

# Big Data to Decisions

Guillermo Sapiro

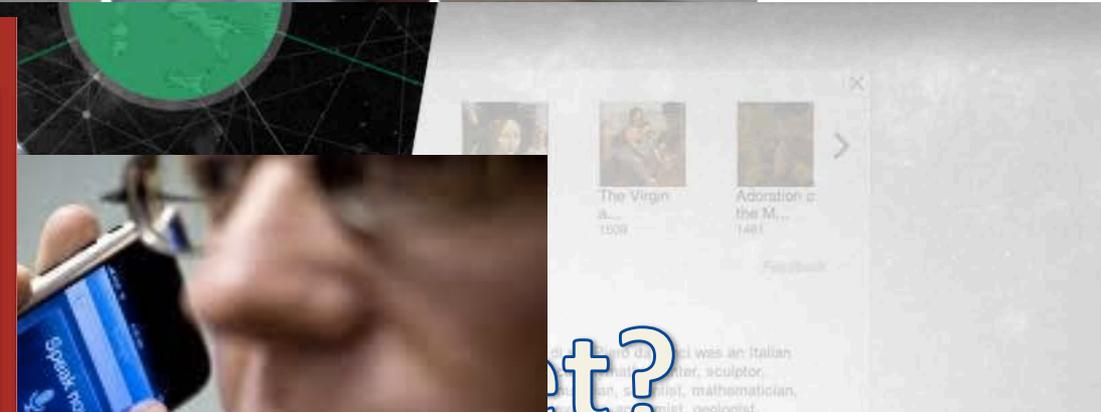
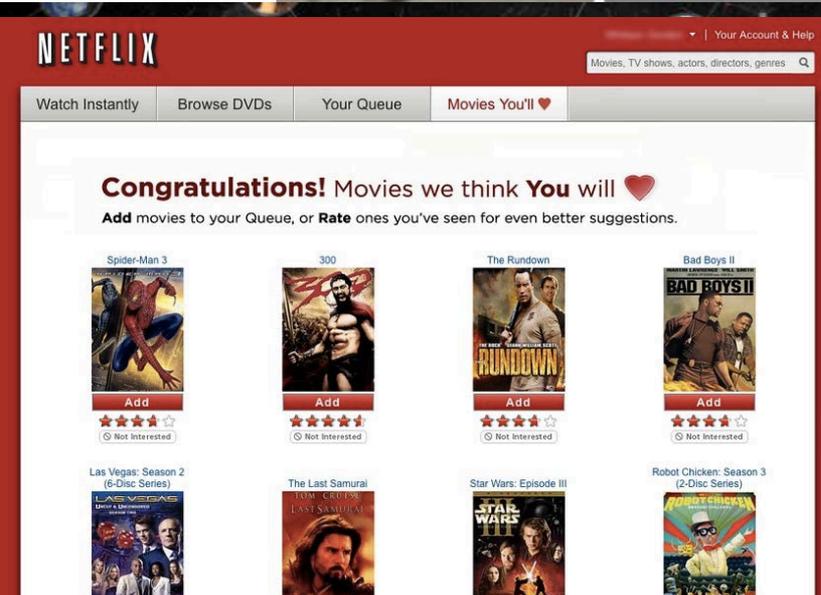


# Some Basic Concepts

- **Data:** facts and statistics collected together for reference or analysis
- **Big Data:** data sets that are too large and complex to manipulate with standard methods or tools
- **Decision:** A conclusion or resolution reached after consideration
- **Information:** facts provided or learned about something or someone



# Examples of Big Data and Decisions: Industry/Academia Today

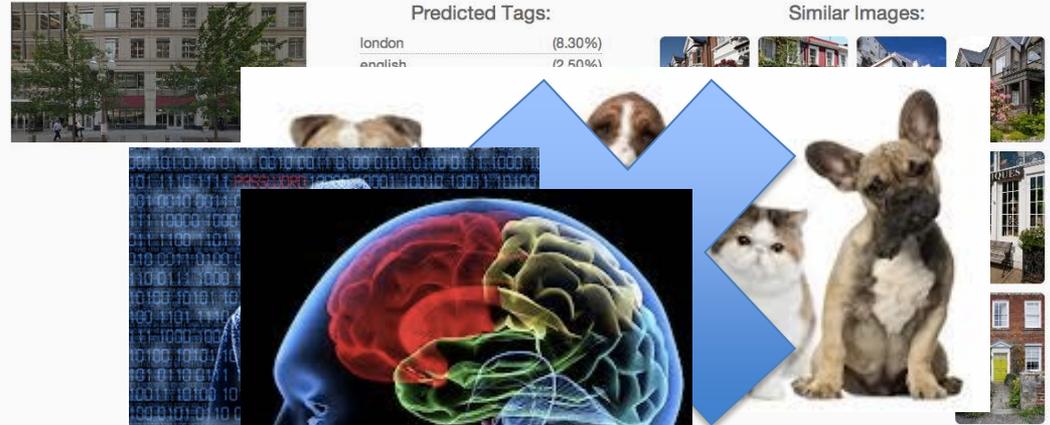


A screenshot of the Amazon.com website's "Recommended for You" section. The Amazon logo is at the top left, and the text "Recommended for You" is at the top right. Below this, a line of text reads "Amazon.com has new recommendations for you based on [items](#) you purchased or told us you own." Below this text, there are three book covers displayed in a row. Each cover has a "LOOK INSIDE!" banner at the top. The first book is "Google Apps Deciphered: Compute in the Cloud to Streamline Your Desktop". The second book is "Google Apps Administrator Guide: A Private-Label Web Workspace". The third book is "Googlepedia: The Ultimate Google Resource (3rd Edition)". Below each book cover is its title and subtitle in blue text.

