Categorizing and Inferring the Relationship Between the Text and Image of Twitter Posts

Alakananda Vempala

Daniel Preoţiuc-Pietro

TechAtBloomberg.com

Motivation

What's the largest difference in
Twitter content in 2010 and 2019?

2010

2019

Motivation

What's the largest difference in Twitter content in 2010 and 2019?

Many more tweets contain images

2010

2019

Motivation

What's the largest difference in Twitter content in 2010 and 2019?

Many more tweets contain images

Approximately 12% tweets are now accompanied by images
> 50M/day

Very little is known about the text-image relationship

2010

2019

Motivation

Text and image in a tweet can be related in several ways:

1. Text is the caption of the image

Bloomberg

Engineering

Motivation

Text and image in a tweet can be related in several ways:

2. Text is a comment on the image

Bloomberg

Engineering

Motivation

Text and image in a tweet can be related in several ways:

3. Image only illustrates part of the tweet text

Bloomberg
Engineering

Applications

Automatically hiding images that don't add additional content to a Twitter post to maximize screen estate

Image can be hidden
Two extra tweets can be displayed

Image has additional content

Bloomberg

Engineering

Applications

- Automatically identify tweets that contain images and their captions

- Data can be used for distant supervision for image classification

- Identify tweets that contain creative descriptions for images

Bloomberg

Engineering

Data – Task Definition – Text Task

Aim: Determine whether there is semantic overlap between the context of the text and the image

Some or all of the content words in the
text are represented in the image
(Text is represented)

None of the content words in the text are
represented in the image
(Text is not represented)

Data – Task Definition – Image Task

Focuses on the role of the image to the semantics of the tweet
Aim: Identify if the image's content contributes with additional information to the meaning of the tweet beyond the text

Image has additional content that represents the meaning of text & image (Image adds)

Image does not add additional content that represents the meaning of text & image (Image does not add)
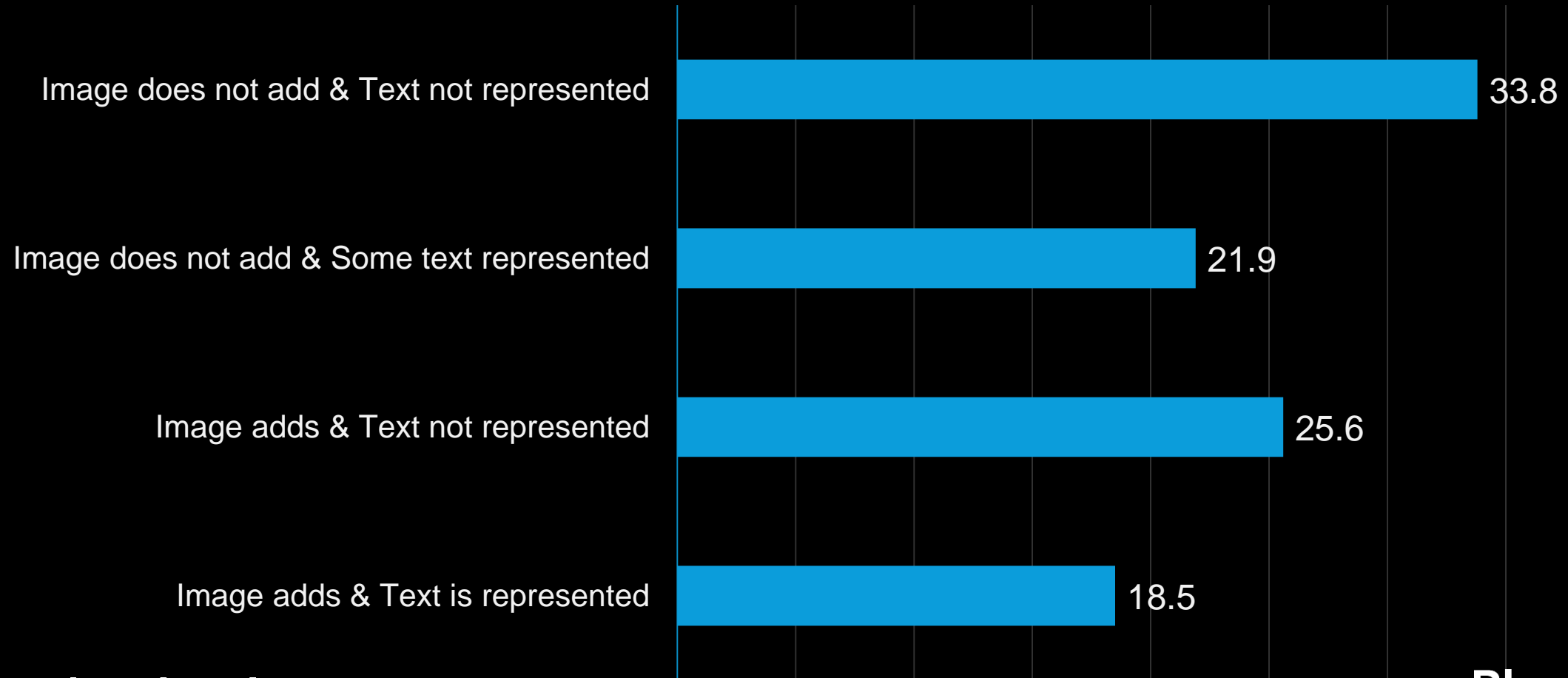
Data – Annotation

