
Emotional analysis prediction using qualitatively representations of multiple psycho-physiological time-series signals.

Guillem Hernández¹

Scientific director: ²Beatriz López

¹G. Hernandez, BDBI ESCI-UPF, Barcelona, Spain, on an internship at the University of Girona, Spain (guillem.hernandez@alum.esci.upf.edu); ²B. Lopez, PhD, eXit research group, University of Girona, Spain. (beatriz.lopez@udg.edu).

Abstract

Motivation: This report provides techniques to deal with time-series emotional data. Transforming time-series data into qualitatively representations enables to reduce the complexity to mine the data while not altering the dimensionality by having a vocabulary. The method uses several individuals with an audiovisual stimulus and records multiple psychophysiological signals. Signals can be retrieved with a simple wristband, providing in time data and auto supervised, the best possible scenario for supervised and deep learning to take part.

Results: Multiple classifications and regression techniques are used to predict different emotions and accomplish a proper classification, reducing the error of prediction progressively. The methods are based on the mining of sequence information from the representations, giving insight into pattern recognition and sequence alignment to take knowledge from the data.

Supplementary information: Supplementary data are available at GitLab link, https://gitlab.com/ghg98/tfg-exit-emotional_prediction.git

1 Introduction

Emotions are mental states of living beings associated with thoughts, feelings or responses to different stimuli from the environment of living beings. Are responsible for the government of psychological changes influencing the behaviour of individuals. From an evolutionary point of view, emotions such as fear or anger could have had an essential role in natural selection, inducing certain adaptive behaviours in favour of species survival. The fact that there is no cognitive process required to have emotional states makes most species prone to have emotional pre-cognitive responses.

Psycho-physiological data has been successfully used to pinpoint the origin of some familiar, easily-distinctive emotions such as fear, anger, love and happiness. Nevertheless,

there is still much more work to do to have a precise mapping of the emotional spectra of species.

In fact, there is no learned principle to mine this kind of signals, always falling in models with no statistical significance, thus relying on simple preprocessing techniques and feature-driving methods. Preprocessing steps are an essential key to get a significant result. Simplistic normalisation procedures do not perform well when dealing with differences among subjects and when isolating the significant signal variations. The fact that emotions involve subjective experience, cognitive processes, psychophysiological changes and expressive behaviour entails unavoidable individual variability. The weak differentiation among responses leads to rigid models of emotions, which become more abstract and predictive intercorrelated. Possibilities to have more than one emotion at a time and varying them among individuals are not often

taken into account. There are easily predictable signals and more significant stimulus.

This report aims at establishing a method to mine psycho-physiological data to improve emotional prediction.

1.1 Related work

Studies correlate Heart Rate Variability and *Power-Spectra-Density* (PSD) analysis with the effect of emotions to characterise physiological illnesses such as depressions or panic-disorders (McCraty et al., 1995). Signals retrieved with wristbands, such as *electrodermal activity* (EDA) or skin temperature are key factors to predict depression patterns following *Hamilton Depression Rating Scale* (HDRS) (Ghandeharioun and Fedor et al., 2017). Beyond studies on depression, wearables have been useful in detecting relapse prevention (Kerz et al., 2016), anxiety (McGinnis et al., 2017) drug addiction (Fletcher et al., 2011), stress (Pino 2014, Choi 2017) or schizophrenia (Torous et al., 2018). Such gadgets provide individualised data that allow the modelling of individuals mental Health. The electrodermal activity has been used to identify excitement, and respiratory volume is used to provide insights into negative-valence emotions (Wu et al., 2011).

1.2 Objectives

This work aims to establish a method to learn emotional patterns from psycho-physiological data to enable their prediction. To that end, a wide range of problems has been addressed. Within-between variability among samples, dimensionality and feature-driving methods (no adapted to time-series data) makes the mining task difficult, needing new methods to have reliable results. Time series is preprocessed and discretised in qualitative representations. Different methods relying on distance sequence classification and pattern mining are used as classifiers. Results are compared and parameters tested to evaluate the reliability of classifications.

To cope with these problems, the objective of this project is to propose a method capable of shrinking the complexity of the data with its major drawbacks while enabling an adequate classification.

2 Materials Methods

Figure 1 shows an overview of the workflow conducted to address the different problems towards the objective, assembling the different methods used.

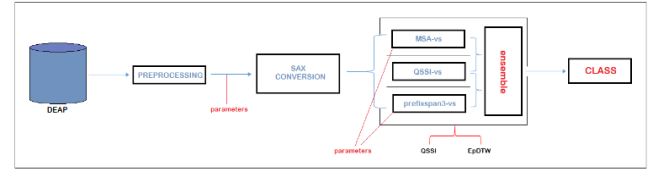


Fig. 1. Pipeline. Preprocessing and SAX conversion before different classification strategies to improve the classifiers.

2.1 Data

The psycho-physiological data used in this study belong to the DEAP data set (Database for Emotional Analysis using Physiological Signals). This is a multimodal database for the analysis of human affective states induced by audiovisual inputs (forty-one-minute long videos). The dataset encompasses the records of thirty-two healthy participants, with the same number of females and males.

To quantitatively describe emotions, the dataset uses a modification of Russell’s valence-arousal scale by placing emotional states on a two-dimensional plane with arousal and valence in the axis (Barrett, L. F., & Russell, J. A. et al., 1999). While arousal and valence explain most of the variation in the emotional field when included dominance and liking as variables to the model, the results improved. Arousal, valence, dominance and liking scales have a continuous classification range continuously from 1 to 9. The results are obtained from each participant self-assessment after each stimulus presentation.

The dataset gathers 32 data sources, one for each participant, that collects 40 physiological signals for each stimulus.

DEAP data set, although vast in terms of mixed signals recorded, provide a small number of them that can be used to predict emotions in non-experimental environments. Since the ultimate goal is to predict emotions in a daily life environment, only signals that can be retrieved by a wristband such as “empática-e4”: *Galvanic skin resistance* (GSR), *electrocardiogram* (ECG) and *skin temperature* (TEMP) have been used.

2.2 Preprocessing

In a first preprocessing step, while participants are watching the video clips, the study records physiological signals. A smoothing and downsampling from 512Hz to 128Hz reduces computation time.

In a second step, a class value is generated from the arousal, valence, dominance and liking information: class “0” for low/inactive (1-3.666) response, “1” for a neutral response (3.667-6.333) and “2” for high/active response (6.334-9).

An oversampling of the class ‘1’ is observed. Therefore a subsampling method is applied to reach a much more equally distributed training data. The test-train partitioning of the data with 10-fold cross-validation gives reliability to the results when performing this procedure.

