

**ana3-BG-Entropie (ana3-BG-transition-predictability)**

Analyse der Vorhersagbarkeit von Akkordübergängen

- Vgl. der Entropie einer Akkordstichprobe aus allen Akkorden mit Akkorden mit bestimmter Eigenschaft (alle aus demselben Datensatz)
- Vgl. der Entropie von Akkordstichproben eines Datensatzes mit Akkorden eines anderen
  - Normal, Differenzierung aller Akkordtoken (wie in ana3a)
  - Reduktion auf Stufen (I, II, ..)

**ana3a-BG-transition-predictability.R**

Dieses Programm führt einen Hypothesentest über die Vorhersagbarkeit bestimmter Akkordübergänge durch. Dies über Untersuchung der normalisierten bedingten Entropie. Für einen Übergang  $A \rightarrow B$  erfasst Sie die Unsicherheit über den folgenden Akkord  $B \in Z$  der Zielakkordmenge unter der Annahme, dass  $A$  bekannt ist. Sie lässt sich als Erwartungswert der Unsicherheit für den Übergang  $A \rightarrow \dots$  verstehen und ist definiert als

$$\bar{E}(B = c_i \in Z | A = c) = \frac{\sum_{i=1}^n p(c \rightarrow c_i) \times \log_2 \frac{1}{p(c \rightarrow c_i)}}{\log_2 |Z|}$$

Es soll untersucht werden, ob die norm. bed. Entropie von Akkorden mit bestimmten Eigenschaften (Vorhalte, Umkehrungen, Alterationen, Orgelpunkte, Tonikalisierungen) von der norm. bed. Entropie zufälliger Akkordstichproben abweicht.

Für den Hypothesentest können folgende Hypothesen formuliert werden:

$H_0$ : Die Entropie von Akkorden mit Eigenschaft ist nicht signifikant anders.

$H_1$ : Die Entropie von Akkorden mit Eigenschaft ist grösser oder kleiner.

Liegt der p-Wert  $\pi_M = \frac{2}{k} \times \min(u_M, o_M)$  über der gewählten Irrtumswahrscheinlichkeit kann  $H_0$  angenommen werden, liegt er darunter, kann  $H_1$  angenommen werden.

Das Programm ist folgendermassen aufgebaut:

(0) Zuerst der Untersuchungsgegenstand und Parameter festgelegt:

(0a) Die zu untersuchenden Dateien

(0b) Die zu untersuchende Akkordeigenschaft

```

57  chord_feature <- "Vorh"
..
73  feature_filter <- function(i){
74      if (!is.na(specific_harmonic_tab$changes[i]) &&
          specific_harmonic_tab$changes[i] != "" &&
          specific_harmonic_tab$changes[i] != 0 && !any(grepl("\\+",
          specific_harmonic_tab$changes[i]))) {
75          1
76      } else {
77          0
78      }

```

(0c) Es besteht die Möglichkeit, nach Tongeschlecht der Tonartabschnitten zu filtern

(0d) Hier können Umfang und Anzahl der Stichproben für den Hypothesentest festgelegt werden

```

66  proben_umfang <- 100 # Grösse (Umfang) einer Probe
67  anz_proben <- 10000 # Anz. Proben

```

(A-B) Dann ana1a-UG-rank-freq.R genutzt, um ein Dataframes zu erstellen, welches jedem Akkord (chord) einen Häufigkeitsrang (rank) und die relative Häufigkeit (relative\_frequency) zuordnet

(A) 4a-combined-data.R um eine grosse .tsv Tabelle über alle Stücke und Sätze zu erhalten

(B) Erstellt ein Rang-Häufigkeits-Dataframes

(1) Nimmt Stichproben aus allen Akkorden und berechnet die norm. bed. Entropie für diese

(1a) Erstellt eine umfassende Matrix aller Akkordübergänge

```
185 all_transition_matrix <- matrix(0, nrow = length(all_chords),
  ncol = length(all_chords), dimnames = list(all_chords, all_chords))
188 for (file in tsv_files) {
  ..
  194   for (i in 1:(length(specific_harmonic_tab$chord) - 1)) {
  195     current_chord <- specific_harmonic_tab$chord[i]
  196     next_chord <- specific_harmonic_tab$chord[i + 1]
  ..
  200     if (current_chord %in% all_chords && next_chord %in% all_chords) {
  201       all_transition_matrix[current_chord, next_chord] <-
         all_transition_matrix[current_chord, next_chord] + 1}
```

