Categorising Behavioural States of People with Profound Intellectual and Multiple Disabilities

Matej Cigale
Jožef Stefan Institute, Jožef Stefan International Postgraduate School
Jamova cesta 39
1000 Ljubljana
matej.cigale@ijs.si

Gašper Slapničar
Jožef Stefan Institute
Jamova cesta 39
1000 Ljubljana
gasper.slapnicar@ijs.si

Mitja Luštrek
Jožef Stefan Institute, Jožef Stefan International Postgraduate School
Jamova cesta 39
1000 Ljubljana
mitja.lustrek@ijs.si

ABSTRACT

People with profound intellectual and multiple disabilities (PIMD) are a diverse group of individuals. They face extreme difficulties in communicating to the outside world. This paper presents two specialised machine learning (ML) methods that attempt to classify the behavioural states and communication attempts of people with PIMD based on annotations of non-verbal signals (NVS) and expert knowledge. The first is based on the idea of unique NVS that classify the behavioural state, e.g., a smile in happy healthy individuals. The second uses the arousal-valence model as a scaffold to generate a value for valance based on a group of NVS.

Keywords

Behavioural state, Communication, PIMD, Prolog, Intent detection

1. INTRODUCTION

People with profound intellectual and multiple disabilities (PIMD) are a heterogeneous group, suffering from different ailments and conditions, making them face extreme difficulties in their daily lives. Severe cognitive, motor and sensory disabilities makes this population reliant on outside care for most daily tasks, and thus extremely vulnerable. While these individuals are exactly the ones that would benefit most from intelligent systems in their vicinity, they are unable to use them due to relative high complexity. The main problem is their lack of symbolic communication – they are unable to express their desires in a consistent manner.

Most assistive technology relies on some form of symbolic communication, which makes it unusable for people with PIMD. The INSENSION project aims to develop a system that will observe behavioural state and non-symbolic communication attempts of people with PIMD and interpret them to people in the vicinity in order to allow them to render assistance or support if needed; and even automatically control their environment using external services. This will in turn help the people with PIMD to have greater agency in their environment and help people who are not as familiar with them to assist them in meting their daily needs.

The first step of the INSENSION system is to recognise non-verbal signals (NVS) expressed by people with PIMD (e.g., certain gestures [1] and facial expressions [7]) and important features of their environment (e.g., presence of a caregiver and objects, temperature). Afterwards, these are interpreted as behavioural states and communication attempts, and provided to a caregiver or external services. This paper deals with the interpretation of NVS once they are recognised by the sensoric and machine vision sub-systems.

Interpretation of NVS of people with PIMD is a challenging task, since each individual is unique with different abilities and signals. Thus, no general-purpose system can be developed and personalised classification methods must be used. Mappings between certain NVS and behavioural states are known to those close to the specific person, and this expert knowledge should be used in the decision making process. Due to no possible generalization, we are also dealing with low amounts of data per subject, as collecting a large set of annotated data for each individual is neither practical nor feasible. Finally, as context is important for interpretation of behavioural states, a database of relevant contexts should be built.

This paper is organised as follows. In Section 2 we present the related work on the subject. Section 3 discusses the data collected so far. Section 4 presents the two ML methods: the Unique Non-Verbal Signals model optimised for extremely small data sets, and the Valence model that works better with limited but somewhat larger data sets. Section 5 looks at the results and discusses current and future work. Finally, Section 6 draws the conclusions based on this paper.

2. RELATED WORK

Since our focus is on recognising the ambiguous feelings, desires and intentions of people with PIMD, which are expressed in unique and unexpected ways, we focus on work dealing with detection and understanding of human feelings. Arousal and valence are the standard metrics that are used to map human feelings onto a 2D plane. Arousal can be understood as the strength of a feeling, while valence is the positive or negative connotation of the feeling. There are several ways to map discrete feelings to this 2D space and the actual mapping is not agreed upon, leading to some ambiguity on this subject, but it is at this point one of the standard models [6].