Last pre-processing is related to obtain new features from blood volume pressure. *Heart Rate (HR)* and *Heart Rate Variability (HRV)* has shown to produce good results in terms of prediction (McCarty, Rollin et al., 1995). A MAX30100 pulse oximeter filter algorithm extrapolates from the blood volume pressure the HR / HRV, detecting peaks of higher volume. The algorithm consists of a “dc removal” and a “mean-median filter”. Classification strategies take place with the raw ECG signal and the filtered HRV.

2.3 Qualitative representations

Qualitative symbols, also called primitives by some authors (Rengaswamy, Venkatasubramanian, et al., 1995) represent time series while enabling the extraction of features that allow the following classification using other tools, mainly pattern recognition procedures.

These representations enable not only to reduce the complexity of the system but do not alter the dimensionality by having a vocabulary (a finite set of qualitative descriptions). They are assignments of qualitative symbols from a precomputed dictionary to equally sized blocks of time series.

In particular, this work uses *Symbolic Aggregate approXimation (SAX)* (Lin et al., 2003). The symbolic representation procedure has two steps (see Figure 2):

- (1) Time series is normalised using a mean filter through a simple spanning window technique.

A time series C of length n can be represented in a w dimensional space by a vector $\bar{C} = \bar{c}_1, \dots, \bar{c}_w$. The element i^{th} of \bar{C} is calculated by the following equation

$$\bar{c}_i = \frac{w}{n} \sum_{j=\frac{n}{w}(i-1)+1}^{\frac{n}{w}i} C_j \quad (1)$$

Eq. 1. PAA. Calculation of dimensionality reduction blocks in PAA segments.

To get w dimensions’ transformation of the time-series of n dimensions, data is divided into w equal sized “blocks” called episodes (Williams et al., 1986) defined by a time interval and a qualitative context. The mean value of the data falling within a block is calculated, and a vector of these values becomes the dimensionality-reduced representation

This normalisation method is called *Piecewise Aggregate Approximation (PAA)* representation (Keogh et al., 2001; Yi & Faloutsos, 2000).

- (2) Breakpoints produce equal-sized areas under Gaussian distribution, dividing the space of the data. When the PAA segment falls into the area inherits the corresponding letter from a precomputed vocabulary (PAA coefficients have associated symbols determined by breakpoints). The time series is, therefore, discretised into a string or word of length “ w ”.

The characteristic function can be other rather than Gaussian, but since time series are physiological signals, the natural distribution can better explain the fluctuation.

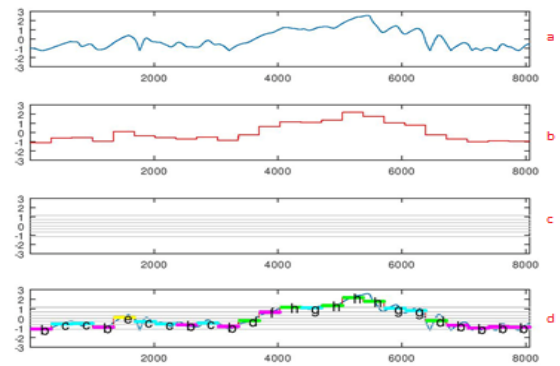


Fig. 2. Steps performed by SAX algorithm. (a) Raw time-series. (b) PAA formation and normalisation. (c) Gaussian distribution of space. (d) qualitatively discretisation.

Although the SAX representations are widely used, they show a particular drawback. Parameters must be set manually to minimise the error (see Figure 3). Alphabet length, PAA segments length, slopes (prominence of the slopes in the result) and stiffness (percentage of the length “ n ” of the time series to discretise) are parameters that must be set before the discretisation. That involves the expertise of data scientists since the settings may change in different datasets.

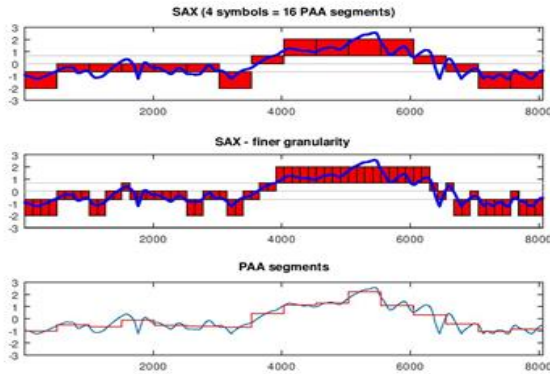


Fig. 3. Granularity effect. (a) If the reduction scale is large (n/w), the chance of information loss increases. (b) if it is small, the PAA segments show noise and do not adequately reduce dimensionality.

Several authors have adopted SAX since its creation. With the symbolic representations, PCA enables the reduction to 1-dimensional time series and find the characteristic patterns (Tanaka et al., 2005). *Minimum Description Length* (MDL) has been used to extract the most frequently occurring pattern over SAX sequences (Lin et al., 2001). Using PLA to compare patterns in a postprocessing step with SAX-representations has good results (Hung and Anh et al., 2007).

2.4 Distance Sequence Classification

There is no learned principle to mine time-series data in an efficient way. Calculating the distance between two time-series could be costly due to dimensionality. Extraction of symbolic representations enables efficiency while still gathering primary information.

Euclidean, Hamming or Manhattan distance measurements are sensible to distortions in time, making them unreliable when comparing sequences.

Alternative to them, QSSI is an algorithm used in engineering environments to compare two-time series to classify voltage sags (Gamero Argüello et al., 2019). The method proposes two different distance measurements to calculate over SAX representations and generate a distance

matrix, from where *one-neighbour classifier* and *svm* performs the classification.

The method uses two distance measurements:

(1) EpDtw:

Dynamic time warping (DTW) (Sankoff and Kruskal et al., 1983). It uses dynamic programming to align the time series to minimise distance (Bellman and Stuart, 1962; Sakoe and Chiba, 1978; Silverman and Morgan, 1990). With the time series using PAA segments, the algorithm uses a variation (*Piecewise dynamic time warping, ePDTW*). It finds the alignment by stretching or shrinking the time series along the time axis, warping them non-linearly (can be used to find similarity regions). Distance is the accumulation of local distances between corresponding elements.

A significant drawback is that the user must choose a compression rate for the dimensionality reduction.

(2) qssi:

Qualitatively Sequence Similarity Index (qssi) algorithm matches similar items (Cuberos et al., 2002). Allows greater flexibility by formulating the alignment in terms of similarity rather than in terms of the distance between sequences. Not aligning the sequences over the full length, makes the algorithm faster.

QSSI - vital signs, proposed in this work, uses as mainframe QSSI to generate an algorithm specially designed for psycho-physiological signals, enabling to perform classification over multiple signals at a time for the same stimulus. The algorithm needs a series of parameters for SAX representations specially tested to reinforce the model, adapting it to the type of signals.

One-nearest-neighbour classification and *svm* are not enough if several psycho-physiological signals need to be combined to avoid drawbacks of emotional data. Therefore, new classification methods are proposed, based on a consensus analysis over the classification of the different signals to give a final result.

