 .018 | au5 .018 | au20 .018 | au7 .055 |  
 JEB FH .018 | au12 .018 | JEB ED .018 | au4 F .018 | au6 .127 | au4 .218 | au1 .018 |  
 au4-9 .018 | au9 .109 | au18 .055 | au16 .018 | au30 .018 | au10 .055 au17 .036 |

... 423 more

Figure 1: Excerpt of the grammar induced for the complete set of sequences.

## Appendix C. Details of the Empirical Evaluation

Table 5: Description of the random samples for 5-fold cross-validation with number of subjects (#sub), number of sequences (#seq), mean, minimum and maximum sequence lengths ( $\bar{l}$ ,  $\min_l$ ,  $\max_l$ ), mean, minimum and maximum number of different terminals (AUs) ( $\bar{t}$ ,  $\min_t$ ,  $\max_t$ ), and total number of AUs ( $t$ ).

Group	#sub	#seq	$\bar{l}$	$\min_l$	$\max_l$	$\bar{t}$	$\min_t$	$\max_t$	$t$
overall	86	347	4.03	1.00	17.00	3.54	1.00	13.00	76
batch1	47	70	4.61	1.00	14.00	3.94	1.00	13.00	39
batch2	46	70	3.94	1.00	16.00	3.51	1.00	13.00	43
batch3	46	69	3.59	1.00	13.00	3.22	1.00	10.00	38
batch4	44	69	4.10	1.00	17.00	3.55	1.00	11.00	42
batch5	43	69	3.90	1.00	13.00	3.48	1.00	9.00	38

Table 6: Precision for the grammars learned by 5-fold cross-validation

1	2	3	4	5	avg
0.6857143	0.6285714	0.6376812	0.6666667	0.6376812	0.6512968