A step towards understanding feelings that is closest to what INSENSION will use (from video and audio) was done by Metallinou et al. [3]. They use a different space, also includ-

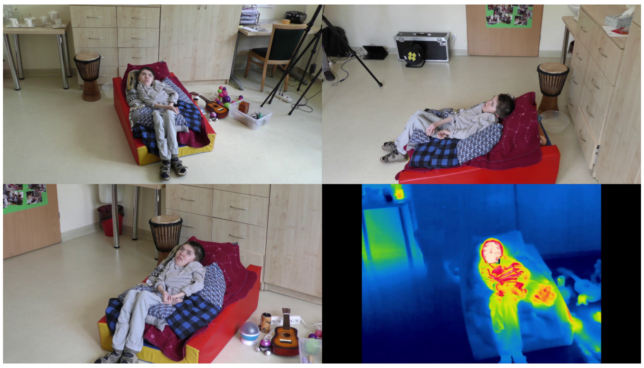

Figure 1: An example of the recorded videos. Computer vision algorithms are ran on all the streams and the results collated based on probability.

ing the dominance dimension. They used USC CreativeIT database consisting of acted-out scenes.

When it comes to extracting the context of the interaction, there are several approaches that produce interesting results. Probabilistic Event Calculus [4] is one of the approaches that can be used and extended to the case at hand. Case-based reasoning (CBR) [2] is another paradigm, in which knowledge is represented as a set of cases – events that happened, and the solutions that were used to solve the problem. Events that are detected are conformed into the closest case that is stored in the database and the solution of the problem is used. The solution is then evaluated and stored in the system based on the success of the solution.

3. DATA

Five PIMD people are currently involved in the INSENSION project. Expert knowledge was collected from their caregivers, who know them well, in the form of an extensive questionnaire. This data was then incorporated into the behavioural state recognition to improve decisions.

Visual data was collected in the facilities where the people with PIMD are cared for, and took the form of multiple-angle recordings with normal and thermal cameras (see Figure 1 for examples). Videos were annotated by hand, using the ELAN [5] software. Annotators were asked to input suitable pre-defined annotations and note any special cases that might play a role in the behavioural state of the subjects. Any state that was not specifically marked was considered *neutral*.

In our experiments the annotations of behavioural states were considered as ground truth, as we feel that people tasked with annotation were familiar enough with their subjects so that they could render as accurate picture of their behavioural state as is possible [8].

4. METHODOLOGY

Our methods assume that the person with PIMD has distinct NVS that correlate to his internal behavioural states. Each of the detected signals can have a meaning, but that is not guaranteed. The NVS can have no meaning or the same NVS can be used to convey multiple dissimilar meanings.

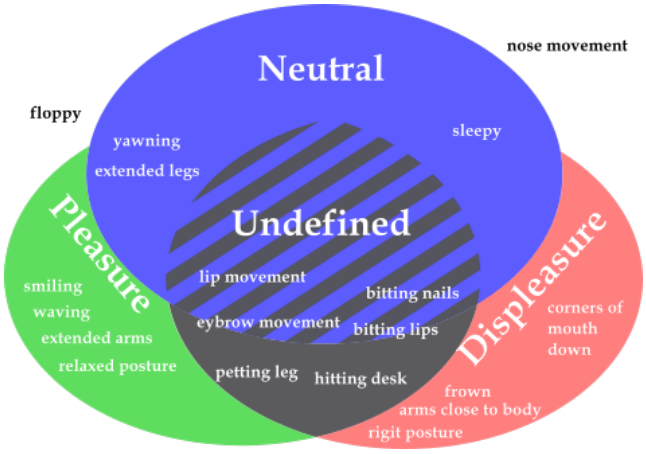

Figure 2: The visualisation of the NVS set interactions.

These signals do not necessary follow social conventions, for instance, lifting the corners of the mouth up can signify pain not pleasure as in normative individuals.