QSSI-vs knn majority voting

It is designed to take as predicted the class with more entities among the “ k ” nearest neighbours. Posterior consensus analysis takes the majority class among the predictions of all different signals.

QSSI-vs knn aggregational voting

The algorithm is based on the time of arrival (distance). A sumatorium of the distances of each possible outcome is generated.

Consensus analysis makes a normalisation between classifications of all the signals. It takes the minimal distance predicted class.

2.5 Pattern sequence classification (text mining)

Similarity sequence classification is based on the alignment of sequences to evaluate the similarity from a structural point of view rather than distance in (Agrawal and Srikant, 1995). Two approaches are explored in this work: *Prefixspan3* and *MSA_2*.

2.4.1 Prefixspan3

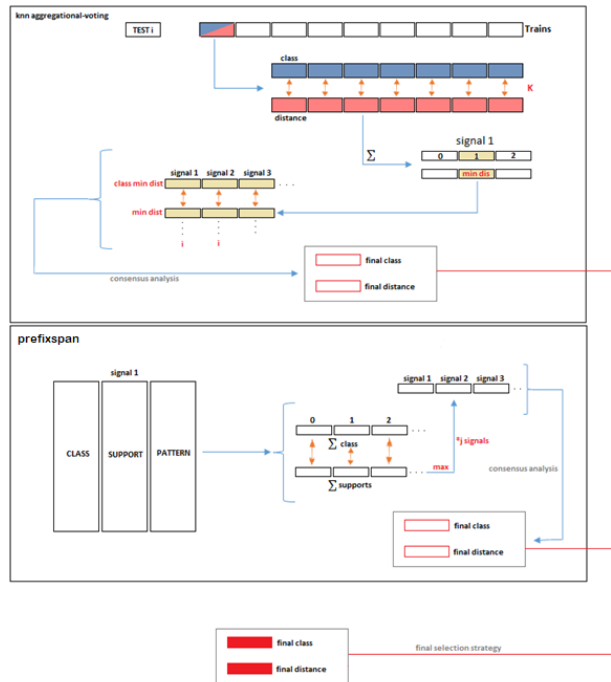


Fig. 4. Pipeline prefixspan3. The pipeline of the prefixspan3 algorithm gathered with the knn-aggregational voting workflow.

Prefixspan3 is based on *prefix-projected Sequential pattern mining*, *Prefixspan* (Pei J. et al., 2001). It is used on qualitative representations of signals (sequence of symbols).

The algorithm is used to discover all sequential patterns under a predefined support threshold (percentage of sequences presenting the pattern). The algorithm performs well for small datasets. In vast datasets, the time complexity is huge $O(n^2)$.

Prefixspan3 is a modification of the *prefixspan* algorithm. The raw algorithm was too flexible, enabling infinite time misalignments, which makes the prediction unreliable (the pattern could be found over the total length since the first symbol is matched). Moreover, it only detects if a pattern is present.

The *prefixspan3* version allows (see Figure 4):

- Setting a maximum distance to find the pattern, based on the percentage of the total length “ n ” of the sequence.
- Detecting the number of appearances for each pattern, not only whether it is present. That implies giving more reliability to repeated patterns.

The *prefixspan3* algorithm allows for a searching method similar to “*bag of words algorithm*”, but since sequences are not real words with a robust meaning, certain flexibility can be allowed. “*Bag of words algorithm*” search the number of appearances of words in different classes, and assigns a class to a new instance based on the matching of the words on it. In a similar method, *prefixspan3* finds the discriminant patterns between classes that have higher frequency to compare with each instance; it is a ‘bag of flexible patterns’ algorithm.

In this work, the algorithm is used on multiple signals. For each signal separates the balanced training set among the different classes, and finds the patterns of each class. Afterwards, it finds the discriminant patterns among the different classes of a signal (the patterns that are only present in each class). A database of discriminant patterns is performed for each class, allocating the pattern and the support.

Classification for each test is achieved by calculating the number of outcomes of each pattern. The similarity index is computed using the number of outcomes and the support for each pattern. The index is used with the QSSI-aggregational voting to perform a classification before the consensus analysis.

2.4.2 MSA_2

Multiple sequence alignment, MSA, is a method for aligning a database of different related sequences.

The algorithm pipeline is the following:

- (1) Computing a matrix of distances between all pairs of input sequences (based on similarity index)

- (2) Estimation of phylogenetic guide tree based on a distance matrix
- (3) Progressive global alignment according to a guided tree
- (4) Guided tree reestimation and realignment
- (5) iterative refinement
- (6) Post-processing and visualisation.

Global sequence alignment uses the NW algorithm (Needleman & Wunsch et al., 1970) for aligning sequences over their entire length to build up an optimal alignment using previous solutions for alignments of smaller subsequences. The algorithm uses dynamic programming (Gusfield et al., 1997) to find alignment maximising the similarity score.

Aligning is stated in terms of transforming sequence A into B while using a set of operations. Algorithm rewards match and penalises indels (insertions/deletions). Producing a matrix for a posterior trace-back. More research has been carried out to create gap-costs that allow block insertions and deletions (Gotoh, 1982; Sankoff & Kruskal, 1983)

MSA_2 uses *Multiple Sequence Alignment* as a mainframe. For each signal, sequences are separated regarding their class. A database of aligned sequences is computed for each class (alignment profile). For each query, the sequence is aligned to each class MSA profile and computes the similarity. Class is inherited from the class profile giving the minimum similarity measurement. Afterwards, a consensus analysis is performed between all signals.

3 Results and Discussion

A number of experimental methods have been carried out to test the different steps of the methodology.

3.1 Preprocessing

3.1.1 Classification strategy

Classification tends to use binary models (low/empowered) In the emotional analysis field; a binary model is useless. This means the individual is always in a reinforced or ignored state, not allowing for a neutral state. For example, a stimulus could like/dislike but also triggers no response at all.

Using these claims, the algorithm discretises the continuous scale of classification into a 3 class model. This model shows certain drawbacks. Having more classes will produce a high probability to fall in incorrect classes in each

classification that would be reduced if the classification was binary.

3.1.2 ECG analysis

Studies show that *Heart Rate Variability* and *Power-Spectral-Density* (PSD) analysis are correlated with the effect of emotions and are able to characterise physiological illnesses such as depressions or panic-disorders (McCraty et al., 1995). *Heart rate* (HR) and *Heart Rate Variability* (HRV) can be extrapolated from electrocardiogram signals using a MAX30100 Pulse Oximeter.

A MAX30100 Pulse Oximeter filter has been tested. The algorithm uses “*mean-median filter*” along with a “*dc removal*” to produce a much smooth signal from which HR can be discretised.

The classification does not improve with the discretisation, and the filter has to be improved since it is normalising the signal in such a way that useful information is being lost.

