BERT

2019年8月15日 14:21

- Bidirectional -> Deep Bidirectional
  - Transformer only, self attention
  - BERT based on AE model can’t fix [MASK] mismatch
    - XLNet based on AR model
- Transformer for music
  - GPT-2: MuseNet
    - A tokenized method for music
- VideoBERT Tasks:
  - Text-only, video-only, linguistic-visual alignment
- Dataset preprocessing
  - Audio to midi
  - Vector content
  - Tokenized, granularity

2019-08-16

2019年8月15日 21:37

230k hours for VideoBERT

YouTube (crawler & 8M):

- Official MVs
- Mashup for songs & films

TikTok (crawler):

- Crawl by Music/User
- Music clip with dance or short plot, which means one clip corresponding to multiple videos
- User who focus on dance or short music video

2019-08-19

2019年8月18日 21:35

Clips selection criteria:

☑ 1. Fitting
☑ 2. Music quality
3. Sound effect & vocal noise?
4. Ez to transcript or sythesize?

- Q1: How to discriminate music soundtrack?
- Q2: Could we erase nosie like sound effect & human voice? NO
- Q3: Fusing speech & music?

1. 8M subset
   a. Music video (116098)
   b. Trailer (59695)
   c. Advertisement (4898 + 2686)
2. MV selection
   ☑ a. Story
   □ b. Performance
   ☑ c. Dancing
   □ d. Animation
3. Movie clip selection
   a. Home-made videos -> 1000 movie clips
4. Mashup keywords

bilibili	混剪 45.9w 1000	踩点 25.5w 1000	
YouTube	Movie mashup 5w 300	Trailers mashup 5w 300	Movie dance mashup 5w 150

5. Advertisement
   a. Q1

- Using one model to learn composition rules, like tempo, beats, velocity
- Another model to learn how to make rules into real music