CONSTRUCT MINING PIPELINE CHEAT SHEET

— — —

EXERCISE 1: SENTENCE EMBEDDINGS

1. from sentence_transformers import SentenceTransformer
2. model = SentenceTransformer('stsb-mpnet-base-v2')
3. data = pd.read_csv('data/strategies_raw_en.csv')
4. strategies = data.text.to_list()
5. strategy_embeddings = model.encode(strategies, show_progress_bar=True)

EXERCISE 2: MEASURING ITEM BIAS

1. Look at the table for the item bias statistic (next page). For which items does a bias exist and in which direction?

Bias exists for all items, because all confidence intervals exclude zero. The bias is in the expected direction, that is emotion regulation suggestions lie closer to their respective vignette than to all other vignettes.

EXERCISE 3: REDUCING ITEM BIAS

1. [Please keep in mind that there are multiple solutions and that the suggested solution is not ideal. The masked words are directly translated from German, which is the original language of the study.]

Vignette texts

1) Your workmate fails to deliver an important piece of **information** on time, causing you to fall behind **schedule** also.

2) You are accepted for a highly sought after **contract**, but have to fly to the location. You have a **phobia of flying**.

3) You answer the phone and hear that close relatives are in **hospital** critically ill.

4) You find out that some **members** of your social **sports team** have been saying that you are not a very good player.

5) You have just gone back to **university** after a lapse of several years. You are surrounded by younger **students** who seem very confident about their ability and you are unsure whether you can compete with them.

6) A demanding **client** takes up a lot of your time and then asks to speak to your boss about your performance. Although your boss assures you that your performance is fine, you feel upset.

7) Your access to essential **resources** has been delayed and your work is way behind **schedule**. Your **progress report** makes no mention of the lack of resources.

8) You are having a large **family gathering** to **celebrate** you moving into your new home. You want the day to go smoothly and are a little nervous about it.

9) You and your colleague usually go to a **cafe** after the working week and chat about what's going on in the company. After your colleague's job is moved to a different **section** in the company, he/she stops coming to the cafe. You miss these Friday talks.

2. masks = {
1: ['information', 'schedule'],
2: ['contract', 'phobia of flying'],
3: ['hospital'],
4: ['members', 'sports team'],
5: ['university', 'students'],
6: ['client'],
7: ['resources', 'schedule', 'progress report'],
8: ['family', 'gathering', 'celebrate'],
9: ['cafe', 'section']
}

EXERCISE 4: REDUCING DIMENSIONALITY AND CLUSTERING

2. Ns=[10, 30, 50]

EXERCISE 5: CHECKING ROBUSTNESS

1. pd.Series(no_clusters).value_counts()
2. no_clusters == 37

EXERCISE 6: VALIDATING CLUSTERS

2. ax.axhline(y=0.75); Clusters 1, 9 and 23 fall below the threshold.

EXERCISE 7: INTERPRETING CLUSTERS

1. See below.

4: Distraction
5: Relaxation techniques
9: Suppression
17: Withdrawal
20: Emotional expression
23: Self-enhancement
28: Social support