3.1.3 Balancing

Balancing does not produce better results. In all the datasets there is an oversampling of class “1”, neutral. Probably, there is more probability for a stimulus to produce a neutral response than to give an empowered or lowered response.

For that reason, when balancing with a subsampling, getting rid of useful information (is more likely for a specific stimulus to give a neutral response) produces worst results.

3.2 Results of QSSI algorithm

The percentage of error produced by the QSSI algorithm for single signals with a one-nearest neighbour classifier shows randomness in almost all the classifications (results over 0.66, as the method has three classes (see Table 1).

Table 1. % error on *QSSI-Inn* for single signals. Mean (Error - Arousal, Error - Dominance, Error - Liking and Error - valence).

SIGNAL	qssi	EpDTW
ECG	0,66015	0.61714
GSR	0,61327	0.61912
TEMP	0,66992	0.64253

3.3 Results of QSSI-vital signs

Figure 5 demonstrates the overlapping of emotional data. To improve the results, further classification with consensus analysis of each vital signal has been performed to improve the overall prediction.

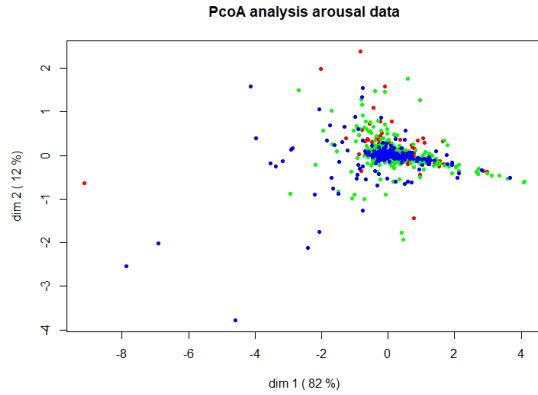


Fig. 5. PCoA on DEAP arousal data. Principal components analysis of arousal data. Demonstrating overlapping of the data

3.2.1 QSSI-vs majority voting

Using a knn classificatory with multiple signals, results are improved in all classifications. Knn does not define the classification based only on one instance, which makes a more robust model.

Table 3. % error on *QSSI-vs-knn_majority_voting* ($k = 35$) using multiple signals (ecg,gsr,temp).

SIGNAL	qssi	EpDTW
Arousal	0.546875	0.5859375
Dominance	0.5078125	0.4921875
Liking	0.578125	0.5625
Valence	0.5859375	0.546875

3.2.2 QSSI-vs aggregational voting

A knn based on the time of arrival noticeably improves the result in EpDTW distance matrix. Pondering the time of arrival to give prominence improves the result.

Table 4. % error on *QSSI-vs-knn_aggregational_voting* ($k = 35$) using multiple signals (ecg,gsr,temp).

SIGNAL	qssi	EpDTW
Arousal	0.5859375	0.6015625
Dominance	0.65625	0.5859375
Liking	0.578125	0.59375
Valence	0.6796875	0.671875

This method shows the worst results when compared with majority voting. Nevertheless, in balanced data, aggregational voting improves majority voting results, but never reaches the ones achieved with the unbalanced dataset.

3.4 Results of prefixspan3

The classification on the structure of the sequences provides the most common patterns included in a database of sequences.

By producing a database of each class in every signal, a support threshold of 0.75 (the minimum percentage of sequences in the database presenting the pattern) is used to make a database of patterns for each class and signal. Discrimination of the patterns get rid of patterns not specific of a class and shrinks the initial database, getting rid of small uninformative patterns.

The pattern database, with the pattern and a similarity index (calculated using the support and the number of outcomes), is used to dynamically find the pattern in each query.

Results are shown in Table 5. They have been obtained with a support threshold of 0.75. When setting support higher than 0.75, although getting more significant patterns, the discrimination lets some of the class pattern databases without a single instance. In the other hand, when support is smaller than 0.75, prefixspan is very expensive in terms of time complexity, and the patterns do not have sufficient support to serve as predictors.

We need a support threshold smaller than 0.75 to get a large enough pattern database (with discriminant entries for each class) but is impossible in terms of computational resources. For that reason, if there is no pattern in the class database, the algorithm performs *35knn_aggregational_voting* from QSSI-vs to make the consensus.

Table 5. % error on *prefixspan3-35knn_aggregational_voting*

SIGNAL	QSSI	EpDTW
ECG	0.6875	0.6875
GSR	0.63281	0.63281
TEMP	0.65625	0.65625
MULTIPLE	0.64063	0.64063

The algorithm does not improve the previous results. Further experimentation with different SAX sequences will probably improve the results.

3.5 Results of MSA_2

MSA takes care of the patterns while also taking care of the order of appearance. Prefixspan3 finds the more significant patterns but does not retrieve the location on the sequence. Since the location of patterns can be used as a predictor, MSA implementation alienates the sequence while giving insight on the position of alignment. Results are shown in Table 6.

Table 6. % error on *MSA_2* for arousal prediction.

SIGNAL	QSSI	EpDTW
ECG	0.65625	0.65625
GSR	0.7265625	0.7265625
TEMP	0.640625	0.640625
MULTIPLE	0.6484375	0.6484375

3.6 Participant analysis

Emotional data is very subjective and depends on many parameters. The patient must perform a self-classification of the emotions, and the scales may change from one patient to another. Because of this, and the fact that there are individual variations within psychophysiological signals of patients makes the classification even more difficult

To solve these significant drawbacks, the model must adopt representations able to handle individual differences in responses and account for essential signal variations, while including flexible sequential behaviour. That improvement could be used to make a self-learning method that uses the individual's data, rather than other data, to continually improve. By doing that, differential behaviour patterns among individuals are eliminated (Henriques et al., 2014).

To minimise within-between individual differences, each individualised dataset can be split between predictor and predicted. Inputs are always the same for the 32 participants.

Results are shown in table 6 and 7. Best results are achieved using that technique, demonstrating that the main source of noise responses differences between individuals. The better predictor in all cases is QSSI-vital signs with knn majority voting (using as $k=35$)

Table 6. % error on participant data using EpDTW distance measurement.

	QSSI-1nn	QSSI-vs-Knn-MJ	QSSI-vs-Knn-AG	MSA_2
Arousal	0.52343	0.4843	0.625	0.7096
Dominance	0.5859	0.42128	0.6718	0.6826
Liking	0.61718	0.46875	0.67187	0.625
Valence	0.648437	0.515625	0.67187	0.6774

Table 7. % error on participant data using QSSI distance measurement.

	QSSI-1nn	QSSI-vs-Knn-MJ	QSSI-vs-Knn-AG	MSA_2
Arousal	0.48437	0.468775	0.63281	0.7083
Dominance	0.57812	0.42968	0.6875	0.7019

Liking	0.57812	0.46687	0.73437	0.625
Valence	0.65625	0.54687	0.625	0.6774

The best prediction is achieved using knn_majority voting (taking as $k = 35$). When comparing the g with the whole dataset, results when using only participant data are better.