- We used Figure Eight (formerly CrowdFlower)

- Text task
  — Redundancy of 3
  — Krippendorf's Alpha = 0.71 (Text Task)
  — Annotators maintained > 85% accuracy over test questions

- Image task
  — Redundancy of 5
  — Krippendorf's Alpha = 0.46 (Text Task)
  — Annotators maintained > 75% accuracy over test questions

- Adjudication by majority vote

Bloomberg

Engineering

Data – Collection

- Collected annotations for 4,888 tweets
  — All tweets posted in the same year (2016)
  — Split across original posts, retweets and favorited posts
  — Deliberately sampled from users with known demographic traits
 — Tweets are all in English

- Available online: https://github.com/danielpreotiuc/text-image-relationship

Bloomberg

Engineering

Data – Distribution

Both task labels are combined to assign one of four classes to each text-image pair

| Class	Value
Image does not add & Text not represented	33.8
Image does not add & Some text represented	21.9
Image adds & Text not represented	25.6
Image adds & Text is represented	18.5

Bloomberg

Engineering

Analysis – Text Task

- Univariate Point-Biserial Correlation between unigram features and text task outcome
- Age is correlated with text being represented in image
    - Especially when image also adds information
    - More *traditional* type of relationship
- Simple tweet metadata not correlated
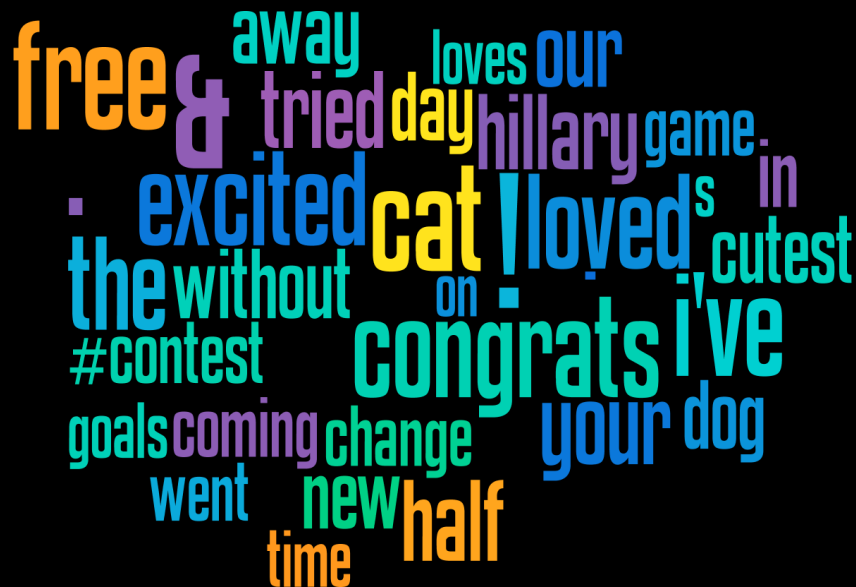
Some or all of the content words in the text are represented in the image
(Text is represented)
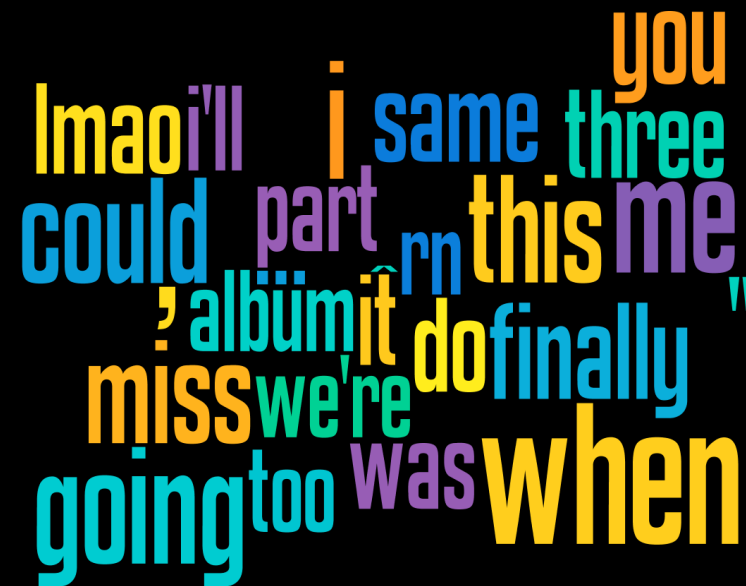
None of the content words in the text are represented in the image
(Text is not represented)

