Memory content

R. Cassano-Coleman

2025-02-25

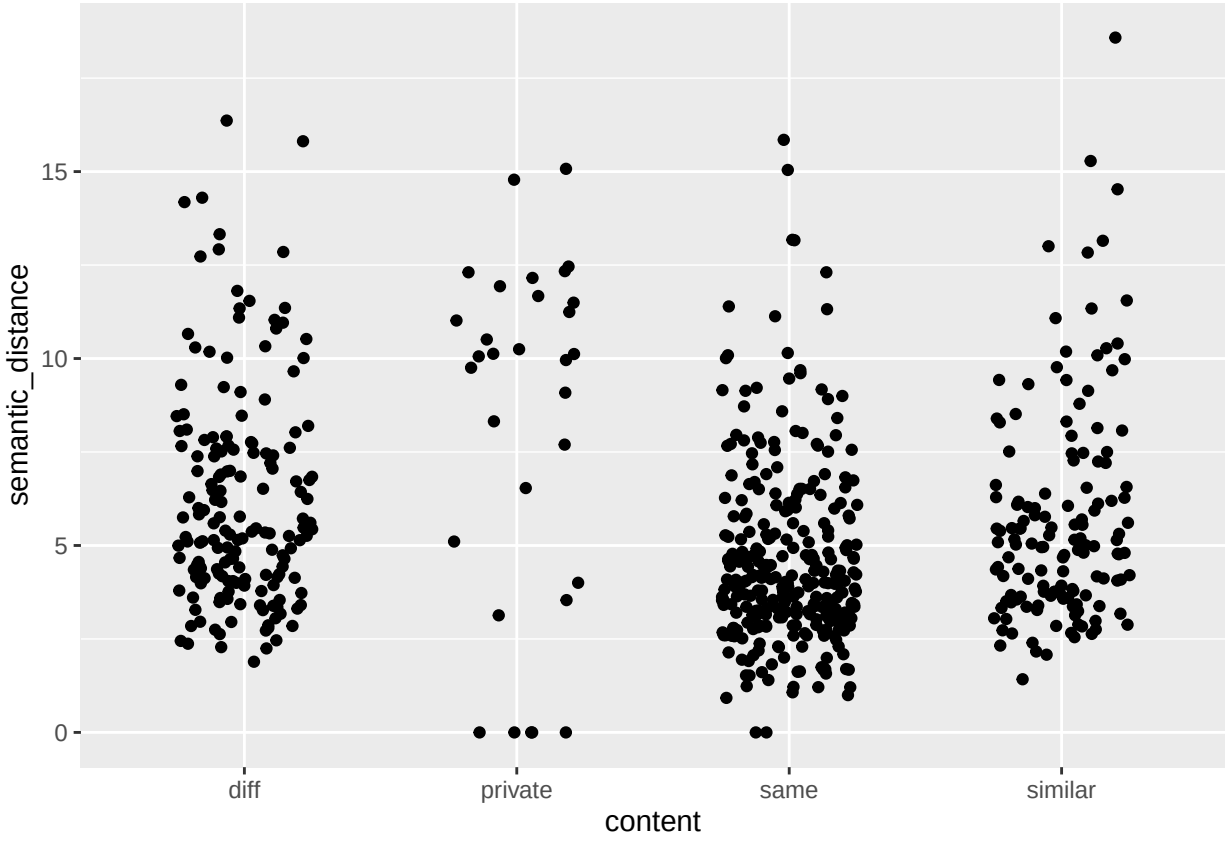
This notebook looks at memory content for pairs of original-cover memories, as predicted by perceptual similarity of the clips evoking the memories.