4.1 Classical Machine Learning

Deep architectures are not feasible in our case due to the small amount of available data. In our experiments we segmented the data into 3-second windows, owing to the fact that this was the window size used by annotators. Due to small number of examples leave one out approach was used and several training rounds were used to determine the accuracy of classification. Only present states were used in the classification.

Several methods were tested (nearest neighbors, linear SVM, RBF SVM, Gaussian process, decision tree, random forest, neural net, AdaBoost, naive Bayes, QDA), the decision trees providing the best results. The average classification accuracy for pleasure and displeasure is 63.8%. While the results do not seem bad at first glance, we can do better. We would like to make use of expert knowledge and perhaps even have access to the model and tweak it if the experts say that it does not make sense.

4.2 The Unique Non-Verbal Signals method

The Unique Non-Verbal Signals method encompasses the idea that there exists a NVS that will signify a specific behavioural state, but will never be used to signify any other behavioural state. This means that in order for us to robustly detect, for example, pleasure, we must remove all NVS that are associated with displeasure or neutral state. This leaves us with a set of NVS that uniquely represent the behavioural state of pleasure. Additionally, if experts annotated that a certain NVS corresponds to a behavioural state, we must take that into account. This approach is illustrated in Figure 2.

To decide the state we check if there are NVS specific to pleasure, either from the expert knowledge or from annotated examples (Listing 1). The term *window(Interval, NVS, _)* unifies for any 3-second window. Term *assessment(State, NVS)* will unify if experts noted that a NVS signifies a state.

```
decide_state(Interval, 'Pleasure') :-
    pleasure_marker(Pleasure),
    window(Interval, NVS, Annotation),
    member(NVS, Pleasure).

pleasure_marker(NVS) :-
    assessment('Pleasure', NVS).
pleasure_marker(NVS) :-
    window(Interval, NVS, _),
    window(Interval, 'Pleasure', _),
    not(displeasure_marker(NVS)),
    not(neutral_marker(NVS)).
```

Listing 1: Querying the behavioural state.

Failing that, *pleasure_marker* will check if there is a NVS in the annotations associated with pleasure and not with displeasure or neutral state.

4.3 The Valence method

The second method treats the significance of the NVS as an indicator of behavioural state on a continuous scale. We assume that each NVS has a certain correlation with valence. In our case valence is a number that is correlated with the three behavioural states *(displeasure, neutral, pleasure)*, a simplified case of mapping feelings to an arousal-valence plane. Valence is assumed to be a value in [-1, 1] interval, where displeasure is associated with negative and pleasure with positive numbers. If there is little or no correlation between pleasure and the expression, it should gravitate towards negative values, as shown in Listing 2. Inverse must be true for displeasure. *correlation_set(NVS, Behavoural_state, Num_correlations)* returns the number of all annotated intervals that contain a NVS at the same time as the behavioural state. The *intervals(Behavioural_state, Num_examples)* returns the number of all annotated intervals of a certain behavioural state.

```
valence(NVS, Valence) :-
    correlation_set(NVS, Pleasure, NVS_P),
    correlation_set(NVS, Displeasure, NVS_D),
    correlation_set(NVS, Neutral, NVS_N),
    intervals(Pleasure, P_Set),
    intervals(Displeasure, D_Set),
    intervals(Neutral, N_Set),
    Valence_direction is NVS_P/P_Set
                       - NVS_D/D_Set,
    Valence_strength is 1 + P_Set
                      + N_Set+D_Set,
    Valence is Valence_direction/Valence_strength.
```

Listing 2: The function that calculates the valence.

We determine the behavioural state based on the value of the sum of valence scores (Listing 3). The *calculate_valence* is a recursive function that sums the valence of a set of NVS, and returns 0 for an empty set.

