JHU vision lab

Information Maximization Perspective of Orthogonal Matching Pursuit with Applications to Explainable AI

Aditya Chattopadhyay[1*], Ryan Pilgrim[1*], René Vidal[2]

[1]Johns Hopkins University, [2]University of Pennsylvania

* Equal Contribution

Overview

- Information Pursuit (IP)[1] is a classical greedy algorithm for active testing.
- IP predicts a variable by sequential asking queries about an input in order of information gain.
- Difficult to implement in high dimensions.

- Orthogonal Matching Pursuit (OMP)[2] is a classical greedy algorithm for sparse coding.
- OMP encodes a signal by sequentially selecting dictionary atoms in order of correlation gain.
- Easy to implement in high dimensions.

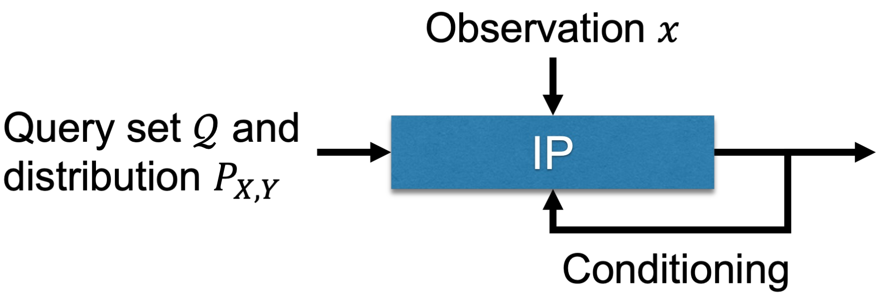
1. D. Geman and B. Jedynak. An active testing model for tracking roads in satellite images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 18(1):1–14, January 1996.
2. Y.C. Pati, R. Rezaiifar, and P.S. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In Asilomar Conference on Signals, Systems and Computers, pages 40–44, 1993.

Contributions

- We formally prove a connection between IP and OMP.
  - OMP can be seen as a special case of IP (modulo a normalization factor).

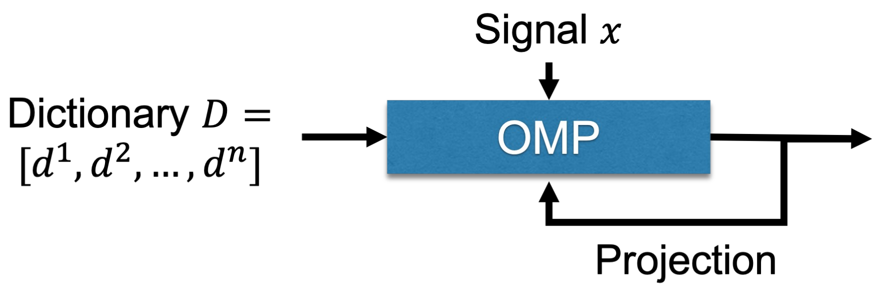
- We propose a computationally simpler alternative to IP for explainable AI that is based on OMP.

Primer: IP and OMP



$$Q_{k+1} = \arg\max_{Q \in \mathcal{Q}} I(Q; Y \mid \mathcal{H}_{1:k}(x))$$

History



$$d_{k+1} = \arg\max_{d^j \in D} |\langle d^j, \Pi_{D_k}^{\perp} x \rangle|$$

Residual

IP vs. OMP

- In IP, the queries and the prediction are all random variables.
- In OMP, the signal and dictionary atoms are all vectors (not random).
- **Contribution 1:** We show that despite these differences, one can obtain the OMP algorithm (up to a normalization factor) from IP by carefully selecting the set of queries and prediction variables.

OMP from IP

- Take each query $Q^i \in \mathcal{Q}$ as a random projection of dictionary atom d^i onto a standard normal Z, that is, $Q^i := \langle d^i, Z \rangle$.
- Take the prediction variable Y to be a random projection of the observed signal x onto Z, that is $Y := \langle x, Z \rangle$.
- ***Theorem (Informal):*** *The query selection step for IP with this choice of $\mathcal{Q}$ and Y coincides with the atom selection step in OMP up to a normalization factor.*

IP-OMP

- More precisely, IP proceeds as follows,

$$Q_1 = \arg\max_{Q^j \in \mathcal{Q}} I(Q^j; \langle x, Z \rangle) = \arg\max_{d^j \in D} \frac{|\langle d^j, x \rangle|}{\|d^j\|_2 \|x\|_2}$$

$$Q_{k+1} = \arg\max_{Q^j \in \mathcal{Q}} I(Q^j; \langle x, Z \rangle \mid \mathcal{H}_{1:k}) = \arg\max_{d^j \in D, \|\Pi_{D_k}^{\perp} d^j\|_2 \neq 0} \frac{|\langle \Pi_{D_k}^{\perp} d^j, \Pi_{D_k}^{\perp} x \rangle|}{\boxed{\|\Pi_{D_k}^{\perp} d^j\|_2 \|\Pi_{D_k}^{\perp} x\|_2}}.$$

- Main distinction with OMP is this normalization factor—we call this IP-derived algorithm *IP-OMP.*

- We empirically show that IP-OMP and OMP have similar success rates for sparse code recovery using random Gaussian dictionaries.

