Protocol Weka

# Prerequisite

* Weka Version: 3.8.0
* Interruption Data Set with Normalized Features

# Feature Selection

***Open Weka “Explorer”***

## Feature Selection: 2-State

1. I used CfsEvalSubset with Cross-Validation.
2. This led to unexpected behaviour: Only a single feature was selected.
3. I reran CfsEvalSubset on the whole Data Set to be able to inspect the Merit Score
   1. This led to 0 Merit for the selected features.
   2. Interpretation: CfsEvalSubset does not detect any “relatedness” between dependent and independent variables

Ein Bild, das Text, Screenshot, Schrift enthält.

Automatisch generierte Beschreibung

1. Switching to Wrapper-based Feature Selector to reduce feature set for better comparability:

**Initial Assumption**: 10s Time Windows with NaiveBayes achieve best results (Time Windows will be tested later on).

Ein Bild, das Text, Screenshot, Zahl, Schrift enthält.

Automatisch generierte BeschreibungEin Bild, das Text, Screenshot, Zahl, Display enthält.

Automatisch generierte Beschreibung

### Selected Features:

1. delta
2. theta/gamma
3. alpha/gamma
4. num\_of\_peaks
5. mean\_temp

## Feature Selection: 5-State

*\*\*Repeat Steps from 2-State Classification\*\**

1. I used CfsEvalSubset with Cross-Validation.
2. This led to unexpected behaviour: Only a single feature was selected.
3. I reran CfsEvalSubset on the whole Data Set to be able to inspect the Merit Score
   1. This led to 0.173 Merit for the selected features.
   2. Interpretation: CfsEvalSubset does not detect close to no “relatedness” between dependent and independent variables

Ein Bild, das Text, Screenshot, Schrift enthält.

Automatisch generierte Beschreibung

1. Switching to Wrapper-based Feature Selector to reduce feature set for better comparability:

**Initial Assumption**: 10s Time Windows with NaiveBayes achieve best results (Time Windows will be tested later on).

Ein Bild, das Text, Screenshot, Zahl, Schrift enthält.

Automatisch generierte BeschreibungEin Bild, das Text, Screenshot, Zahl, Display enthält.

Automatisch generierte Beschreibung

### Feature Selection:

1. alpha/gamma
2. max\_peak\_amplitude
3. pnn50

# Building Classifier

Initial Test of Models using Simple 10-Fold Cross Validation.

## 2-State

**Performance Majority Classifier (ZeroR):** 60.9756 %

Performance Naive Bayes: 74,3902% (10-k CV / Seed = Matriculation Number:1551574)

Ein Bild, das Text, Screenshot, Schrift, Zahl enthält.

Automatisch generierte Beschreibung

## 5-State

**Performance Majority Classifier (ZeroR):** 35.3659 %

Performance Naive Bayes: 46.3415 % (10-k CV / Seed = Matriculation Number:1551574)

Ein Bild, das Text, Screenshot, Schrift, Zahl enthält.

Automatisch generierte Beschreibung

# Manually Creating Leave One Subject Out Folds

pID: Needs to be converted from Numeric to Nominal first! (ID:39 in the Screenshots points at Column with pID. (Note: The column ID changes after feature selection.)

Train Set: All participants except one. Test Set: Left out participant.

*Repeat for 10 participants.*

Ein Bild, das Text, Screenshot, Display, Zahl enthält.

Automatisch generierte BeschreibungEin Bild, das Text, Screenshot, Display, Zahl enthält.

Automatisch generierte Beschreibung

Figure 1 Filtering for Test Set Figure 2 Filtering For Train Set

Remove pID Column.

Saving 10 Training Sets and 10 Test Sets.

Repeat for 5 State Data Set.

# Validation

## Comparing Classification Results

For validation of the classifier performance, it is unclear what the base performance is. From the original paper it seems as if they took the performance of a majority classifier being used once as the baseline.

Therefore, I assume they did a one-sided T-test when testing for a significant difference in performance between the baseline and the accuracy samples of the cross-validations.

Another possible approach would have been to build a majority classifier for each cross-validation fold as a baseline model and do a (corrected) two-sided T-Test instead.

## Repeated Cross Validation

10 times 10-Fold CV

Validating the Naïve Bayes Classifier against a Majority Vote Classifier (ZeroR)

Source Code:

* PerInstanceValidation2State.java
* PerInstanceValidation5State.java

### 10 times 10-Fold CV Results: 2-State

### 10 times 10-Fold CV Results: 5-State

## Per Participant Cross Validation

(Aka Leave One Subject Out)

Source Code:

* PerParticipantValidation2State.java
* PerParticipantValidation5State.java

## Summary of Cross Validation Results

(\*One Sample T-Test P-Value < 0.05)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **States** | **Per Instance CV** | | | **Per Participant CV** | | | **Majority Classifier** |
| *Accuracy* | *Cohen’s Kappa* | *Stdev* | *Accuracy* | *Cohen’s Kappa* | *Stdev* |
| Two | 76%\* | 0.48 | 13.1 | 71% | 0.11 | 16.4 | 60.98% |
| Five | 46\* | 0.26 | 14.8 | 48% | 0.22 | 21.6 | 35.37% |

**Results of the Original Study:**

Ein Bild, das Text, Screenshot, Schrift, Zahl enthält.

