2018 SDML HW3

Prof. Shou-De Lin
TA: Yen-Ting Lee, Liang-Hsin Shen

Introduction: Generate a lyric sentence based on the previous lyric sentence

- Task 0
    - Corpus: Raw Lyrics (simplified character)
        - Will be released before 11/12
    - Goal: Training it with two sequene-2-sequence models
        - one without attention, one with attention

Task 0 (Warm up)

- Requirements
  - predict.sh <input file path> <output file path>
  - Test script will be announced later
    - You will need to run the script in class
    - Time limit is 20 minutes
  - No Kaggle this time
  - Third-party packages can be used
  - Cannot use extra data

Data Format

- Chinese using simplified characters
- Lyric_no, Lyric_segments
- Sample data

    lyric_no,lyric_segments
    0, 颜色 多 美丽
    0, 而 我 的 星球 自行 旋转
    (The lyrics from the same song have same lyric_no)
    (lyric_segments is separated by single whitespace)

Submission

- Need to use our script to produce the result in class
  - In the class on 11/22 (first break 20 min)
  - A different testing set will be downloaded during that period
    - The testing set would be about 5000 sentences
    - No header, no index, segments separated by white space
- Submit the result to Ceiba soon after the result is produced
  - No late submission is allowed
- Output Format
  - Output sentence: segments joint by white spaces
  - Each line should contain a generated output sentence, i.e. the number of output lines should match that of input lines
  - No header is needed

Grading Policy

- 5% of HW3
- You can get full point as long as you submit your output with correct format on time
- Content is not considered in this part of homework, but in HW3-1.
- We will feedback whether the quality of your result is acceptable in the upcoming homework.

TA Information

- 沈亮欣 r06922011@ntu.edu.tw
- 李彥霆 r06922008@ntu.edu.tw
- TA hour: Monday 10:00 ~ 11:00 @ 302
- If you ask something we have already mentioned in class, we won't reply.