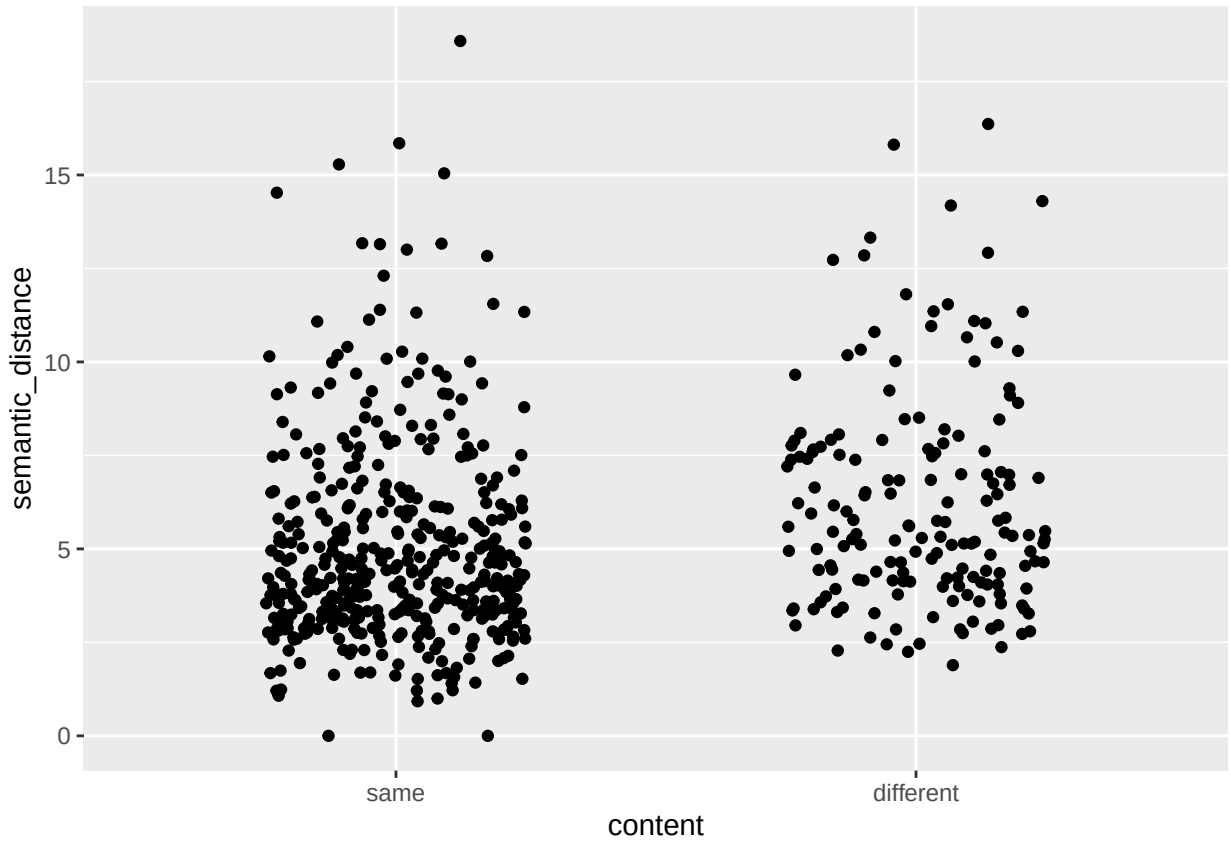
Load the data.

Load just the features and similarity for song-wise analysis.

Make same/different coding a factor. (Filter out private, lump similar in with same.)

Look at same/different coding vs semantic distance from sentence embeddings.





