Touché at CLEF 2023

Matching Images and Keywords with CLIP

Christiane Gütter, Fatihah Ulya Hakiem, Maaz Hussain, Nazifa Kazimi, Elisabeth Alice Rahn

Bauhaus-Universität Weimar

Example: Pro vs Con Keywords

Topic: Do we need cash?

PRO

cash is inclusive

CON

digital payment is better than cash

Task: Image Retrieval for Arguments

- We need to retrieve images that support a pro or con stance towards some controversial topics
- For each 50 topics, the title, description, and narrative is provided.

Number	100
Title	Do we need cash?
Description	A user pays with a card for the first time in their life. Impressed by the simplicity, they wonder whether they need cash at all.
Narrative	On-topic images address, for example, the advantages, problems or inconveniences of using cash, or with alternatives for small or large purchases. Pro images might be shared on social media with this text (no irony, no link): We need cash

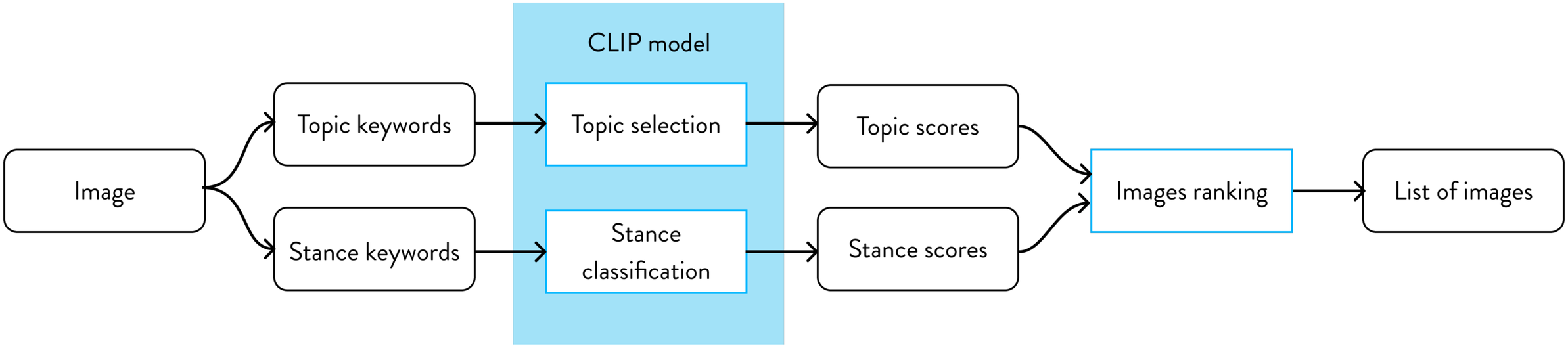
Example of a topic

Evaluation Criteria

- Topic Accuracy: Are images assigned to their correct topics?
- Argumentativeness: Are argumentative images assigned to a pro or con stance?
- Stance Relevance: Are images assigned to their correct pro/con stance?

Approach

- CLIP (Contrastive Language-Image Pretraining) model to connect images and their corresponding labels
- Given an image and corresponding text, the model calculates the similarity score



