CCBER - Software Description Document

Brian Vegetabile

June 2, 2018

# Overview of Software

is an R package for the estimation of behavioral entropy rate, developed for the Conte Center @ UCI. The package reads in an individual, or set of, excel files and processes the file to estimate the entropy rate from the supplied data. This document describes usage and input formats that are expected with this software.

See reference: Davis, E.P., Stout, S.A., Molet, J., Vegetabile, B., Glynn, L.M., Sandman, C.A., Heins, K., Stern, H., Baram, T.Z. (in press). **Exposure to unpredictable maternal sensory signals influences cognitive development across-species**. *Proceedings of the National Academy of Sciences*.

An overview of behavioral entropy rate estimation in the context of human behavior is found in the supplemental in the reference above.

The source files for ccber are found at [github.com/bvegetabile/ccber](https://www.github.com/bvegetabile/ccber)

## Installation of ccber

The package devtools is required to install this R package from this Github repository. Install this package first if it is not already installed.

install.packages('devtools', dependencies = TRUE)

Once that package has been installed, use the following to install ccber

devtools::install\_github('bvegetabile/ccber')

Load the package to begin analysis!

library('ccber')

## Function Overviews

### Function : ber\_analyze\_file

The following function from ccber is one of the primary functions for the estimation of entropy rate:

ber\_analyze\_file(f\_loc,  
 plot\_all=F,  
 plots\_to\_file=F,  
 tactile\_padding = 1.0,  
 auditory\_padding = 1.0,  
 behavior\_types=list(  
 "mom\_auditory\_types" = c('Vocal'),  
 "mom\_tactile\_types" = c('TouchBaby',  
 'HoldingBaby'),  
 "mom\_visual\_types" = c('ManipulatingObject'),  
 "baby\_visual\_types" = c('LookAtMomActivity'),  
 "missing\_types" = c('CantTellHolding',  
 'ActivityNotVisible',  
 'CantTellLooking')),  
 missing\_threshold = 0.1)

Below is a more detailed description of each input:

|  |  |  |
| --- | --- | --- |
| Input | Input Type | Description |
| f\_loc | String | String indicating the location of the file of interest |
| plot\_all | Logical | Indicator of whether to provide visualization to the user. Values : True or False |
| plots\_to\_file | Logical | Indicator of whether or not to save visuals. Currently not implemented. |
| tactile\_padding | Numeric | Value (in seconds) to right pad each tactile point event |
| auditory\_padding | Numeric | Value (in seconds) to right pad each auditory point event |
| behavior\_types | List | List outlining the behavioral states expected. Required : mom\_auditory\_types, mom\_tactile\_types, mom\_visual\_types, baby\_visual\_types, missing\_types. These categories define the sensory domains as in the original paper. See the example code above for the default values expected in each category. |
| missing\_threshold | Numeric | Value (a proportion) that indicators how much missingness is acceptable. This threshold is defined to be the percentage of the behavioral sequence that is represented by missing types defined in missing\_types. |

#### Example Usage

ber\_analyze\_file('./testfiles/Entropy\_6m - 88888HE - Event Logs.xlsx')

#### Output

ber\_analyze\_file returns a R data.frame with the following column headings

|  |  |
| --- | --- |
| Column Header | Description |
| SubjectID | Subject ID found during analysis of file |
| CanEstimateEntropy | Indicator of success or failure in estimating entropy rate |
| EntropyRate | Estimate of entropy rate between 0 and where is the total number of states |
| TotalNumberOfTransitions | The total number of events considered in the final event sequence |
| CombinedVideoDuration | Total duration of the video (endtime + final duration) |
| PercentMissing | Proportion of time represented by missing\_types category |
| AuditoryCounts | Total number of events in auditory category |
| AuditoryTotalTime | Total time representing events in auditory category |
| AuditoryAverageTime | Average duration of events in auditory category |
| VisualCounts | Total number of events in visual category |
| VisualTotalTime | Total time representing events in visual category |
| VisualAverageTime | Average duration of events in visual category |
| TactileCounts | Total number of events in tactile category |
| TactileTotalTime | Total time representing events in tactile category |
| TactileTotalTime | Average duration of events in tactile category |

### Function : ber\_analyze\_dir

The following function is another primary functions for the estimation of entropy rate. Given a directory, the function makes multiple calls to ber\_analyze\_file to analyze the excel files within that directory. Note that there is no plotting capability currently included when operating on directories.