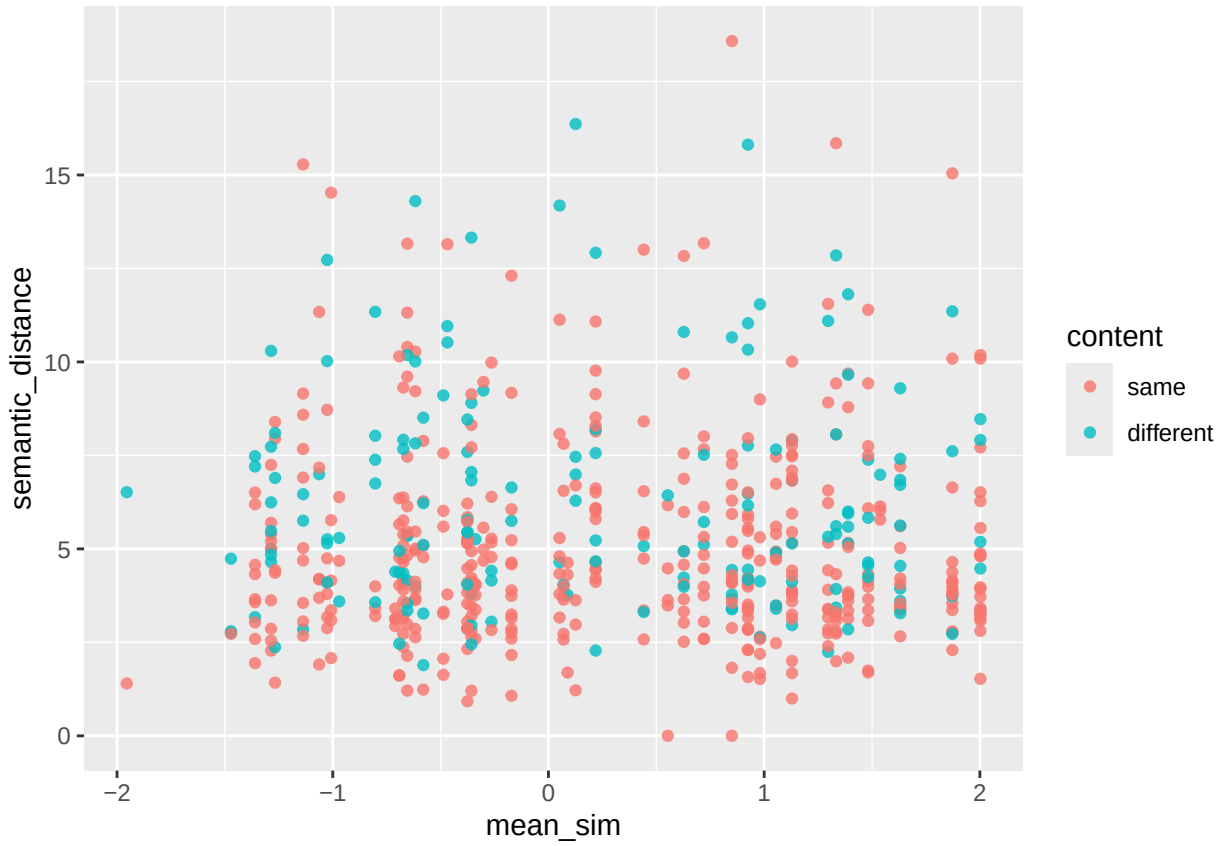
Predict semantic distance from perceptual similarity.

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: semantic_distance ~ mean_sim + (1 | internal_id) + (1 | song_id)
##    Data: data_factor
##
## REML criterion at convergence: 2917.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1880 -0.6487 -0.2439  0.4158  4.7441
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  internal_id (Intercept) 0.5902   0.7683
##  song_id     (Intercept) 0.0000   0.0000
##  Residual                7.2411   2.6909
## Number of obs: 598, groups:  internal_id, 79; song_id, 50
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)   5.2672     0.1569  56.0995  33.576   <2e-16 ***
## mean_sim     -0.0219     0.1114 580.9586  -0.197    0.844
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
```

```
##          (Intr)
## mean_sim -0.215
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

Predict content (same/different) from perceptual similarity.

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: content ~ mean_sim + (1 | internal_id) + (1 | song_id)
##    Data: data_factor
##
##      AIC      BIC   logLik deviance df.resid
##    695.1    712.6   -343.5    687.1      594
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.2147 -0.6228 -0.4389  1.0171  2.7319
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  internal_id (Intercept) 0.8618   0.9283
##  song_id     (Intercept) 0.0000   0.0000
## Number of obs: 598, groups:  internal_id, 79; song_id, 50
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.10555    0.17114  -6.460 1.05e-10 ***
## mean_sim    -0.13664    0.09817  -1.392    0.164
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr)
## mean_sim -0.132
## optimizer (Nelder_Mead) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```



Visualize.

Can we predict semantic distance just for same memories? (expect no)