Analysis – Image Task

- Univariate Point-Biserial Correlation between unigram features and text task outcome
- Simple tweet metadata not correlated
- No demographic user information correlated
- 4-way analyses in the paper

Image has additional content that represents the meaning of text & image (Image adds to meaning)
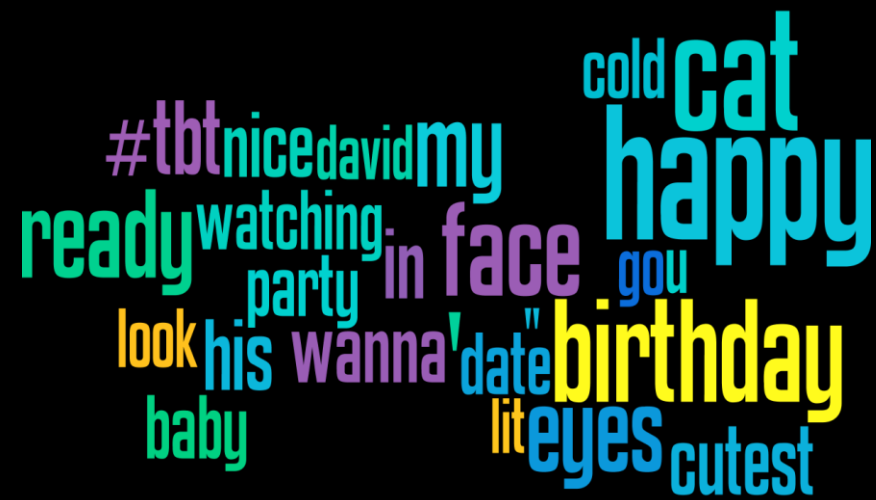
Image does not add additional content that represents the meaning of text & image (Image does not add to meaning)

