FilterAge

Sho

May 10, 2017

Age filter for MetaLab

Script to filter out studies from all datasets that do not at least test two age groups on the samephenomenon of interest

Strategy (by dataset)

1. Automatic filtering

1.1 Order dataset by all identified relevant columns that could vary between rows 1.2 Only keep those rows that have the same content in the relevant columns for different age-groups

2. Manual correction

While I tried to include all relevant columns (by adding additional ones to spec),not always is the difference between two rows deductible. For instance there might be two rows of 4m-olds in a dataset that have identical columns. In that case, something that was varied in the study is not reflected in metalab (probably because it was not vital to the meta-analysis). If the number of rows per age-group still matched up, e.g. 2 rows for 4m and 2 rows for 10m, I included the data without further inspection, assuming the same unknown dimension was varied pairwise. However, for cases where number of rows did not match up, e.g. 3 rows for 4m and 2 rows for 10m, I went back to original paper and identified the row that did not have an age-matched equivalent effect size. For those cases, there is a second documented filtering sequence in the script. I have not done this yet for InWordDB for sheer volume and lack of expertise.

Decisions taken

I decided to define “same age” as the same age in months. In order to do that and to avoid counting samples that have two slightly different age groups in days (e.g. 112 and 114), I needed to round age somehow. Depending on how exactly age groups fell, for some studies floor(mean_age_months) worked better, for others round(mean_age_months). I decided to use the same rounding strategy for all studies (round()), not manually correcting studies where this lead to artificial inclusion of rows (e.g. 6.4m –> 6m, 6.6m –>7). This does not affect a large number of rows (documented below).

```
source("C:/Users/Sho/Documents/GitHub/metalab - Copy/dashboard/global.R", chdi
r = TRUE)
```

```
##
## Attaching package: 'shinydashboard'
```

```
## The following object is masked from 'package:graphics':
##
##     box
```

```
##
## 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
```

```
##
## Attaching package: 'purrr'
```

```
## The following objects are masked from 'package:dplyr':
##
##     contains, order_by
```

```
##
## Attaching package: 'langcog'
```

```
## The following object is masked from 'package:base':
##
##     scale
```

```
## Warning in readLines(input, encoding = "UTF-8"): incomplete final line
## found on '../metadata/spec.yaml'
```

```
## Warning in bind_rows_(x, .id): binding factor and character vector,
## coercing into character vector
```

```
## Warning: Grouping rowwise data frame strips rowwise nature
```

```
## Joining, by = "dataset"
```

```
## Joining, by = "dataset"
```

```
library(dplyr)

all_data <- all_data %>%
  mutate(age_months_floor = floor(mean_age_months),age_months_round = round(mea
n_age_months,0))

#for each dataset, find studies that test at least 2 age groups on critical thi
ng


#filtering is fine
gaze_following_age <- all_data %>%
  filter(short_name=="gaze_following")%>%
  group_by(study_ID,infant_type,cue_type)%>%
  mutate(count_ages = length(unique(age_months_round))) %>%
  filter(count_ages > 1)


#filtering is fine
idspref_age <- all_data %>%
  filter(short_name=="idspref")%>%
  group_by(study_ID,infant_type,speaker,speech_type,dependent_measure)%>%
  mutate(count_ages = length(unique(age_months_round))) %>%
  filter(count_ages > 1)


#filtering is fine
#2nd filtering round to remove studies not captured by filtering:
#Mugitani2009 expt 2 uses English stimuli (as opposed to Japanese in expt 3 an
d 4); only 1 age-group