# Challenges:

## Why is this Not Enough for the Sailor/Marine?

- N=1
- Life-changing decisions
- Data is not information
  - Good data representations
  - Adversary data
- Data integration
  - Multimodal
  - Complex environment
  - Dynamic environment
- Collect less data – Expensive
  - Analytics helps to collect less data
- Time and power-constrained acquisition and
- Same data/hardware and new task
- Whole scene analysis
- Natural query system



- **Navy leading its own data analytics challenges**

# Navy/ONR Examples Moving in This Direction

- Multimodal

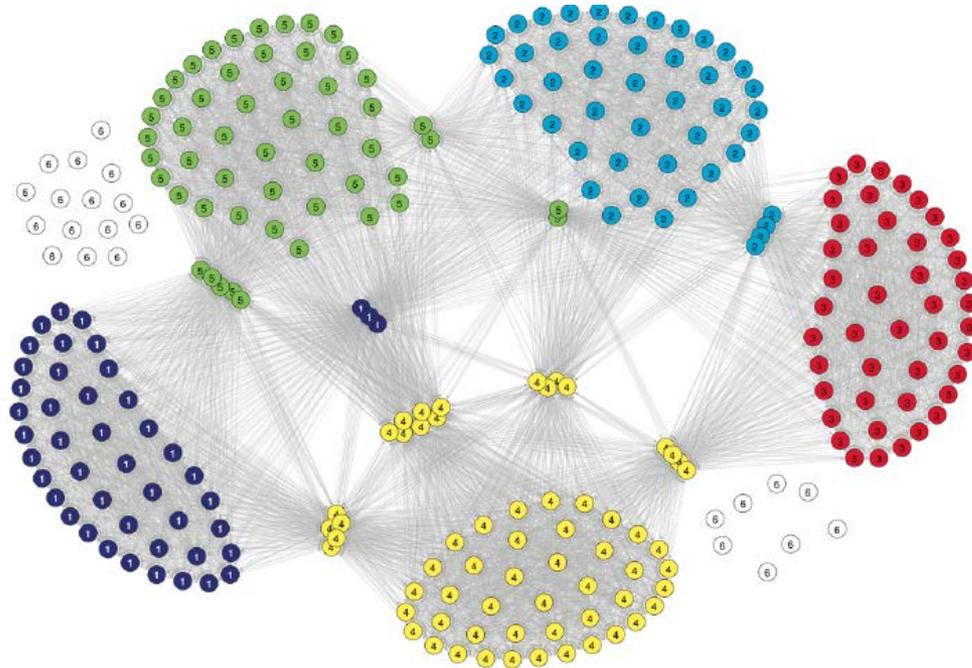


Image from Carin et al.

# Navy/ONR Examples Moving in This Direction

- Data representation and information



Image from Baraniuk et al.



–Given pairwise dissimilarities between data points, we consider the problem of finding a subset of data points called representatives or exemplars that can efficiently describe the data collection.

–We obtain the range of the regularization parameter for which the solution of the proposed optimization program changes from selecting one representative to selecting all data points as the representatives.

–When there is a clustering of data points, defined based on their dissimilarities, we show that for a suitable range of the regularization parameter, the algorithms finds representatives from each cluster.