Prediction – Methods

- Demographics

- Metadata

- Text-Based Methods
  — Surface
  — Unigrams
  — BiLSTM

- Image-Based Methods
  — ImageNet Classes
  — InceptionNet Tuned

- Joint Text+Image approaches
  — Linear ensemble of text and image predictions
  — LSTM + InceptionNet architecture
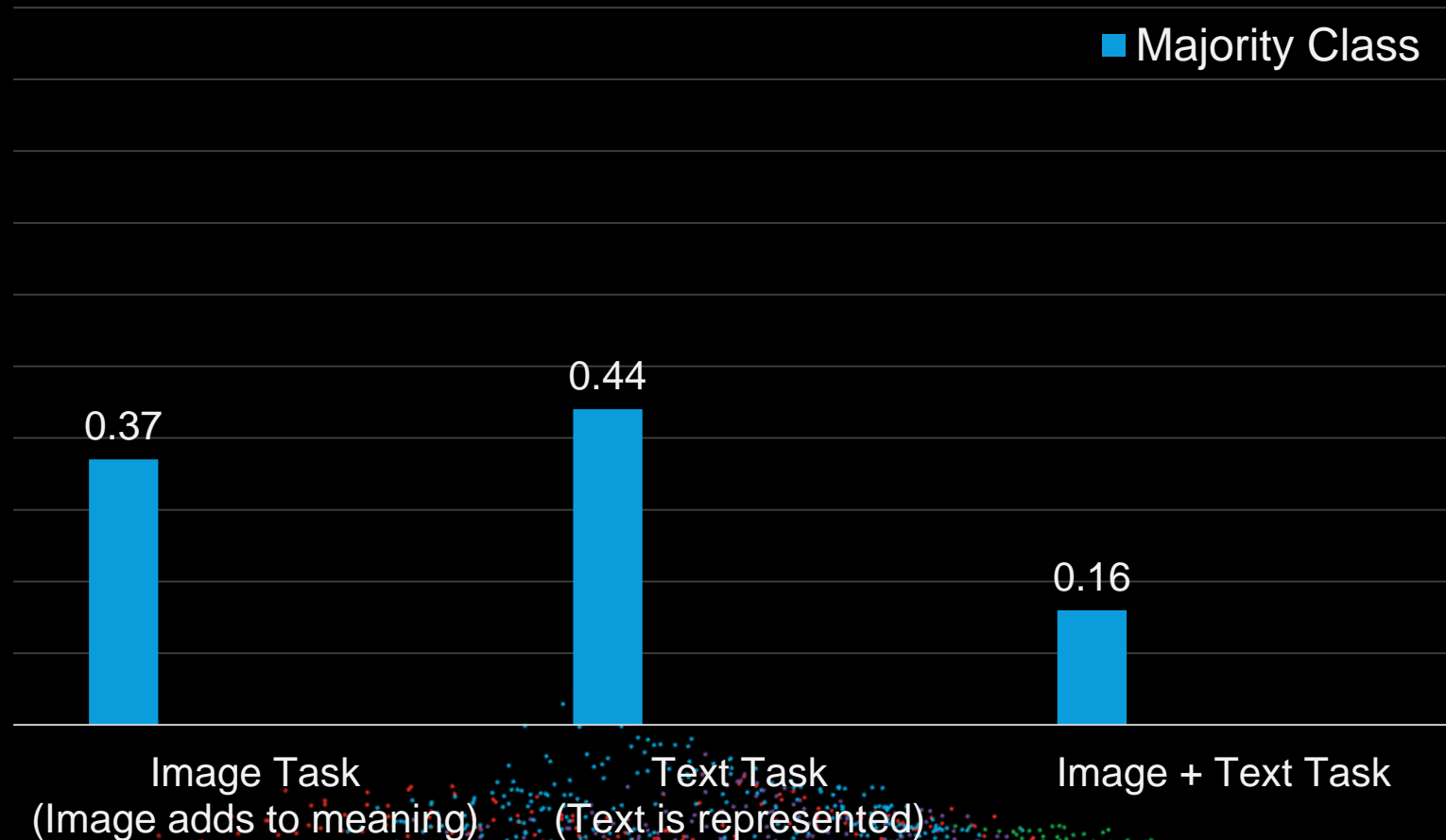
TechAtBloomberg.com