IP-OMP for explainable AI (CLIP-IP-OMP)

- **Contribution 2:** Inspired by recent application for IP to explainable AI, we propose a simple algorithm using IP-OMP for the same.

- **Modelling assumption:** CLIP image embeddings can be expressed as sparse combinations of CLIP embeddings of text concepts.

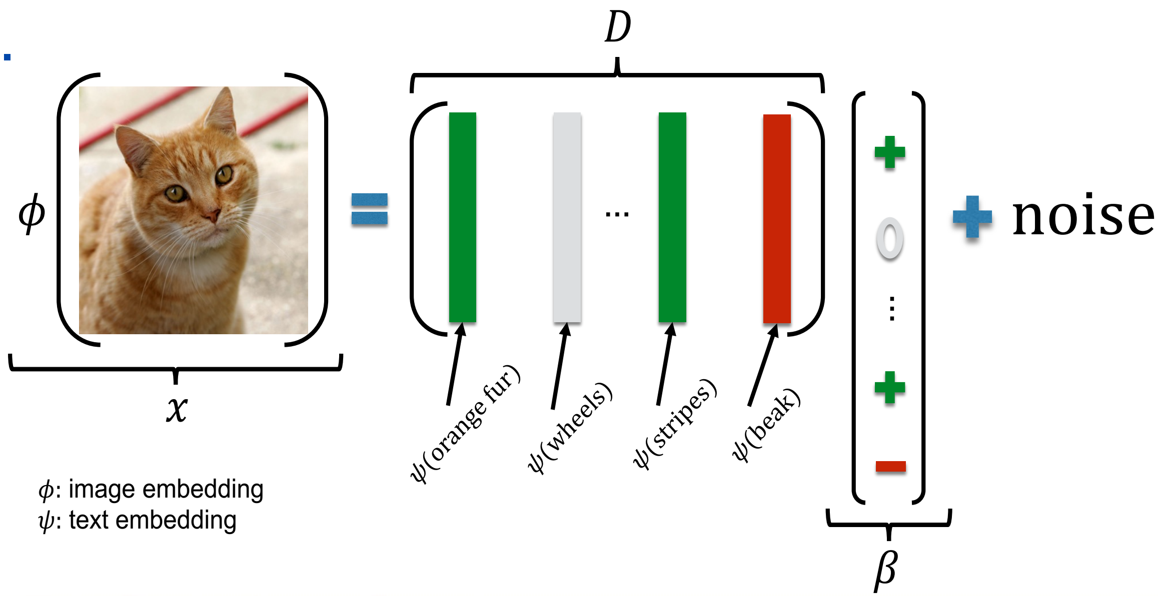


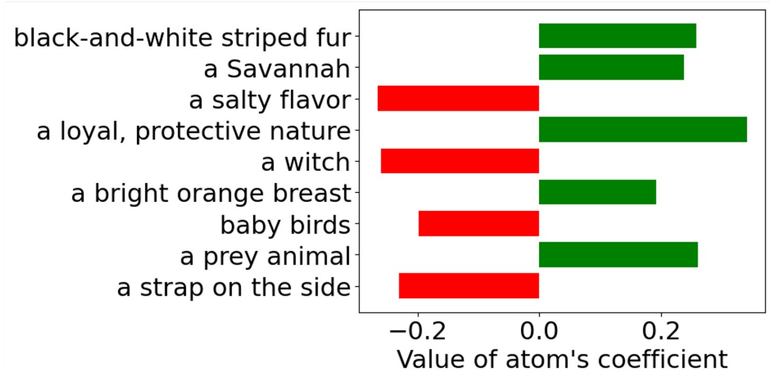
Image credit: *https://en.wiktionary.org/wiki/cat#/media/File:Cat03.jpg*

CLIP-IP-OMP explanations

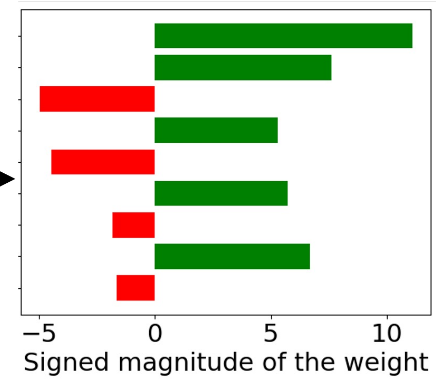
Input image x

Prediction: Tiger

Sparse code $\hat{\beta}_{\text{IP-OMP}}$ after 9 iterations



Weights of linear classifier



- **Idea:**
  - Construct a dictionary of CLIP text embeddings of semantic concepts.
  - Use IP-OMP to sparse-code image embeddings.
  - Train a linear classifier to predict class from sparse code.
  - Explain predictions via the sparse code and classifier weights.

More information

Research supported by the Army Research Office under the Multidisciplinary University Research Initiative contract W911NF-17-1-0304, the NSF grant 2031985 and by Simons Foundation Mathematical and Scientific Foundations of Deep Learning (MoDL) grant 135615.

Vision Lab @ JHU
http://www.vision.jhu.edu

Mathematical Institute for Data Science @ JHU
http://www.minds.jhu.edu

Thank You!

JOHNS HOPKINS
MATHEMATICAL INSTITUTE
for DATA SCIENCE