Title

One-Object Decision-Making model: Fast and Frugal Heuristic for HumanActivity Classification

Authors

Sharma, Karan
Bhandarkar, Suchendra

One-Object Decision-Making model: Fast and Frugal Heuristic for Human Activity Classification

karan sharma

Keysight Technologies, Atlanta, Georgia, United States

Suchendra Bhandarkar

The University of Georgia, Athens, Georgia, United States

Abstract

Consider an uncertain situation where an artificial intelligence (AI) system is called upon to determine a human action or activity in an image or scene. The AI system has not been previously trained to recognize any human action or activity, and has no prior information on pose, parts, spatial layout of the object in an image. In such a situation, what is the AI system supposed to do? Its options are limited, and it must determine the action or activity with the aid of the most probable inanimate object (other than the human actor) that it can detect in the image. The AI system needs to formulate two hypotheses to infer the action or activity in a zero-shot manner; first, that the most probable inanimate object detected in the image is one that is involved in the action or activity, and second, that the most likely action or activity associated with this object in the real world is the one actually occurring in the image. To what extent are these hypotheses valid? We propose that correct detection of the highly probable object and use of natural language word embeddings obtained via training on a general text corpus such as Wikipedia could enable the AI system to determine the underlying human action or activity in an image with reasonable classification accuracy. We conducted studies on the HICO dataset, which is a challenging dataset containing many rare human action/activity categories. Our experimental results show that if the AI system can reliably detect the most probable inanimate object in the image and then infer the corresponding verb in a zero-shot manner using language models trained on general text corpora, then it has a reasonable chance of correctly guessing the underlying action/activity in an image.