The P_Cut and D_Cut variables determine the intervals of pleasure, displeasure or neutral behavioural state. We use constraint logic programming to determine the optimal values for these values based on the training set. At its core it is a minimisation problem where we try to find the thresholds for the intervals that produce the smallest classification error. The pseudo-code of the algorithm is presented in Listing 4. Here *count_errors(Bag, Errors)* is a function that returns the number of missclassifications over the whole domain, and *min(Errors)* is a function that attempts to minimise the variable based on the constraints. It uses standard architecture for constraint logic programming on finite domains from the SWI-Prolog library, adapting it to the problem.

```
behaviour_state(NVS_Set, Decision, P_Cut, D_Cut) :-
    calculate_Valence(NVS_Set, Valence),
        (valence > P_Cut ->
            (Decision = Pleasure);
        (valence < D_Cut ->
            (Decision = Displeasure);
            (Decision = Neutral))).

calculate_valence([], 0).
calculate_valence([NVS | Rest], Valence) :-
    valence(NVS, V1),
    calculate_valence(Rest, V2),
    Valence is V1 + V2.
```

Listing 3: Determining the behavioural State.

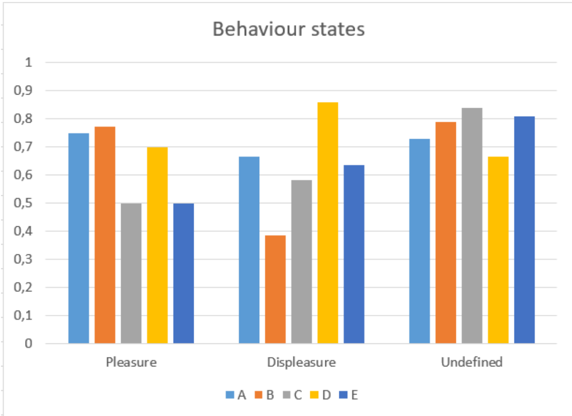

Figure 3: The accuracy from the decision tree classifier.

```
optimal_cut(P_cut, D_cut) :-
    [P_cut, D_cut] ins [inf..sup]
    min(Errors),
    findall(Correct,
            compare(Correct, P_cut, D_cut),
            Bag),
    count_errors(Bag, Errors).

compare(Correct, P, D) :-
    member(
    , All_observations)
    behaviour_state(NVS_Set, Decision, P, D),
    ground_Truth(NVS_Set, Truth),
    Correct = (Decision == Truth).
```

Listing 4: Calculating the minimal error.

5. RESULTS AND DISCUSSION

The benchmark we are trying to beat is based on standard machine learning algorithms from the python libraries. Out of the 10 algorithms tested, decision tree provided the highest accuracy (Figure 3).

The Unique Non-Verbal Signals method works surprisingly well with the limited data available. But it is expected to

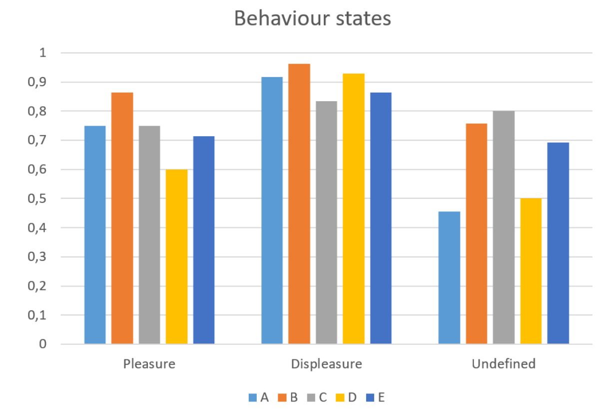

Figure 4: The Unique Non-Verbal Signals method classification accuracy.

