Abstracts

Ken Church

Feb 5, 2025

Abstracts are like talks, papers, etc.

- Say everything 3 times
  - Say what you are going to say (promise)
  - Say what you said (connect the dots)
  - Say what you said (delivery)
- Abstracts
  - Topic Sentence
  - Body
  - Concluding Sentence

BERT

- We introduce a new language representation model called BERT,
  - which stands for Bidirectional Encoder Representations from Transformers.
- Unlike recent language representation models, BERT is designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers. As a result, the pre-trained BERT model can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks, such as question answering and language inference, without substantial task-specific architecture modifications. BERT is conceptually simple and empirically powerful.
- It obtains new state-of-the-art results on eleven natural language processing tasks, including
  - pushing the GLUE score to 80.5% (7.7% point absolute improvement),
  - MultiNLI accuracy to 86.7% (4.6% absolute improvement),
  - SQuAD v1.1 question answering Test F1 to 93.2 (1.5 point absolute improvement) and
  - SQuAD v2.0 Test F1 to 83.1 (5.1 point absolute improvement).

Deep Seek https://arxiv.org/pdf/2501.12948

- We introduce our first-generation reasoning models,
  - DeepSeek-R1-Zero and DeepSeek-R1.
- DeepSeek-R1-Zero, a model trained via large-scale reinforcement learning (RL) without supervised fine-tuning (SFT) as a preliminary step, demonstrates remarkable reasoning capabilities. Through RL, DeepSeek-R1-Zero naturally emerges with numerous powerful and intriguing reasoning behaviors. However, it encounters challenges such as poor readability, and language mixing. To address these issues and further enhance reasoning performance, we introduce DeepSeek-R1, which incorporates multi-stage training and cold-start data before RL. DeepSeekR1 achieves performance comparable to OpenAI-o1-1217 on reasoning tasks.
- To support the research community, we open-source
  - DeepSeek-R1-Zero,
  - DeepSeek-R1, and
  - six dense models (1.5B, 7B, 8B, 14B, 32B, 70B) distilled from DeepSeek-R1
    - based on Qwen and Llama.

Suffix Arrays

- A new and conceptually simple data structure, called a suffix array,
  - for on-line string searches is introduced in this paper.
- Constructing and querying suffix arrays is reduced to a sort and search paradigm that employs novel algorithms. The main advantage of suffixarrays over suffixtrees is that, in practice, they use three to five times less space. From a complexity standpoint, suffix arrays permit on-line string searches of the type, “Is W a substring of A?” to be answered in time $O(P + \log N)$, where P is the length of W and N is the length of A, which is competitive with (and in some cases slightly better than) suffix trees. The only drawback is that in those instances where the underlying alphabet is finite and small, suffix trees can be constructed in $O(N)$ time in the worst case, versus $O(N \log N)$ time for suffix arrays. However, we give an augmented algorithm that, regardless of the alphabet size, constructs suffixarrays in $O(N)$ expected time, albeit with lesser space efficiency.
- We believe that suffix arrays will prove to be better in practice
  - than suffix trees for many applications.

ArtELingo-28

- Research in vision and language has made considerable progress thanks to benchmarks such as COCO.
- COCO captions focused on unambiguous facts in English; ArtEmis introduced subjective emotions and ArtELingo introduced some multilinguality (Chinese and Arabic). However we believe there should be more multilinguality. Hence, we present ArtELingo28, a vision-language benchmark that spans 28 languages and encompasses approximately 200,000 annotations (140 annotations per image). Traditionally, vision research focused on unambiguous class labels, whereas ArtELingo28 emphasizes diversity of opinions over languages and cultures. The challenge is to build machine learning systems that assign emotional captions to images.
- Baseline results will be presented for three novel conditions: Zero-Shot, Few-Shot and One-vs-All Zero-Shot.
- We find that cross-lingual transfer is more successful for culturally-related languages.
- Data and code will be made publicly available.

No Culture Left Behind: ArtELingo-28, a Benchmark of WikiArt with Captions in 28 Languages

Youssef Mohamed[1]* Runjia Li[2] Ibrahim Said Ahmad[3] Kilichbek Haydarov[1]
Philip Torr[2] Kenneth Ward Church[3] Mohamed Elhoseiny[1]*
[1]KAUST [2]University of Oxford [3]Northeastern University

https://www.artelingo.org/

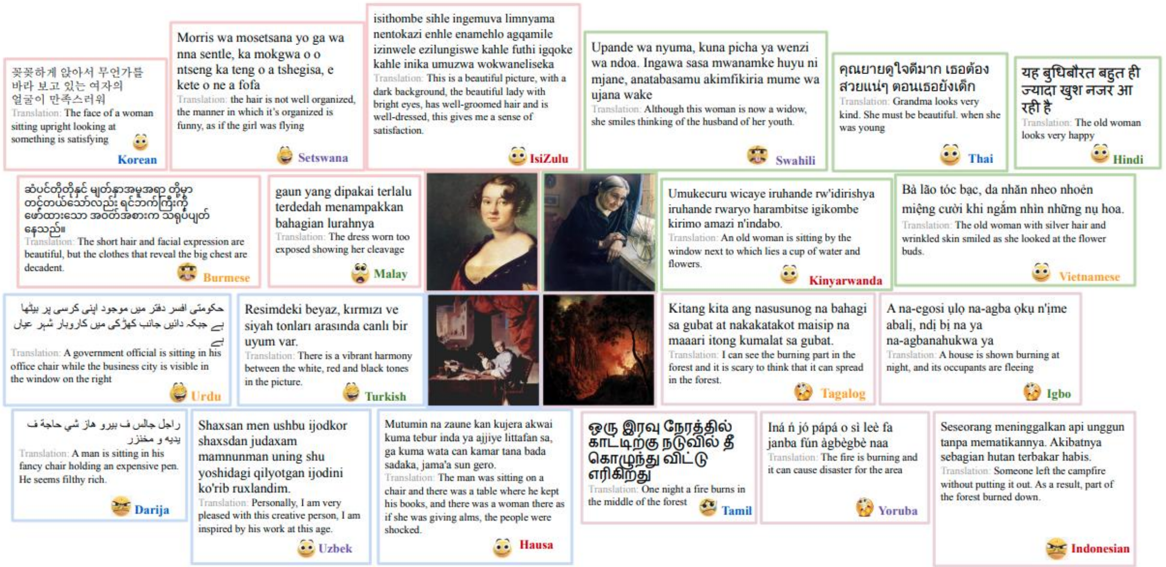

Figure 1: **ArtElingo-28 Benchmark:** 9 emotion labels with captions in 28 languages. The ∼140 annotations per WikiArt image embrace diversity over languages and cultures.

Abstract

Research in vision and language has made considerable progress thanks to benchmarks such as COCO. COCO captions focused on unambiguous facts in English; ArtEmis introduced subjective emotions and ArtELingo introduced some multilinguality (Chinese and Ara-

1 Introduction

A quick review of recent surveys on multimodal AI (Cao et al., 2023; Berrios et al., 2023; Zhang et al., 2023), reveals just how much the literature is focused on English. The literature on benchmarking (Liu et al., 2023c; Li et al., 2023a) provides an