3.7 Parameter variation analysis

A number of parameters testing have been carried out to check the effect of parameter variation on the results.

3.7.1 vocabulary effect

The length of vocabulary is one of the parameters that define the final dimension of qualitative representations. Choosing a proper length is critical to have better results.

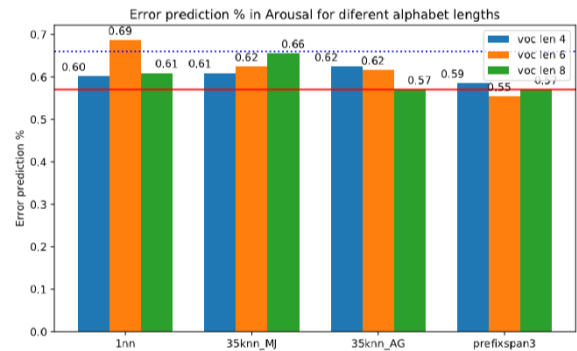


Fig. 6. Vocabulary length effect. Error in Arousal prediction for different vocabulary effects.

Figure 6 shows the result of analysing the impact of different vocabulary sizes, from 4 to 8 symbols. There is no apparent effect of the length of the vocabulary on the predictions (for different predictions and different signals there is not a significant relevance).

3.7.2 PAA effect

Best results are achieved using PAA segments of 64. Since Data is downsampled to 128hz, segments on 64 still gather essential information while reducing substantially the dimensionality getting rid of the granularity problem.

3.7.3 k effect

The k effect demonstrates the hypothesis that when using individualised data, the results are better than using as a train and test the whole dataset.

Since the emotional classification and physiological signals heavily depend on the individual (introducing noise in the

dataset), to reduce noise the noise, and therefore improve the prediction, the model should focus in using as predictors signals from the same individual as the test. This reasoning produces a high improvement in the prediction when using participant data, by dropping the overall error from 0.60939 when using as predictors all individuals to 0.468775.

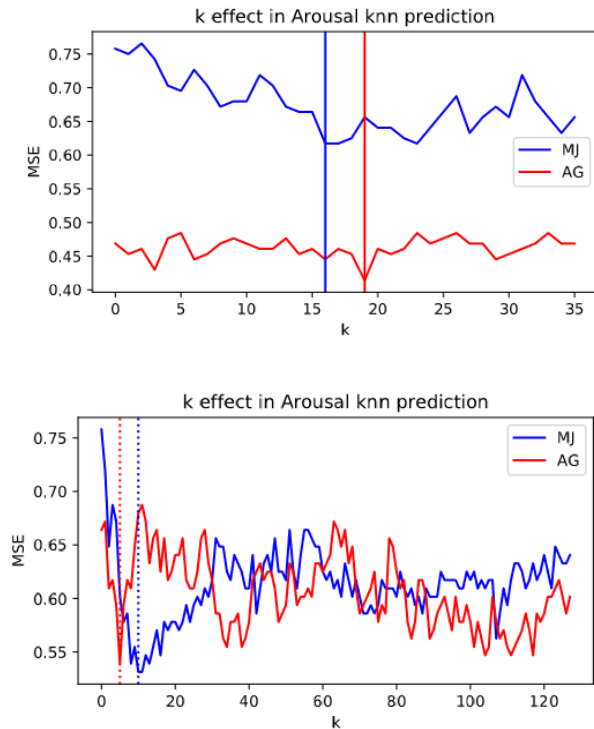


Fig. 7. K effect on prediction error. (1) Effect of k in individualised data. (2) effect of k in the whole dataset.

In some of the predictions (Arousal and liking), for knn methods (it is more noticeable in aggregational voting), the best k in individualised data is very close to the one when using the whole dataset as it is shown in fig. 7. That effect suggests there is much noise in the whole dataset, that is not as heavy in the individualised one. Variability between individuals, produced by different physiological responses to the same stimulus and self-assessment could be the principal source of variation. Models to mine that type of signals must use individualised data to get better results.

4 Conclusions

Emotional data show significant limitations and drawbacks in terms of predictive analysis. Subjectiveness in self-assessment, with no clear empirical support, creates a limitation on the classificatory step. Parameters must be clear, and a method as objective as possible must be chosen to avoid distortions between users. Moreover, the complexity of the data, with time-series structure, have been a limitation and a jigsaw for years in the computational machine learning field.

The need for new methods to reduce the complexity of the signals while enabling for a proper prediction is clearly stated when mining physiological signals.

To tackle this problem, this report provides a methodology based on the simultaneous use of different psychophysiological signals to improve the mining of emotional data. Discretising time-series into character strings reduces the complexity and also provides a valid input for feature-driven models and text mining approaches. This method integrates sequential and feature-driving methods to perform a reliable classification. Several classifiers have been tested to improve the prediction. The best prediction is achieved after a knn classification of a distance matrix by sequence distance methods of QSSI and EpDTW, two methods based on computing distance between sequences after alignment approaches.

In order to improve the results, methods must focus on creating individualised signal profiles to achieve an accurate emotion prediction. Focusing on the structure of the signal as a whole, by aligning and getting the distances, and with features associated with states are the best way to improve the results. The better results are achieved using each individual as an independent instance (responding to an input in an individualised way) rather than treating all individuals equally. Since emotions can be different between individuals, and physiological signals too, the best option is to create an individualised model, refining it progressively to accurate the prediction.

Best results are achieved when using as predictors, sequences from the same individual as the predicted ones. Moreover, the k effect analysis shows clear evidence that the number of predictors to choose tends to be nearby to the number of instances of each individual.

Self-learning approaches, individualised, what integrate a higher variety of signals (that can be provided by a gadget such as a wristband), with an objective self-assessment tool, will be the key to achieve robust models that can mine the signals to achieve reliable emotional prediction tools in a daily environment.

Acknowledgements

Funding

This work has been supported by the eXiT research group, {UdG}.