(1b) Berechnet und erstellt ein Dataframe, welches jedem Akkord (chord) seine bed. Entropie (entropy) und seine norm. bed. Entropie zuordnet (norm\_entropy)

```
226 norm_konst <- log2(ncol(all_transition_matrix))
  ..
  245 if (transition_freq == 0) {
  246   summand <- 0
  247 } else {
  249   summand <- transition_freq * log2(1/transition_freq)
  250 }
  251 entropy <- entropy + summand
  ..
  257 entropy_df$norm_entropy <- entropy_df$entropy / norm_konst
```

(1c) Wählt Stichproben von Akkorden und berechnet den Mittelwert der Entropie der Stichprobenakkorde. Eintrag in Tabelle für jede Stichprobe. Dies mit der 'sample' Funktion, mit 'replace = T' erlauben wir Wiederholungen, durch 'prob' werden die Stichproben wahrscheinlichkeitsgewichtet gewählt.

```
270 probe <- sample(
  271   all_chords,
  272   proben_umfang,
  273   replace = T,
  274   prob = chord_freq_df$relative_frequency)
```

(2) Berechnet die mittlere norm. bed. Entropie aller Akkorde mit Akkordeigenschaft

(2a) Erstellt eine Matrix aller Akkordübergänge, die von einem Akkord mit Akkordeigenschaft ausgehen

(2b) Berechnet und erstellt ein Dataframe, welches jedem Akkord (chord) mit Akkordeigenschaft seine bed. Entr. (entropy) und norm. bed. Entr. zuordnet (norm\_entropy)

(2c) Berechnet den Erwartungswert (wahrscheinlichkeitsgewichteter Mittelwert)

```
377 feat_entropy_df$rel_freq <- rowSums(feat_transition_matrix)/
  sum(feat_transition_matrix)
379 entropy_expected_feat <- sum(feat_entropy_df$norm_entropy *
  feat_entropy_df$rel_freq)
```

(C) Berechnet den p-Wert

```
379 u_M <- sum(sample_df$sample_mean < entropy_mean_feat, na.rm = T)
380 o_M <- sum(sample_df$sample_mean > entropy_mean_feat, na.rm = T)
381 p_M <- (2/anz_proben) * min(u_M, o_M) # p-Wert
```

(D) Zuletzt ist es möglich ein Diagramm mit den mittl. norm. bed. Entr. der Stichproben und der mittl. norm. bed. Entr. der Akkorde mit Akkordeigenschaft ausgeben zu lassen.

ana3b-BG-compare-entropy.R

Dieses Programm funktioniert ähnlich wie `ana3a-BG-transition-predictability.R`, nur werden zwei Hypothesentests durchgeführt. Einmal ob die Entropie des Basisdatensatz mit der des Vergleichsdatensatz abweicht und umgekehrt. Es lässt sich ebenfalls ein Diagramm ausgeben.

### `ana3b-BG-compare-entropy-reduced.R`

Dieses Programm funktioniert ähnlich wie `ana3b-BG-compare-entropy.R`, nur werden nicht die einzelnen Akkordzeichen betrachtet sondern auf die Stufe reduziert. Dazu wird statt dem Feld 'chord' in den Datensätze das Feld 'numeral' genutzt, in welchem ausschliesslich die Stufe (ohne Berücksichtigung von Akkordmerkmalen wie Umkehrungen etc.) erfasst ist. Es lässt sich ebenfalls ein Diagramm ausgeben.

<sup>1</sup> Die Tabellen müssen, damit sie von den Skripten erkannt und eingelesen werden können, wie folgt benannt werden:

`[KOM]-[A][xx]-M[x].tsv`

wobei KOM für das Komponistenkürzel steht (LVB für L. v. Beethoven, WAM für W. A. Mozart), A für die Stückart (S für Sonate, Q für Streicherquartett) mit der üblichen Nummerierung (xx aus 00-99) und M für den Satz mit Nummer (x aus 0-9)

Die Datei zum dritten Satz der ersten Beethoven-sonate heisst also: `LVB-S01-M3.tsv`