Approach

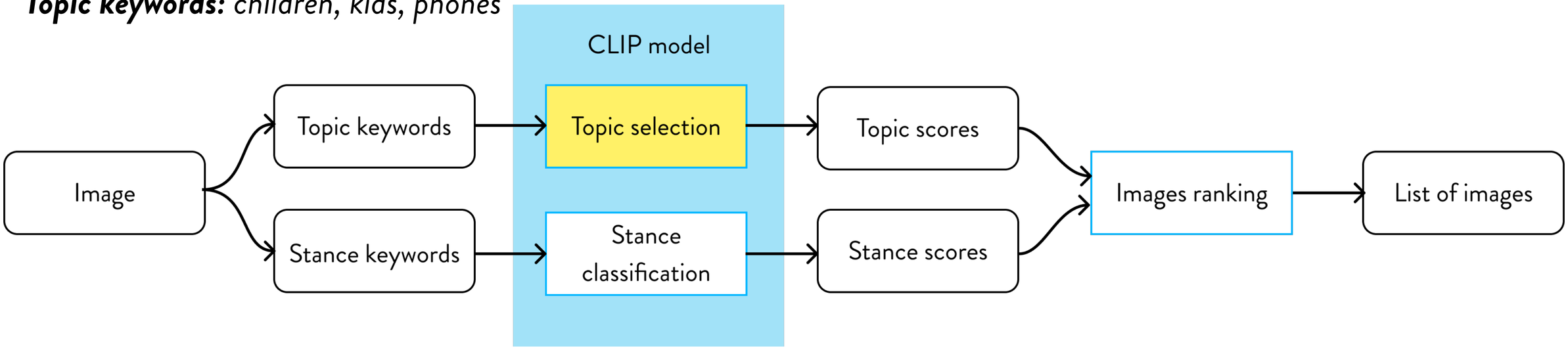
Determining Topic Accuracy

For each image, a list of of 50 topic titles or topic keywords are given to the model. The keywords were generated based on the titles and extended with synonyms and antonyms from WordNet. The ID of the list with the highest score is set as the topic for the image.

Example:

Topic title: *Should children have mobile phones?*

Topic keywords: *children, kids, phones*



Approach

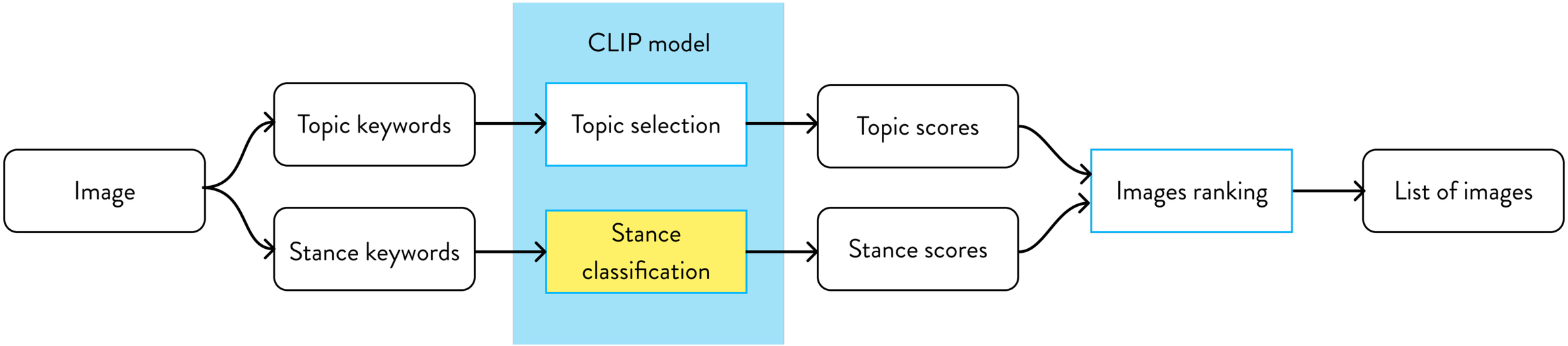
Determining Argumentativeness and Stance Relevance

The model was then given two lists consisting of keywords that support and contradict the topic. The keywords are generated manually. The list with the highest score determines the stance of the image.

Example:

Pro keywords: *mobile phones for children*

Con keywords: *emergency risk of internet access to children*



Result

- The approach based on the topic title and random stance assignment outperforms the one using topic keywords.
- This leads to a higher on-topic accuracy, but results in lower scores for argumentativeness and stance relevance.

Method	On-Topic	Arg.	Stance
Keywords	0.664	0.350	0.185
Topic-title	0.770	0.335	0.179

Future Plan

- Expand the keywords used for stance classification
- Adapt the model with more contextual captions