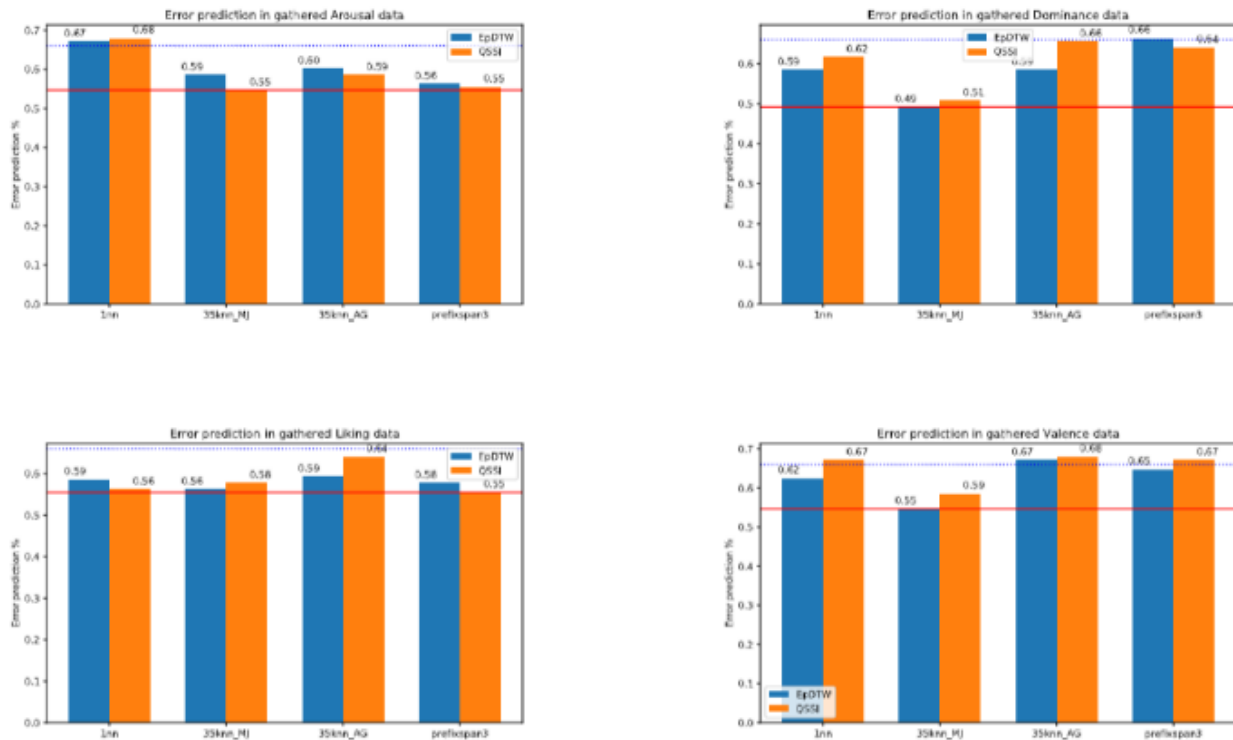
Conflict of Interest: none declared.

References

- Ghandeharioun, A., Fedor, S., Sangermano, L., Ionescu, D., Alpert, J., Dale, C., Son-tag, D., Picard, R. "Objective assessment of depressive symptoms with machine learning and wearable sensors data," International Conference on Affective Computing and Intelligent Interaction (ACII), San Antonio, Texas, October 2017
- Pei J., Han J., Mortazavi-Asl J., Pinto H., Chen Q., Dayal U., Hsu M., "PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth", 17th International Conference on Data Engineering (ICDE), April 2001.
- Agrawal R., Srikant R., "Mining sequential patterns", Proceedings 1995 Int. Conf. Very Large Data Bases (VLDB'94), pp. 487-499, 1999.
- Han J., Dong G., Mortazavi-Asl B., Chen Q., Dayal U., Hsu M.-C., "Freespan: Frequent pattern-projected sequential pattern mining", Proceedings 2000 Int. Conf. Knowledge Discovery and Data Mining (KDD'00), pp. 355-359, 2000.
- Wojciech Stach, "http://webdocs.cs.ualberta.ca/~zaiane/courses/cmput695-04/slides/PrefixSpan-Wojciech.pdf".
- Erken, Eray. How to Measure Heart Rate with MAX30100 Pulse Oximeter. 30 April 2016. <https://www.youtube.com/watch?v=36cELpNGJZY>How to Measure Heart Rate with MAX30100 Pulse Oximeter (accessed December 1, 2016).
- Kennedy, Stephen M. "AN INTRODUCTION TO PULSE OXIMETERS: EQUATIONS AND THEORY." The University of Wisconsin-Madison. 20 April 2015. <http://www.imt.liu.se/FoUtb/kurser/oxikurs/restricted/pulseoximetersequationsandtheory-stephenkennedy.pdf> (accessed January 15, 2017).
- Maxim. "Pulse Oximeter and Heart-Rate Sensor IC for Wearable Health." MaximIn-tegrated. September 2014. <https://datasheets.maximintegrated.com/en/ds/MAX30100.pdf> (accessed November 18, 2016).
- Sang Stroganovs, Raivis. "Implementing Pulse Oximeter Using MAX30100 - MORF - Coding And Engineering". *Morf.Lv*, 2019.
- McCraty, Rollin et al. "The Effects Of Emotions On Short-Term Power Spectrum Analysis Of Heart Rate Variability". 1995.
- Pous, Carles et al. "Modeling Reuse On Case-Based Reasoning With Application To Breast Cancer Diagnosis." 2019.
- Henriques, Rui et al. "Seven Principles to Mine Flexible Behavior from Physiological Signals for Effective Emotion Recognition and Description in Affective Interactions." *PhyCS* (2014).
- H. Xiong et al. "A brief survey on sequence classification." SIGKDD Explorations (2010).
- Espinosa-Carrasco, Jose et al. *Pergola: Boosting Visualization And Analysis Of Longitudinal Data By Unlocking Genomic Analysis Tools*. 2019.
- Koelstra, Sander. "DEAP: A Dataset For Emotion Analysis Using Physiological And Audiovisual Signals". *Eecs.Qmul.Ac.Uk*, (2019), <http://www.eecs.qmul.ac.uk/mmv/datasets/deap/>.
- Gamero Argüello, Fco. Ignacio. "Pattern recognition based on qualitative representation of signals. Application to situation assessment of dynamic systems". *Dugi-Doc.Udg.Edu*, 2019, <https://dugi-doc.udg.edu/bitstream/handle/10256/7039/tfiga.pdf?sequence=5>.
- Aamodt, A., & Plaza, E., (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications*. Retrieved from <http://iospress.metapress.com/index/316258107242JP65.pdf>
- Ang, H. H., Gopalkrishnan, V., Zliobaite, I., Pechenizkiy, M., & Hoi, S. C. H. (2013). Predictive Handling of Asynchronous Concept Drifts in Distributed Environments. *IEEE Transactions on Knowledge and Data Engineering*, 25(10), 2343–2355. <https://doi.org/10.1109/TKDE.2012.172>
- Armengol, E., (2016). Assessment of Diabetic Complications Based on Series of Records. In *First AI4D workshop, ECAI* (pp. 28–30).
- Aung, M. H., Matthews, M., & Choudhury, T. (2017). Sensing behavioural symptoms of mental health and delivering personalised interventions using mobile technologies. *Depression and Anxiety*, 34(7), 603–609. <https://doi.org/10.1002/da.22646>
- Chih-Fong Tsai, "Bag-of-Words Representation in Image Annotation: A Review," *ISRN Artificial Intelligence*, vol. 2012, Article ID 376804, 19 pages, 2012. <https://doi.org/10.5402/2012/376804>.
- McCraty, Rollin & Atkinson, Mike & A.Tiller, William & Rein, Glen & D. Watkins, Alan. (1995). The Effects of Emotions on Short-Term Power Spectrum Analysis of Heart Rate Variability. *The American journal of cardiology*. 76. 1089-93. 10.1016/S0002-9149(99)80309-9.
- Barrett, L. F., & Russell, J. A. (1999). The Structure of Current Affect: Controversies and Emerging Consensus. *Current Directions in Psychological Science*, 8(1), 10–14. <https://doi.org/10.1111/1467-8721.00003>