Bloomberg

Engineering

Prediction – Baseline Methods

- Results in weighted F1, Train (80%), Test (20%), Parameters set via 10-fold CV
- Three tasks: Image Task (binary), Text Task (binary), Image + Text (4-class)
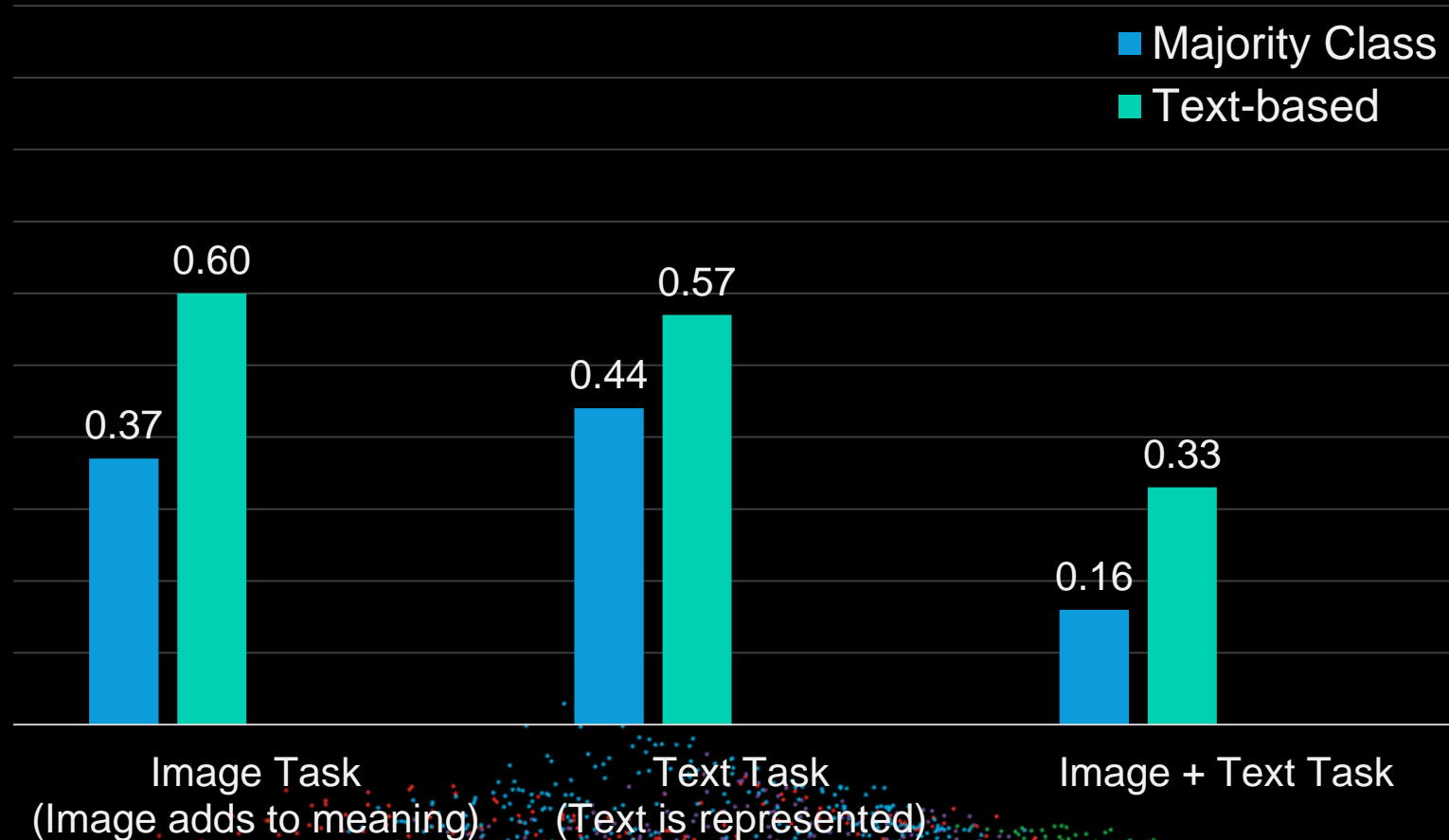- Demographics, tweet metadata features almost no predictive value

Majority Class

0.37	Image Task (Image adds to meaning)
0.44	Text Task (Text is represented)
0.16	Image + Text Task

Bloomberg
Engineering

Prediction – Text-based Methods

- All tasks show improvements on the baseline
  — Higher predictive power on image task (does the image have additional content)
- LSTM models are marginally better than unigrams

