Text2Scene: Discriminator & Brutal-force search

- Explore feature engineering
- Train and optimize the discriminators
- Implement brutal-force generator
- Evaluation and error analysis
- Detailed report available
- Demo available

Text2Scene: Discriminator & Brutal-force search

- The entire dataset will be available tomorrow

Text2Scene: Discriminator & Brutal-force search

- Explore feature engineering
  - Text encoder: tokenizer -> lemmatizer -> tf-idf
  - Picture encoder: categories -> binary vector + paired similarity
  - Joint encoder: paired similarity

- Notes
  - Category keywords can be idf reweighted as well
  - Scaled similarities present dissimilar features

Text2Scene: Discriminator & Brutal-force search

- Consistency discriminator
  - Input: text – picture
  - Output: consistent or inconsistent
  - Encoder: Text; Picture; Joint
  - Dataset: Triplets: text – picture; random text – picture; text– random picture
  - Training: Binary classification – logistic regression
  - Metric: Average precision

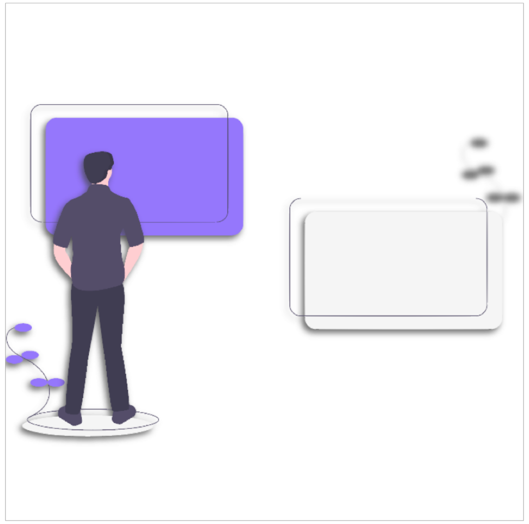
Text2Scene: Discriminator & Brutal-force search

- Reasonability discriminator
  - Input: picture
  - Output: reasonable, unreasonable
  - Encoder: Picture
  - Dataset: Pair: text – real picture; text – fake picture
  - Training: Binary classification – logistic regression
  - Metric: Average precision

- Notes
  - interpretability

Text2Scene: Discriminator & Brutal-force search

- Generator
  - Input: text
  - Output: picture
  - Objective: maximize weighted sum of scores given by discriminators
  - Method: Brutal-force search
  - Metric: category-level average F1 score