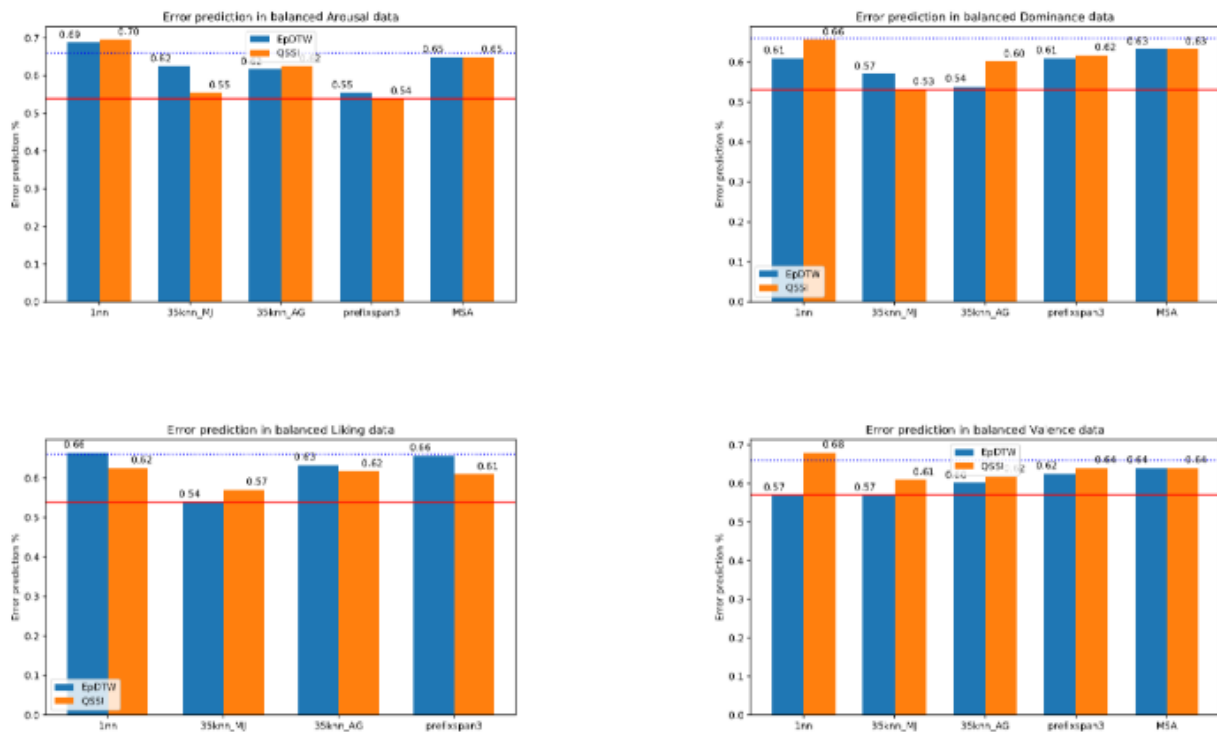
Annexe, plot 1.

Result of the different algorithms on unbalanced data set with QSSI and EpDTW measurements. [% error of prediction, the blue line is the random threshold, the red line is lower prediction error]



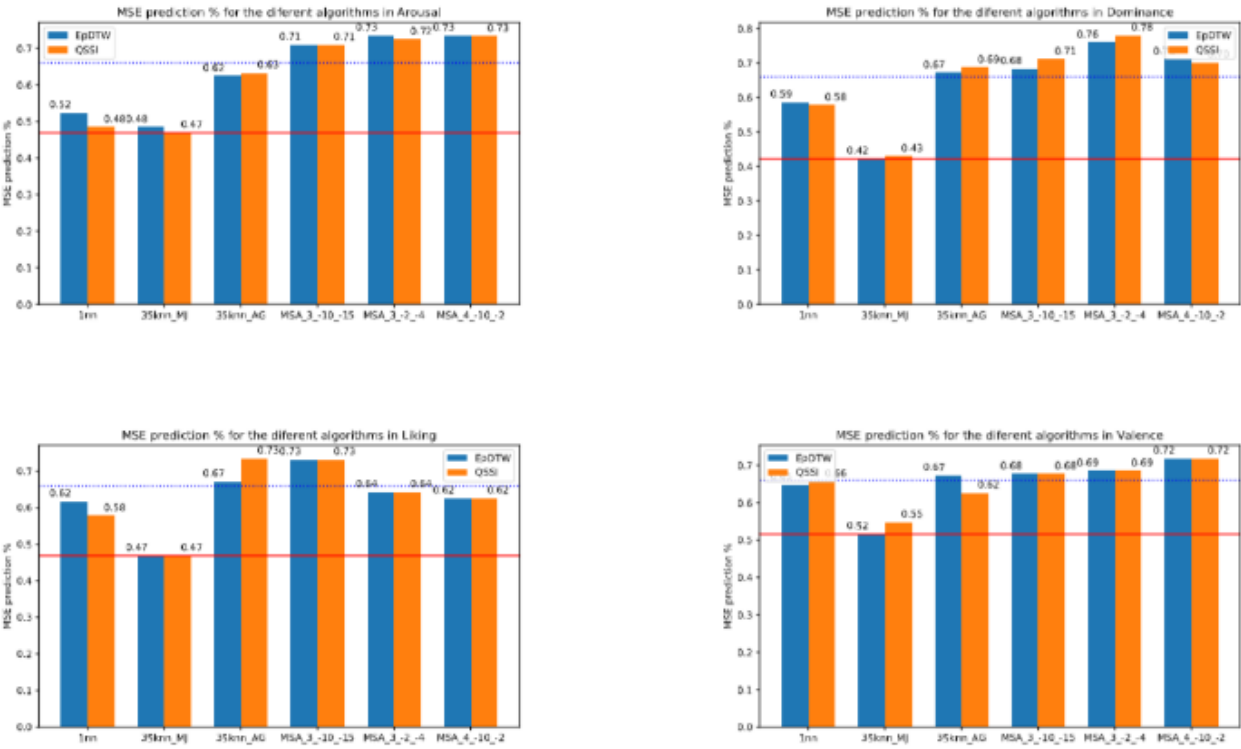
Annexe, plot 2.