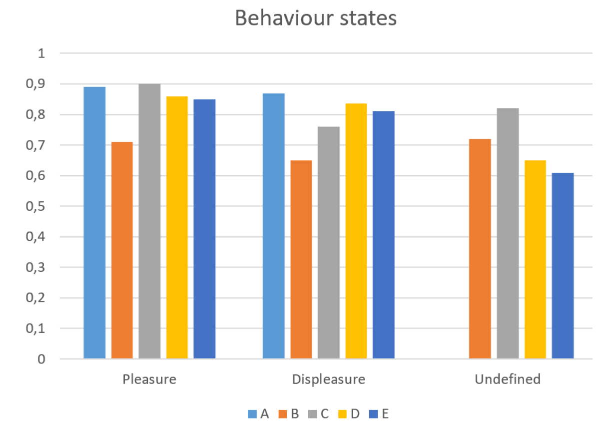

Figure 5: The Valence method classification accuracy.

become less viable with more data. The results in terms of classification accuracy are shown in Figure 4. The average classification accuracy for pleasure and displeasure is 81.8%. Due to uniqueness of the people with PIMD, a model is trained for each individual. The dataset for each individual was small, consisting of less than 10 annotated examples of each behavioural state. In order to evaluate our method, we trained the model on all the examples barring one for each state and compared it to the ground truth – based on annotations. The example here is a contiguous set of windows that annotate pleasure, so the data is not cross contaminated.

Using the same methodology, the Valence method performs worse than the somewhat naive Unique Non-Verbal Signals method, as seen in Figure 5. The average classification accuracy for pleasure and displeasure is 81.3%. Person A has very high miss-classification of neutral state, due to small sample size of this state. The Valence method seems to perform better for subjects with more annotations, perhaps indicating that it does not benefit as much from expert knowledge.

The two methods work on opposite spectrum, as Unique Non-Verbal Signals method with infinite data converges toward expert knowledge, while Valence method diverges from it.

6. CONCLUSIONS

In this paper we presented two ML methods specialised for learning behavioural states of people with PIMD. The advantage over the more common ML methods is the ability to incorporate prior knowledge in the from of assessments made by experts. The developed methods show promising results. The Unique Non-Verbal Signals method achieves 81.8% accuracy in classifying pleasure and displeasure and the Valence method achieves 81.3% on the same classes. However, more data must be collected and used to further validate our proposed methods. The next step is to use the environment of the person with PIMD to extract information on the actions that can be take in order to ameliorate the state our user finds himself in.

7. ACKNOWLEDGMENTS

This work is part of the INSENSION project that has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No. 780819.

8. REFERENCES

[1] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh. Realtime multi-person 2D pose estimation using part affinity fields. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, (Xxx):1302–1310, 2017.

[2] E. Lupiani, J. M. Juarez, J. Palma, and R. Marin. Knowle dge-Base d Systems Monitoring elderly people at home with temporal Case-Based Reasoning. *Knowledge-Based Systems*, 134:116–134, 2017.

[3] A. Metallinou, A. Katsamanis, and S. Narayanan. Tracking continuous emotional trends of participants during affective dyadic interactions using body language and speech information. *IMAVIS*, 31(2):137–152, 2013.

[4] A. Skarlatidis, G. Paliouras, and A. Artikis. Probabilistic Event Calculus for Event Recognition. 2014.

[5] H. Sloetjes and P. Wittenburg. Annotation by category - ELAN and ISO DCR. *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC'08)*, pages 816–820, 2008.

[6] Z. Soroush, B. Brain, M. Z. Soroush, K. Maghooli, S. K. Setarehdan, and A. M. Nasrabadi. A novel approach to emotion recognition using local subset feature selection and modified Dempster-Shafer theory. *Behavioral and Brain Functions*, 4:1–15, 2018.

[7] X. Sun, M. Lv, C. Quan, and F. Ren. Improved facial expression recognition method based on ROI deep convolutional neutral network. *2017 7th International Conference on Affective Computing and Intelligent Interaction, ACII 2017*, pages 256–261, 2018.

[8] P. Vos. *Measuring the Subjective Well-Being of People with Severe and Profound Intellectual Disabilites*. PhD thesis.