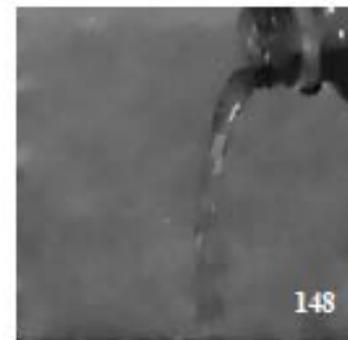
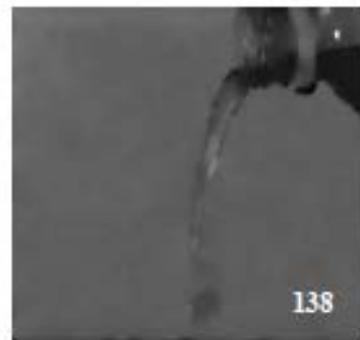
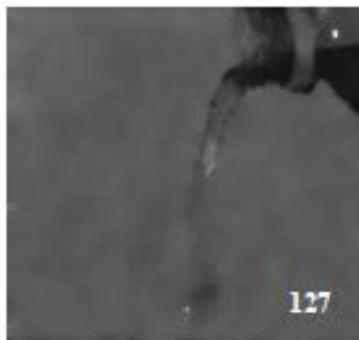
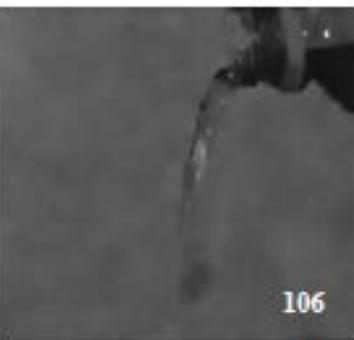
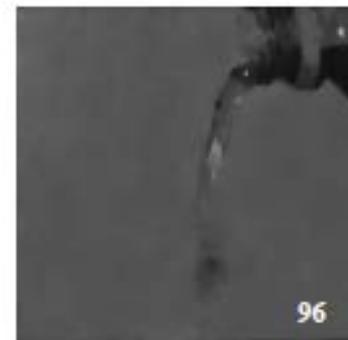
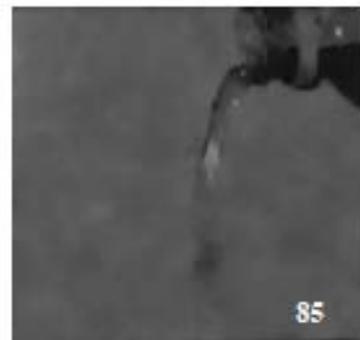
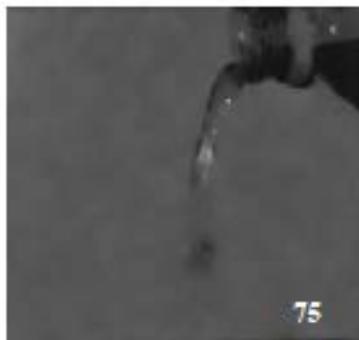
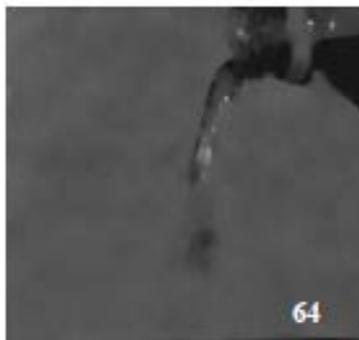
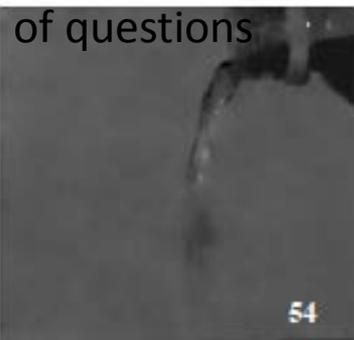
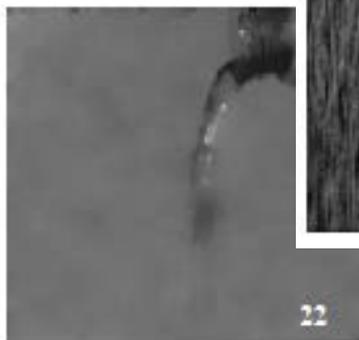
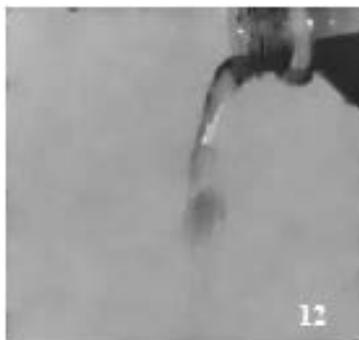
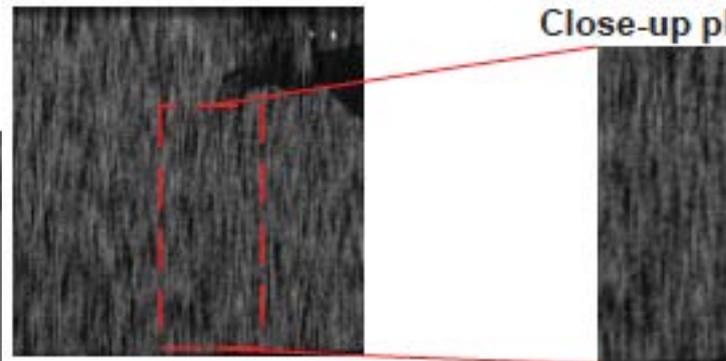
–As the results show the classification performance using the representatives found by our proposed algorithm is close to that of using all training samples.

# Navy/ONR Examples Moving in This Direction

148 frames recovered

Single frame captured

Close-up photo



Reduced number  
of questions

# Navy/ONR Examples Moving in This Direction Scientific Wish List?

- Foundations
  - Coifman, Lafferty, Calderbank, DeVore, Maggioni, Osher, etc.

# Scientific Wish List?

- Complex and dynamic scenarios
- High risk decisions/tasks
- Data representation
- Integration
- Foundations
- Why it works? When it works?
- Real time
- N=1

# Concluding Remarks

- Significant progress in big data
- Lot's still to be done
- Big data + task = success
- Scientific community + Navy = success

Thanks

- Thanks to Navy, ONR, Wen Masters, Carey Schwartz, Wayne Perras