Result of the different algorithms on balanced data set with QSSI and EpDTW measurements. [% error of prediction, the blue line is the random threshold, the red line is lower prediction error]



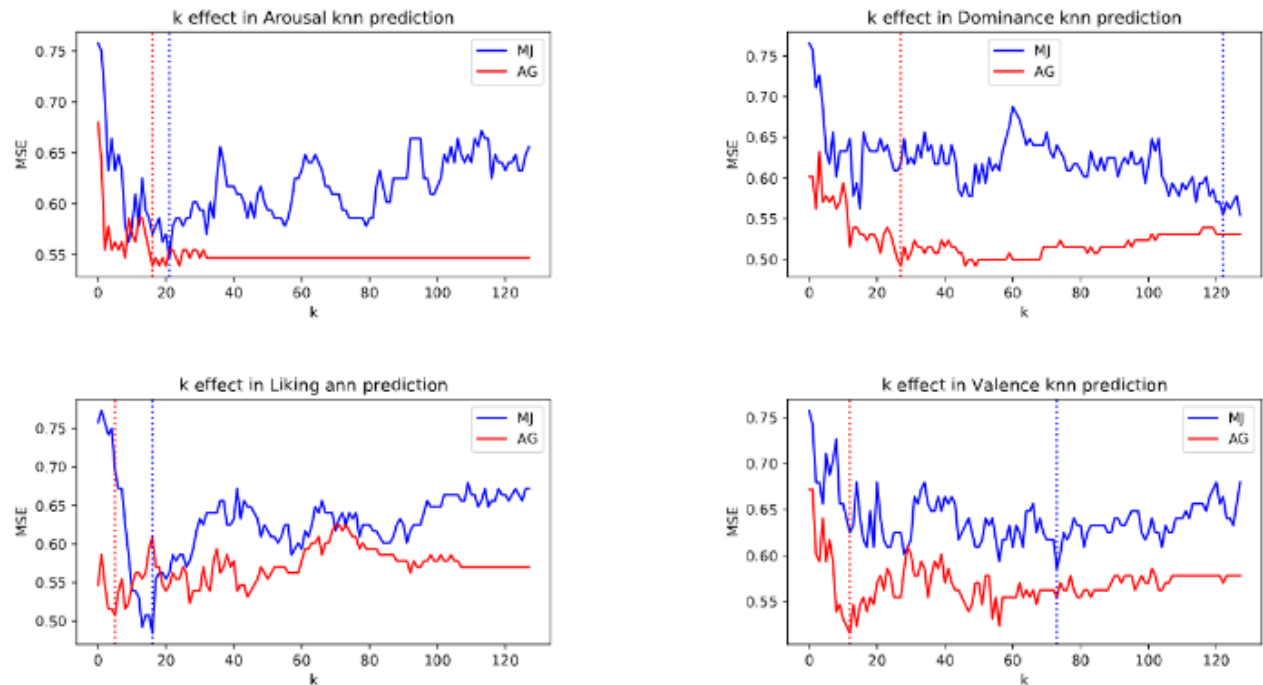
Annexe, plot 3.

Result of the different algorithms on balanced data set with QSSI and EpDTW measurements. [% error of prediction, the blue line is the random threshold, the red line is lower prediction error].



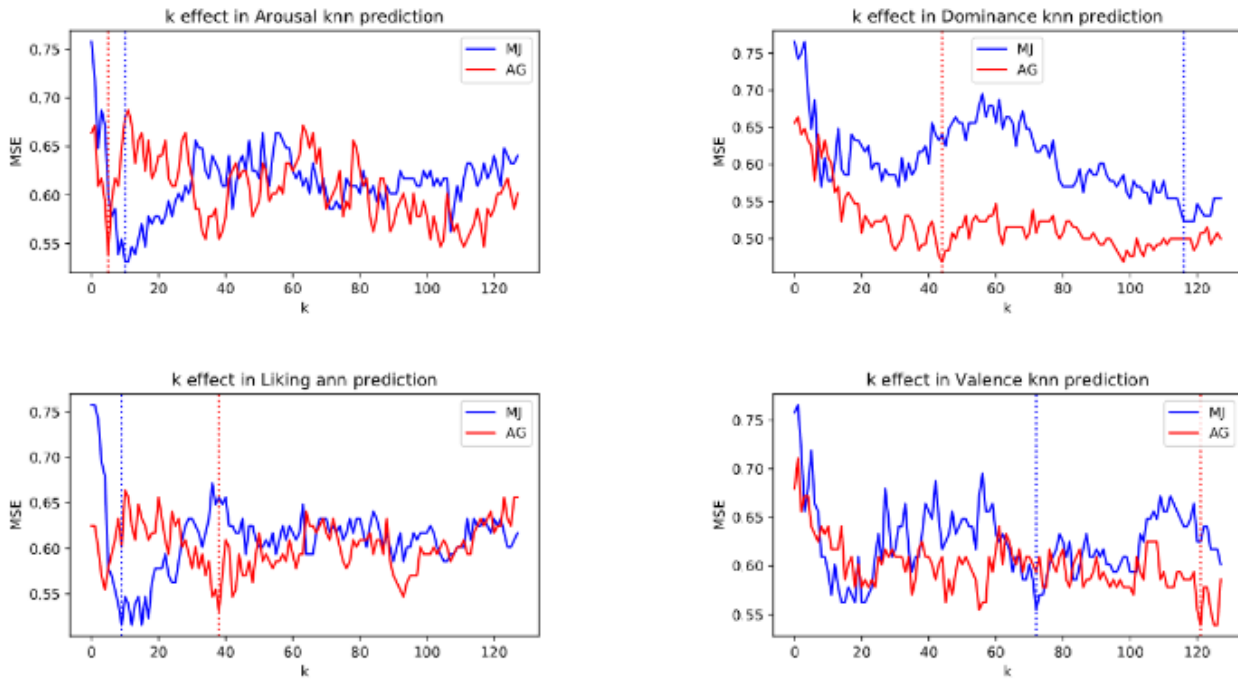
Annexe, plot 4.

Result of the different knn algorithms for different k in the unbalanced dataset. [k value ranges from 1 to 128, algorithms are knn-majority voting and knn-aggregational voting].



Annexe, plot 5.

Result of the different knn algorithms for different k in the balanced dataset. [k value ranges from 1 to 128, algorithms are knn-majority voting and knn-aggregational voting].



Annexe plot 6.

Result of the different knn algorithms for different k in participant data set. Average of the 32 participants. [k value ranges from 1 to 36, algorithms are knn-majority voting and knn-aggregational voting].

