R Notebook for Emotion Rating Database Analysis

Ari Dyckovsky

Categorization Analysis

Load packages

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
library(stringr)
library(ggplot2)
library(hash)
```

```
## hash-2.2.6.1 provided by Decision Patterns
```

```
library(readr)
```

Constants

```
# Prevalence measures
HOMOSEXUAL_PREVALENCE = 0.1
HETEORSEXUAL_PREVALENCE = 0.9
HOMOSEXUAL_TAG = "_ho"

# For formatting doubles
options(digits = 2)
```

Set datapath and load `shlab.imgct`

Currently pulling from local, but will make dynamic for other machines later.

```
# Set the working directory to be part of S Drive (may make dynamic later?)
# Whilst not dynamic, change for own session if mount point is not equivalent on
# local machine
shared_dir <- "~/Projects/shlab/mounts/imgct"
package_dir <- "~/Projects/shlab"
```

```
datapath <- file.path(shared_dir, "csn_images")
imgct_package_path <- file.path(package_dir, "shlab.imgct")

# Make sure that devtools, tidyverse are installed before this call
devtools::load_all(imgct_package_path)
```

```
## Loading shlab.imgct
```

Load Database Imports

```
oasis_df <- shlab.imgct::load_imported_xlsx(datapath, "oasis_emotion_ratings")
iaps_df <- shlab.imgct::load_imported_xlsx(datapath, "iaps_emotion_ratings")
naps_be_df <- shlab.imgct::load_imported_xlsx(datapath, "naps_be_emotion_ratings")
naps_ero_df <- shlab.imgct::load_imported_xlsx(datapath, "naps_ero_emotion_ratings")
```

Handle Lack of General Ratings for NAPS ERO dataset

Provided that NAPS ERO has specified arousal and valence ratings for four groups, but no general rating of arousal and valence, we must choose an averaging strategy. We will attempt a weighted average with rough prevelance estimates, based on these four groups:

1. HoF (Homosexual Female)
2. HoM (Homosexual Male)
3. HeF (Heterosexual Female)
4. HeM (Heterosexual Male)

Constants for prevalence are determined above, and below is a function for weighted ratings based upon the columns of NAPS ERO where "_ho" and "_he" are tags for homosexual and heterosexual, respectively.

```
weighted_rating <- function(ratings) {
  columns <- names(ratings)

  # Weights are distributed to tags such that:
  #   - FALSE = HETEROSEXUAL_PREVALENCE
  #   - TRUE = HOMOSEXUAL_PREVALENCE
  weights <- c(
    HETEORSEXUAL_PREVALENCE,
    HOMOSEXUAL_PREVALENCE
  )

  tags <- stringr::str_detect(columns, HOMOSEXUAL_TAG)
  weights_vec <- weights[1 + tags]

  weighted.mean(
    ratings,
    weights_vec
  )
}
```

Mutate new columns of general arousal and valence ratings for NAPS ERO.

```
naps_ero_df <- naps_ero_df %>%
  dplyr::mutate(
    arousal = dplyr::select(., starts_with("arousal")) %>%
      purrr::pmap_dbl(
        .,
```

```
        ~ weighted_rating(c(...))
      ),
    valence = dplyr::select(., starts_with("valence")) %>%
      purrr::pmap_dbl(
        .,
        ~ weighted_rating(c(...))
      )
  ) %>%
  dplyr::relocate(
    c(arousal, valence),
    .after = image_id
  )
```

Merge dataframes

```
all_emotion_ratings_df <- dplyr::bind_rows(
    list(
      oasis_df,
      iaps_df,
      naps_be_df,
      naps_ero_df
    ),
    .id = "db_id"
  ) %>%
  dplyr::select(db_id, image_id, arousal, valence) %>%
  dplyr::mutate_at(
    vars(db_id),
    ~dplyr::recode(
      .,
      "1" = "OASIS",
      "2" = "IAPS",
      "3" = "NAPS_BE",
      "4" = "NAPS_ERO"
    )
  )

all_emotion_ratings_df
```

```
## # A tibble: 3,621 x 4
##    db_id image_id            arousal valence
##    <chr> <chr>                 <dbl>   <dbl>
##  1 OASIS OASIS_acorns_1.jpg     2.35    4.69
##  2 OASIS OASIS_acorns_2.jpg     2.23    4.52
##  3 OASIS OASIS_acorns_3.jpg     2.31    4.75
##  4 OASIS OASIS_alcohol_1.jpg    2.87    4.69
##  5 OASIS OASIS_alcohol_2.jpg    3       4.25
##  6 OASIS OASIS_alcohol_3.jpg    2.35    4.17
##  7 OASIS OASIS_alcohol_4.jpg    2.87    4.77
##  8 OASIS OASIS_alcohol_5.jpg    3.55    5.33
##  9 OASIS OASIS_alcohol_6.jpg    2.64    4.54
## 10 OASIS OASIS_alcohol_7.jpg    3.13    4.48
## # ... with 3,611 more rows
```

```r
p <- ggplot(all_emotion_ratings_df, aes(x=arousal, y=valence, color=db_id)) +
  geom_point(alpha = 0.5)

p <- p + labs(
    title = "Valence vs. Arousal of Database Images",
    x = "Arousal",
    y = "Valence",
    color = "Source"
  ) +
  theme_classic()

p
```

Save to results as general TSV of arousal and valence ratings

```r
readr::write_tsv(
  all_emotion_ratings_df,
  file.path(datapath, "results", "general_emotion_ratings.tsv"),
  append = FALSE,
  col_names = TRUE
)
```