Automatisch generierte Beschreibung

Comparing the results of the original study with my own, although there is a comparable improvement on baseline accuracy, in my validation the accuracy varied far stronger in the repeated cross validation (Stdev[2-State]: 13,5 vs 0,7; Stdev[5-State]: 16,3 vs 2,9) This means that the final performance of our model may be only a bit better (or worse) than a simple majority classifier.

# Comparing Time Windows

Again, I will need to select the time windows based on wrapper-based attribute selectors, since CFSEval does not work on my Data Set.

I assume they validated all time windows using 10 times 10-fold Cross Validation since the points for the 10 seconds time window on their comparison graphic match the values in the table from the repeated cross validation.

## Wrapper Attribute Selection Protocol

Apply Wrapper-base Feature Selection to each Time Window Data Set.

Ein Bild, das Text, Screenshot, Schrift, Zahl enthält.

Automatisch generierte Beschreibung Ein Bild, das Text, Screenshot, Zahl, Display enthält.

Automatisch generierte Beschreibung

Ein Bild, das Text, Screenshot, Display, Zahl enthält.

Automatisch generierte Beschreibung

Save filtered Data Set as:

"[Project Path]\Normalized Data\[Time Window]\with\_pID\normalized\_interruption\_data\_ALL\_tw\_[Time Window]s\_pID\_3hz\_[2state/5state]\_filtered\_attributes.arff"

## Attribute Selection (TWs)

### Attribute Selection 10s 2-State:

CFSEval (Train Set): (Merit: 0)

Wrapper:

1. delta
2. theta/gamma
3. alpha/gamma
4. num\_of\_peaks
5. mean\_temp

### Attribute Selection 20s 2-State:

CFSEval (Train Set): (Merit: 0,272)

1. alpha/gamma
2. max\_peak\_amplitude
3. sum\_peak\_amplitude

Wrapper:

1. theta
2. delta/theta
3. delta/gamma
4. theta/alpha
5. theta/gamma
6. num\_of\_peaks
7. mean\_temp

### Attribute Selection 30s 2-State:

CFSEval (Train Set): (Merit: 0)

Wrapper:

1. alpha
2. theta/gamma
3. mean\_hr\_in\_bpm
4. num\_of\_peaks
5. mean\_temp

### Attribute Selection 45s 2-State:

CFSEval (Train Set): (Merit: 0)

Wrapper:

1. theta
2. alpha
3. delta/gamma
4. mean\_hr\_in\_bpm
5. num\_of\_peaks
6. mean\_temp

### Attribute Selection 60s 2-State:

CFSEval (Train Set): (Merit: 0)

Wrapper:

1. beta
2. delta/alpha
3. delta/gamma
4. mean\_hr\_in\_bpm
5. sdnn
6. pnn50
7. max\_temp
8. num\_of\_peaks
9. mean\_temp

### Attribute Selection 120s 2-State:

CFSEval (Train Set): (Merit: 0)

Wrapper:

1. delta/alpha
2. theta/beta
3. gamma/alpha
4. num\_of\_blinks
5. hr\_variance\_in\_bpm
6. max\_temp

### Attribute Selection 10s 5-State:

CFSEval (Train Set): (Merit: 0.173)

1. theta/gamma

Wrapper:

1. alpha/gamma
2. max\_peak\_amplitude
3. pnn50

### Attribute Selection 20s 5-State:

CFSEval (Train Set): (Merit: 0)

Wrapper:

1. delta/alpha
2. theta/delta
3. alpha/gamma
4. hr\_variance\_in\_bpm
5. max\_peak\_amplitude
6. sum\_peak\_amplitude
7. pnn50

### Attribute Selection 30s 5-State:

CFSEval (Train Set): (Merit: 0)

Wrapper:

1. theta/delta
2. hr\_variance\_in\_bpm

### Attribute Selection 45s 5-State:

CFSEval (Train Set): (Merit: 0)

Wrapper:

1. delta/theta
2. gamma/beta
3. num\_of\_peaks

### Attribute Selection 60s 5-State:

CFSEval (Train Set): (Merit: 0)

Wrapper:

1. beta/gamma
2. gamma/beta
3. num\_of\_peaks
4. sum\_peak\_amplitude
5. pnn20
6. mean\_temp

### Attribute Selection 120s 5-State:

CFSEval (Train Set): (Merit: 0)

Wrapper:

1. beta/gamma
2. gamma/beta
3. pnn50

### Attribute Selection 180s 5-State:

CFSEval (Train Set): (Merit: 0)

Wrapper:

1. theta/alpha
2. alpha/delta
3. gamma/beta
4. hr\_variance\_in\_bpm
5. max\_temp

## Results: Comparing Time Windows

As observed in the original study shorter time windows (<60s) perform better than longer ones. I would also have the same intuition as the original author, that the shorter time windows probably benefit from having less artifacts. As in the original study this trend is much weaker for the 5-State predictions.

|  |  |  |
| --- | --- | --- |
|  | **2-State** | **5-State** |
| **10s** | 76% | 46% |
| **20s** | 73% | 45% |
| **30s** | 76% | 44% |
| **45s** | 75% | 42% |
| **60s** | 75% | 41% |
| **120s** | 67% | 38% |
| **180s** | 66% | 42% |

Ein Bild, das Reihe, Diagramm, Screenshot, Quittung enthält.

Automatisch generierte Beschreibung

Figure Graph from Original Paper