Legend:
- Majority Class
- Text-based

Image Task (Image adds to meaning): Majority Class 0.37, Text-based 0.60
Text Task (Text is represented): Majority Class 0.44, Text-based 0.57
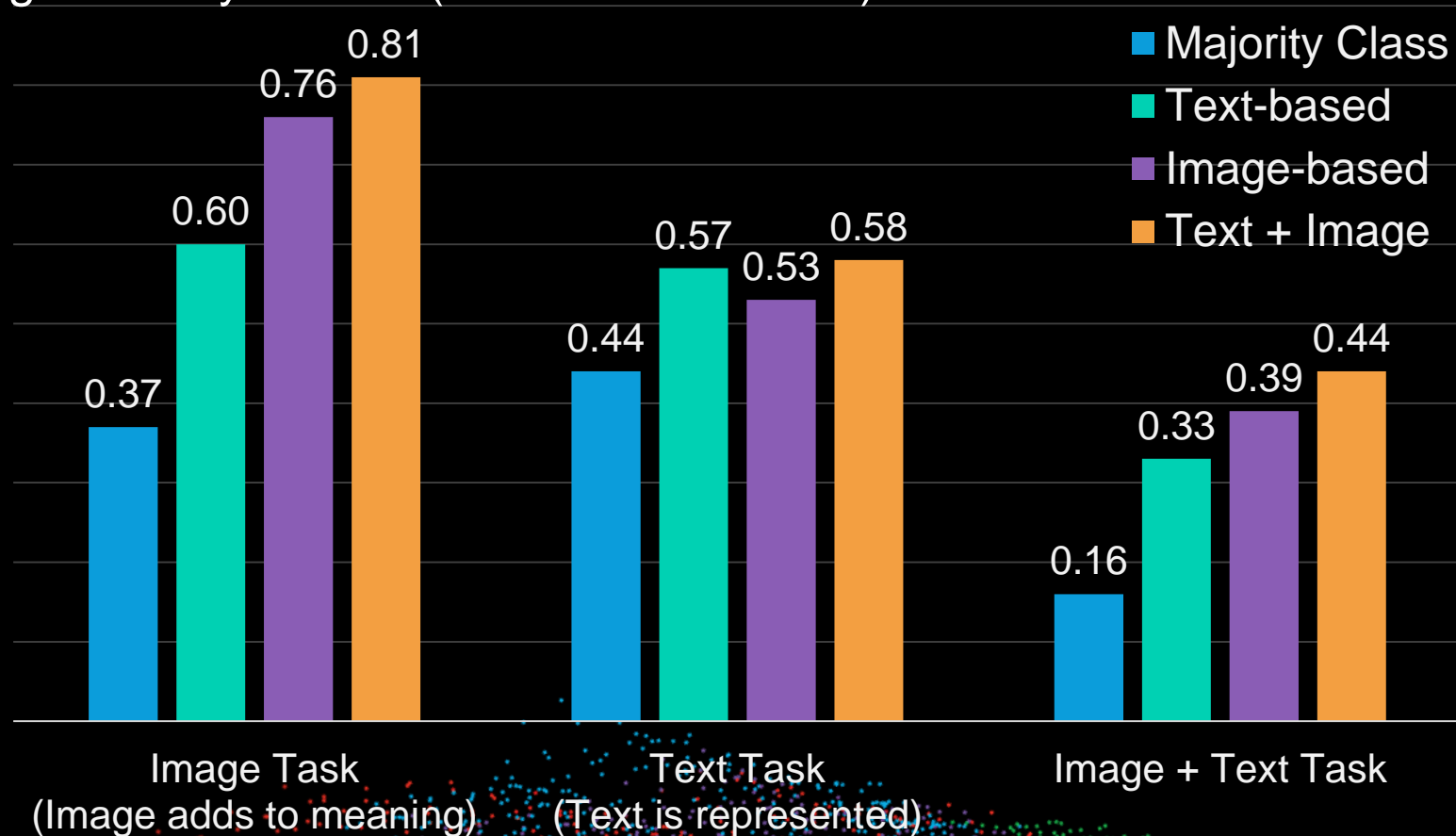Image + Text Task: Majority Class 0.16, Text-based 0.33

Bloomberg
Engineering

Prediction – Joint Text + Image Methods

- Improves over the best text-based or image-based methods
- LSTM + InceptionNet performs better
- Image task is much more predictable using models
- Text task is significantly harder (similar to humans)

Legend:
- ■ Majority Class
- ■ Text-based
- ■ Image-based
- ■ Text + Image

Image Task (Image adds to meaning): 0.37, 0.60, 0.76, 0.81

Text Task (Text is represented): 0.44, 0.57, 0.53, 0.58

Image + Text Task: 0.16, 0.33, 0.39, 0.44

Bloomberg
Engineering

Takeaways

Text-Image relationship in tweets is complex:
- Text does not always describe the image
- The image does not always illustrate text

Text-image relationship is likely useful for downstream applications

New classification schema and data set for text-image relationships on Twitter
- https://github.com/danielpreotiuc/text-image-relationship

Relationship type is predictable from both text and image
- Best results on each subtask are obtained by methods using different modalities (text or image)

We are hiring:
- NYC – http://careers.bloomberg.com/job/detail/74022
- London – http://careers.bloomberg.com/job/detail/74154

Bloomberg
Engineering