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: semantic_distance ~ mean_sim + (1 | internal_id) + (1 | song_id)
##    Data: .
##
## REML criterion at convergence: 2061.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.9963 -0.6378 -0.2730  0.4244  5.0312
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  internal_id (Intercept) 0.15932  0.3992
##  song_id     (Intercept) 0.09654  0.3107
##  Residual                6.88808  2.6245
## Number of obs: 429, groups:  internal_id, 74; song_id, 50
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)  4.960573   0.154089 29.284616  32.193   <2e-16 ***
## mean_sim     0.003126   0.138054 34.835421   0.023    0.982
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Correlation of Fixed Effects:
##          (Intr)
## mean_sim -0.256
```

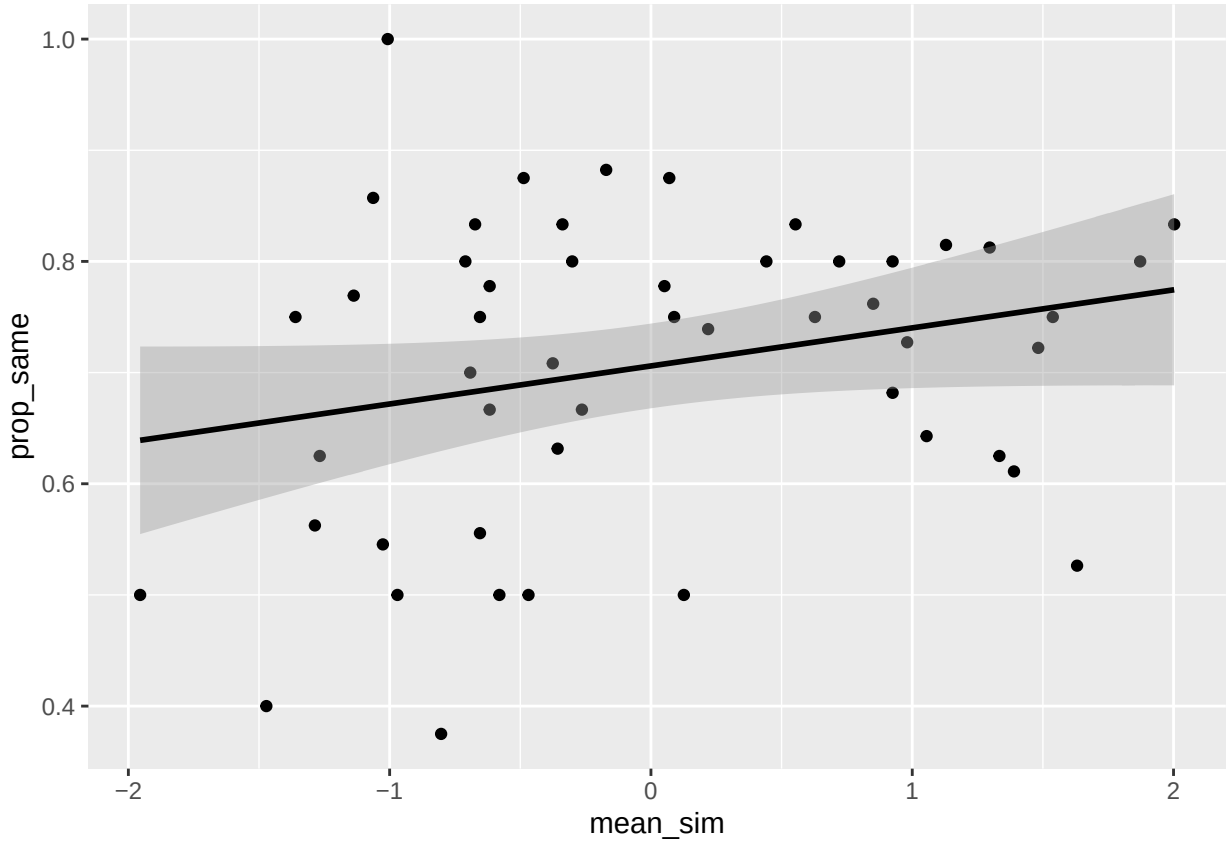
Can we predict semantic distance just for different memories? (expect yes??)

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: semantic_distance ~ mean_sim + (1 | internal_id) + (1 | song_id)
##    Data: .
##
## REML criterion at convergence: 832.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.4646 -0.7193 -0.2074  0.4978  3.3808
##
## Random effects:
##  Groups      Name        Variance  Std.Dev.
##  internal_id (Intercept) 1.463e+00 1.210e+00
##  song_id     (Intercept) 1.102e-09 3.319e-05
##  Residual                7.020e+00 2.650e+00
## Number of obs: 169, groups:  internal_id, 49; song_id, 49
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)   6.12401    0.29155  43.36182  21.005   <2e-16 ***
## mean_sim      0.06805    0.21015 162.94987   0.324    0.746
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr)
## mean_sim -0.156
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

Compute proportion of same or similar memories per song.

```
##
## Call:
## lm(formula = prop_same ~ mean_sim, data = data_by_song)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.30351 -0.09942  0.02410  0.08765  0.32848
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.70599    0.01895  37.249   <2e-16 ***
## mean_sim     0.03422    0.01915   1.787   0.0802 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.134 on 48 degrees of freedom
## Multiple R-squared:  0.0624, Adjusted R-squared:  0.04287
```

```
## F-statistic: 3.194 on 1 and 48 DF,  p-value: 0.0802
```



Visualize.