#Sato2010 expt 4 uses natural stimuli (as opposed to edited ones in expts 1
-3); only 1 age-group
#Age rounding errors: Tsujisubmitted
inphondb_age <- all_data %>%
  filter(short_name=="inphondb-native"|short_name=="inphondb-nonnative")%>%
  group_by(study_ID,native_lang,infant_type,contrast_sampa)%>%
  mutate(count_ages = length(unique(age_months_round))) %>%
  filter(count_ages > 1)

inphondb_age <- inphondb_age%>%
  mutate(stud_expt = paste(study_ID,expt_num)) %>%
  filter(stud_expt!= "Mugitani2009 2",stud_expt!= "Sato2010 4")


#manual filtering still needs to be done on following studies:
# MattysJusczyk2001
# HoustonJusczyk2003
# Nazzietal2006
```

```
# WillitsSeidenbergSaffran2009
# Kim2012
# NishibayashiGoyetNazzi2014
# HaryuKajikawa2016
# Flocciaetal2016
#Age rounding errors: HoustonSantelmann2004Jusczyk,ShiCutlerWerkerCruickshank20
06, Kim2012
inworddb_age <- all_data %>%
  filter(short_name=="inworddb")%>%
  group_by(study_ID,native_lang,infant_type,Linguistic,words_to_passage,edge_al
ignment,familiarization_voice,familiarization_register,familiarization_affect,t
est_voice,test_register,indexical)%>%
  mutate(count_ages = length(unique(age_months_round))) %>%
  filter(count_ages > 1)


#filtering is fine
#2nd filtering round to remove studies not captured by filtering:
#robinson2007 expt 2 varied the stimuli and is only conducted on 1 age-group
labadv_age <- all_data %>%
  filter(short_name=="labadv")%>%
  group_by(study_ID,infant_type,expt_condition,audio_condition,condition_type)%
>%
  mutate(count_ages = length(unique(age_months_round))) %>%
  filter(count_ages > 1)

labadv_age <- labadv_age%>%
  mutate(stud_expt = paste(study_ID,expt_num)) %>%
  filter(stud_expt!= "robinson2007 2")


# filtering is fine
#Age rounding errors:frank2015,horst2008
mutex_age <- all_data %>%
  filter(short_name=="mutex")%>%
  group_by(study_ID,infant_type,expt_condition,object_stimulus)%>%
  mutate(count_ages = length(unique(age_months_round))) %>%
  filter(count_ages > 1)

#filtering is fine
#2nd filtering round to remove studies not captured by filtering:
#chambers2011 expt 1 only tests one age-group
phonotactics_age <- all_data %>%
  filter(short_name=="phonotactics")%>%
  group_by(study_ID,native_lang,infant_type,expt_condition,rule.type,repetition
s,types,tokens)%>%
  mutate(count_ages = length(unique(age_months_round))) %>%
  filter(count_ages > 1)
```

```
phonotactics_age <- phonotactics_age%>%
  mutate(stud_expt = paste(study_ID,expt_num)) %>%
  filter(stud_expt!= "1002 1")


#filtering is fine (no candidate data points)
pointing_concurrent_age<- all_data %>%
  filter(short_name=="pointing_concurrent")%>%
  group_by(study_ID,infant_type,motive)%>%
  mutate(count_ages = length(unique(age_months_round))) %>%
  filter(count_ages > 1)


#filtering is fine (no candidate data points)
pointing_longitudinal_age<- all_data %>%
  filter(short_name=="pointing_longitudinal")%>%
  group_by(study_ID,infant_type,motive)%>%
  mutate(count_ages = length(unique(age_months_round))) %>%
  filter(count_ages > 1)


#filtering is fine
#Age rounding error: yoshida2009
sounds_age<- all_data %>%
  filter(short_name=="sounds")%>%
  group_by(study_ID,infant_type,expt_condition,native_lang,test_lang)%>%
  mutate(count_ages = length(unique(age_months_round))) %>%
  filter(count_ages > 1)


#filtering is fine
symbolism_age<- all_data %>%
  filter(short_name=="symbolism")%>%
  group_by(study_ID,infant_type,expt_condition,native_lang,method,expt_conditio
n2,word_round,word_spiky)%>%
  mutate(count_ages = length(unique(age_months_round))) %>%
  filter(count_ages > 1)


#filtering is fine
word_recognition_age <- all_data %>%
  filter(short_name=="word_recognition")%>%
  group_by(study_ID,infant_type,native_lang)%>%
  mutate(count_ages = length(unique(age_months_round))) %>%
  filter(count_ages > 1)
```