ber\_analyze\_dir(dir\_loc,  
 tactile\_padding = 1.0,  
 auditory\_padding = 1.0,  
 behavior\_types=list(  
 "mom\_auditory\_types" = c('Vocal'),  
 "mom\_tactile\_types" = c('TouchBaby',  
 'HoldingBaby'),  
 "mom\_visual\_types" = c('ManipulatingObject'),  
 "baby\_visual\_types" = c('LookAtMomActivity'),  
 "missing\_types" = c('CantTellHolding',  
 'ActivityNotVisible',  
 'CantTellLooking')),  
 missing\_threshold = 0.1)

|  |  |  |
| --- | --- | --- |
| Input | Input Type | Description |
| dir\_loc | String | String indicating the location of the directory of interest |
| tactile\_padding | Numeric | Value (in seconds) to right pad each tactile point event |
| auditory\_padding | Numeric | Value (in seconds) to right pad each auditory point event |
| behavior\_types | List | List outlining the behavioral states expected. Required : mom\_auditory\_types, mom\_tactile\_types, mom\_visual\_types, baby\_visual\_types, missing\_types. These categories define the sensory domains as in the original paper. See the example code above for the default values expected in each category. |
| missing\_threshold | Numeric | Value (a proportion) that indicators how much missingness is acceptable. This threshold is defined to be the percentage of the behavioral sequence that is represented by missing types defined in missing\_types. |

#### Example Usage

From the directory where the files of interest are located type the following:

ber\_analyze\_dir('.')

#### Output

ber\_analyze\_dir returns a R data.frame with column headings as in ber\_analyze\_file. Each row of the data.frame represents the results from calling ber\_analyze\_file on a file.

## Input file formats

This section describes the expected input file format required for the software. It is also identifies important columns utilized in the estimation of entropy rate.

#### File Format Type

At the moment, the required format is an Excel .xlsx file. In the future this may be modifed to include .csv files, by altering specific lines of code in ber.R. Specifically, line 37 within ber\_analyze\_dir should be altered to accept different string patterns:

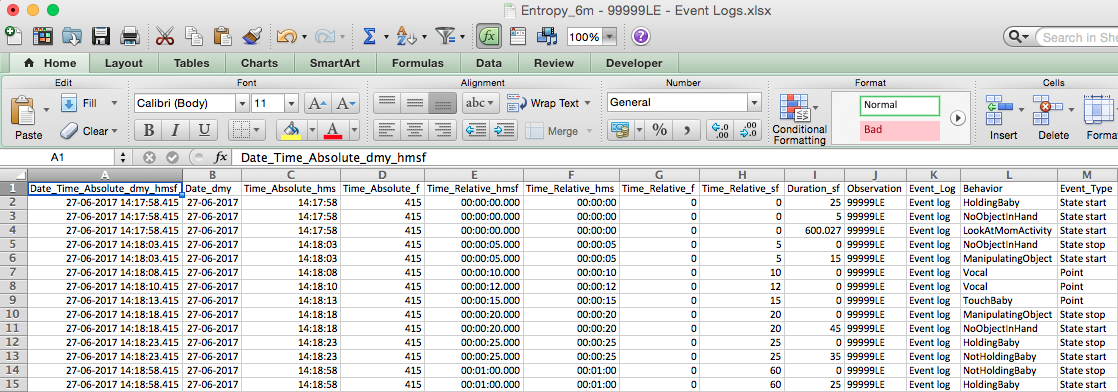
all\_files = list.files(dir\_loc, pattern="\*.xlsx")

and line 116 within ber\_analyze\_file should be changed read .csv files:

behavior\_data <- data.frame(readxl::read\_xlsx(f\_loc))

#### File Organization

Below is a visualization of a current file that successfully is processed by ccber located within the testfiles subdirectory of the source files:



Screenshot of the test file “Entropy\_6m - 99999LE - Event Logs.xlsx”

The expected column headings are:

|  |  |  |
| --- | --- | --- |
| Column Heading | Used? | Description |
| Date\_Time\_Absolute\_dmy\_hmsf | No | Absolute Time - Day-Month-Year Hour:Min:Sec.Millisec (Military Time) |
| Date\_dmy | No | Absolute Time - Day-Month-Year |
| Time\_Absolute\_hms | No | Absolute Time - Hour:Min:Sec |
| Time\_Absolute\_f | No | ??? Appears to be millisecond part ??? |
| Time\_Relative\_hmsf | No | Relative Time - Hour:Min:Sec.Millisec |
| Time\_Relative\_hms | No | Relative Time - Hour:Min:Sec.Millisec |
| Time\_Relative\_f | No | ??? Appears to be millisecond part ??? |
| Time\_Relative\_sf | Yes | Relative Time - Seconds. Used as starting point for each instance in the video |
| Duration\_sf | Yes | Relative Time - Seconds. Duration of the the instance. The time of an event is Time\_Relative\_sf + Duration\_sf |
| Observation | Yes | Typically Subject ID. Cell J1 is used as the Subject ID that is reported from ber\_analyze\_dir and ber\_analyze\_file |
| Event\_Log | No | Unknown |
| Behavior | Yes | Used to find and match against behavior\_types specified in ber\_analyze\_dir and ber\_analyze\_file |
| Event\_Type | Yes | State start indicates cells which will be used as *events*. Point instances that are included will be right padded to become events. State stop rows are ignored. |

The order of the rows should **not** matter, but it is best to be safe and keep the organization that is above for ease of processing.

### Input File Tests

when ber\_analyze\_file is run, a series of tests is performed to check the input file.