DESSERT: An Efficient Algorithm for Vector Set Search with Vector Set Queries

By Joshua Engels, Benjamin Coleman, Vihan Lakshman, Anshumali Shrivastava



Traditional Vector Search



Vector Set Search

Less obvious which set is the most similar to the query



Vector Set Search Problem

- Given set-to-set similarity function F(X, Y)
- E.g.

$$F(X,Y) = \sum_{x \in X} \left[\max_{y \in Y} \left(\sin(x,y) \right) \right]$$

- Given *N* vector sets S_i and query set Q
- Find:

$$S^{\star} = \operatorname*{argmax}_{i \in \{1, \dots, N\}} F(Q, S_i)$$



Why do we care?

- Many objects can be better represented as collections of vectors!
- Document search: embed each word in each document
 - ColBERT (PLAID) gets SOTA
- Problem: No existing efficient approximate algorithm like for single vector search



The lk system sources documents and multimedia information from a variety of web resources. This data is compiled by web crawlers and is sent to database storage systems. The free-text terms are indexed, and the vocabulary is sorted, both using automated or manual procedures. Sept. 2003

upGrad
 https://www.upgrad.com + Blog + Data Science
 information Retrieval System Explained: Types, Comparison ...

People also ask

Google

What is the basic process of information retrieval?

What is an example of an information retrieval system?

How does document retrieval work?

How does AI works in information retrieval?

GeeksforGeeks
https://www.geeksforgeeks.org>what-is-information-...

What is Information Retrieval?

Sep 19, 2023 — Information Retrieval (IR) can be defined as a software program that deals with the organization, storage, retrieval, and evaluation of ... Components Of Information... \circ Difference Between... \circ User Interaction With...

Wikipedia https://en.wikipedia.org > wiki > Information_retrieval

Information retrieval

Information retrieval (IR) in computing and information science is the process of obtaining information system resources that are relevant to an information ... Overview - History - Applications - Timeline

Q Quora https://www.quora.com - What are the components-of-t...

What are the components of the information retrieval system? An information retrieval system(RRS), is used to perform search queries and reports(e.g. SQL reports). It's like having someone open the file cabinet and ... I answer: 0 votes: The main components of an information retrieval system are: 1. Document ...

Coveo https://www.coveo.com > blog > information-retrieval

What Is Information Retrieval?

Nov 8, 2022 — Information retrieval systems act as a bridge between people and data repositories. At a high-level, an information retrieval system includes ... The 3 Parts Of Any... · Machine Learning And Ir... · Probabilistic Models

Solution: Dessert



Theoretical Query Latency

Bruteforce:

$$O(m_q m N d)$$

• Ours:

$$O\left(m_q \log(Nm_q m/\delta)d + m_q N \log(Nm_q m/\delta)\right)$$

- 1 $\boldsymbol{\delta}$ is probability \cup **Ordentified** hashing cost,
- Elided dataset dependently constituints
- 1δ ved dependence on m and d

Cost to query all N sketches with each of the m_a query vectors

Empirical Results: Synthetic Data



Empirical Results: MS MARCO (passage retrieval)

	Method	Latency (ms)	MRR@10
Results optimized for $k = 10$:	DESSERT	9.5	35.7 ± 1.14
	DESSERT	15.5	37.2 ± 1.14
	PLAID	45.1	39.2 ± 1.15

	Method	Latency (ms)	R@1000
Results optimized for $k = 1000$:	DESSERT	22.7	95.1 ± 0.49
	DESSERT	32.3	96.0 ± 0.45
	PLAID	100	97.5 ± 0.36

Empirical Results: LoTTE (passage retrieval)



Future Work

- Can we build a vector-set search algorithm that is sublinear in N?
- What other data domains beyond passage retrieval can we try?
 - Image search: use pre-pooled embeddings
 - E-commerce: embed each item in a basket
 - Clustering: find most similar cluster to existing cluster
- What additional problems can be sped up with approximate set similarity calculations?

Team and Contact

Josh Engels: josh.adam.engels@gmail.com Previously at ThirdAI, now a PhD student at MIT

Ben Coleman: <u>benjamin.ray.coleman@gmail.com</u> *Previously at ThirdAI, now a research scientist at Google DeepMind*

Vihan Lakshman: <u>vihan@thirdai.com</u> AI Engineer at ThirdAI

Anshumali Shrivastava: <u>anshumali@rice.edu</u> CEO of ThirdAI, associate professor at Rice University