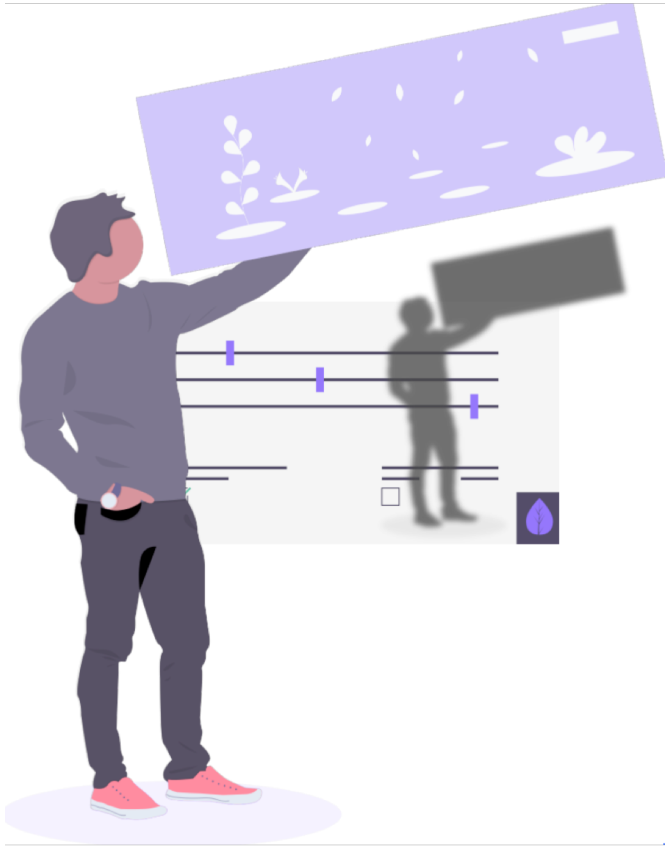


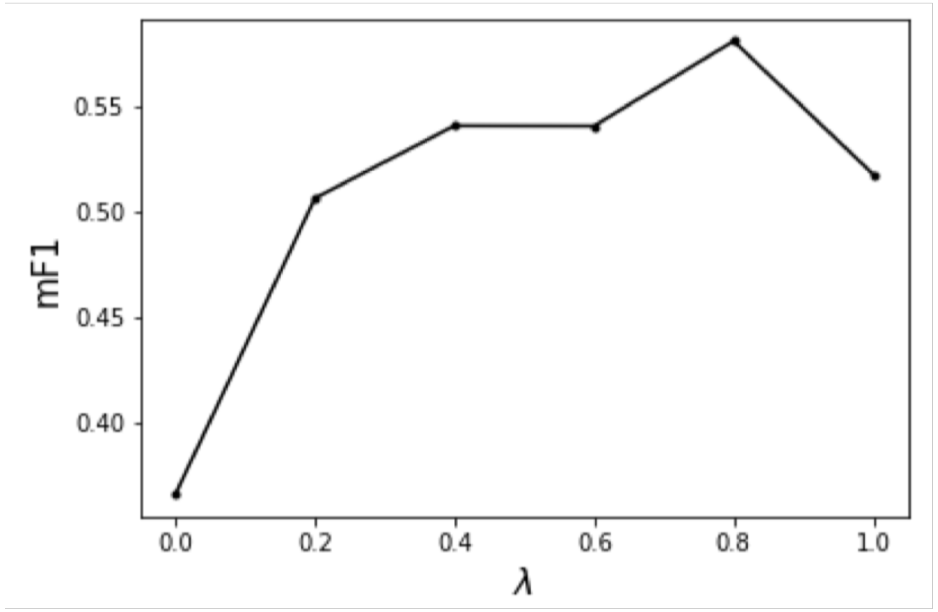
A3212 Character – Person – Stand - Back
A2121 Surrounding – Room – Abstraction - Chart

[0, 1, 1, 0 | 0, …, 1, …,1, …, 1, …, 0, …. | …, 1, …,1, …, 1, …, 0, ….]

Text2Scene: Discriminator & Brutal-force search

- Results
  - Input: “A man is showing a chart”
  - Output:





Text2Scene: Discriminator & Brutal-force search

- Conclusion
  - Joint features are critical to the decisions
  - Context information may not be that important as for current short sentences

- Problem
  - Imbalance
    - Some highly frequent keywords dominate the decision through single-token features, and give rise to fallaciously high performance
    - A majority of keywords are too rare to be learned with non-trivial weights
  - Similarities are constrained to keywords with the same POS
  - No joint feature for bigrams
  - No extra care for unseen words
  - Random states strongly affect the performance, which probably attributes to the dataset

Text2Scene: Discriminator & Brutal-force search

- To-do
  - K-folds validation to yield a more robust performance index
  - Optimize a congregated loss function rather than train two discriminators separately
  - Explore more methods to encode joint information
  - Replace rare tokens with their hypernyms to increase recurrence
  - Replace unseen tokens with their most similar keywords in the vocabulary
  - Pixel-level picture encoder

Text2Scene: Discriminator & Brutal-force search

- Thought
  - Sophisticated embedding methods, e.g. word2vec, multi-layer perceptron, and subsequent models, e.g. sequence, GAN will produce better results?

Text2Scene: Discriminator & Brutal-force search

- Requirement
  - Dataset