Identifying weird and bizarre news in online news media

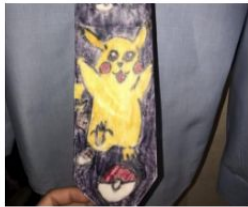Vijayasaradhi

Definition and Motivation

- Bizarre news is news so strange and unusual that readers might question the claims presented in the news.
- Usually very rare, silly, strange and unbelievable.
- Time agnostic, fun to read even long after the event.
- **Use bizarre news as anchors to gain readers attention and increase viewership.**

Some examples

Bear breaks into Colorado home, plays piano
June 5 (UPI) -- Police in a Colorado city said a resident checked her security camera after an apparent burglary and discovered the culprit was actually a piano-playing bear.

Dad attends son's graduation wearing tie he made in first grade
June 5 (UPI) -- A Texas dad warmed the hearts of Twitter by attending his son's graduation proudly wearing the "Pokemon" tie the boy made for him in first grade.

Squirrels steal Texas veteran's flags to build patriotic nest
June 4 (UPI) -- A Texas veteran's heartbreak at having several U.S. flags stolen from his front yard turned to laughter when he discovered the culprits -- squirrels.

Police: Papa John's workers delivered cocaine in pizza boxes
June 4 (UPI) -- Authorities in Washington state said employees at a Papa John's pizzeria were selling cocaine at the restaurant and delivering the drugs in pizza boxes.

More examples…

- A goat has been arrested in India and it's facing seven years in prison
- Florida man dies after eating roaches and worms in contest
- Alligator bites off the arm of a Florida man fleeing police
- Student left 30 minutes from death after tapeworm burrowed into his brain
- Woman shot in the foot by pet dog named Trigger

Solution Overview

- Built a bizarre news detection model for classifying news as bizarre or not.

- Used various structural, linguistic,and semantic features for classification.

- Used GloVe pre-trained vectors as features.

- **Achieved classification accuracy of 91% with Random Forest using all features**

Related work

- No prior work relating to weird and bizarre news.

- Clickbaits – A close hit.

  - Abhijnan Chakraborty, Bhargavi Paranjape, Sourya Kakarla, and Niloy Ganguly. 2016. Stop Clickbait: Detecting and Preventing Clickbaits in Online News Media.

  - Martin Potthast, Sebastian K¨opsel, Benno Stein, and Matthias Hagen. 2016. Clickbait Detection.

Conventional News vs Clickbait vs Bizarre News

Attribute	Conventional	Clickbaits	Bizarre News
Target	Inform the user	Lure to click	Inform the user
Liked by	Majority	None	Specific group
Benefit	Knowledge	Disappointment	Entertainment
Publishers interest	Viewership	Revenue	Viewership
Usefulness	Reduces with time	Time agnostic	Time agnostic
Headline	Gives summary	Lures the user	Increases the curiosity

Dataset

- Collated around 73K weird and bizarre news headlines. Gold standard for bizarre category.

Source	#News items
http://www.odditycentral.com/	6840
http://in.reuters.com/news/archive/oddly EnoughNews/	7785
http://www.upi.com/Odd News/	58300
Miscellaneous sources	2000
Total	~73000

- Crawled 75K news headlines from conventional news sites. Gold standard.
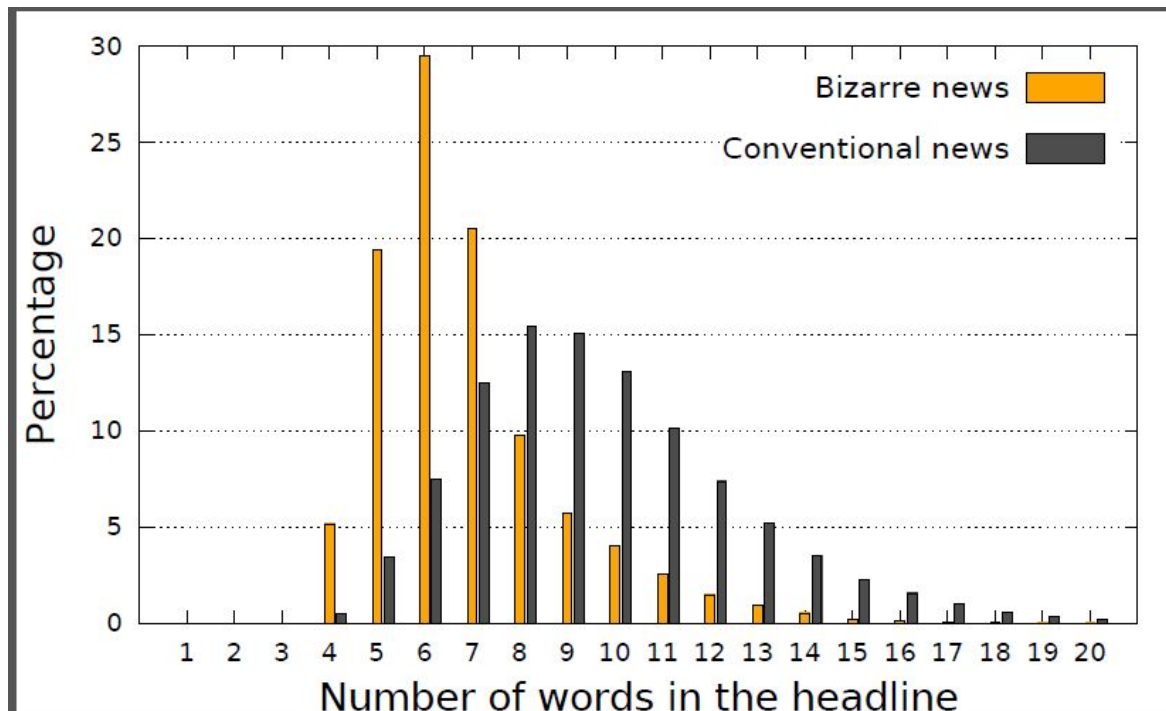
Features

- Structural and Punctuation Features
- Linguistic Features
- Named Entity and POS features
- Distributed Word Representations - GloVe
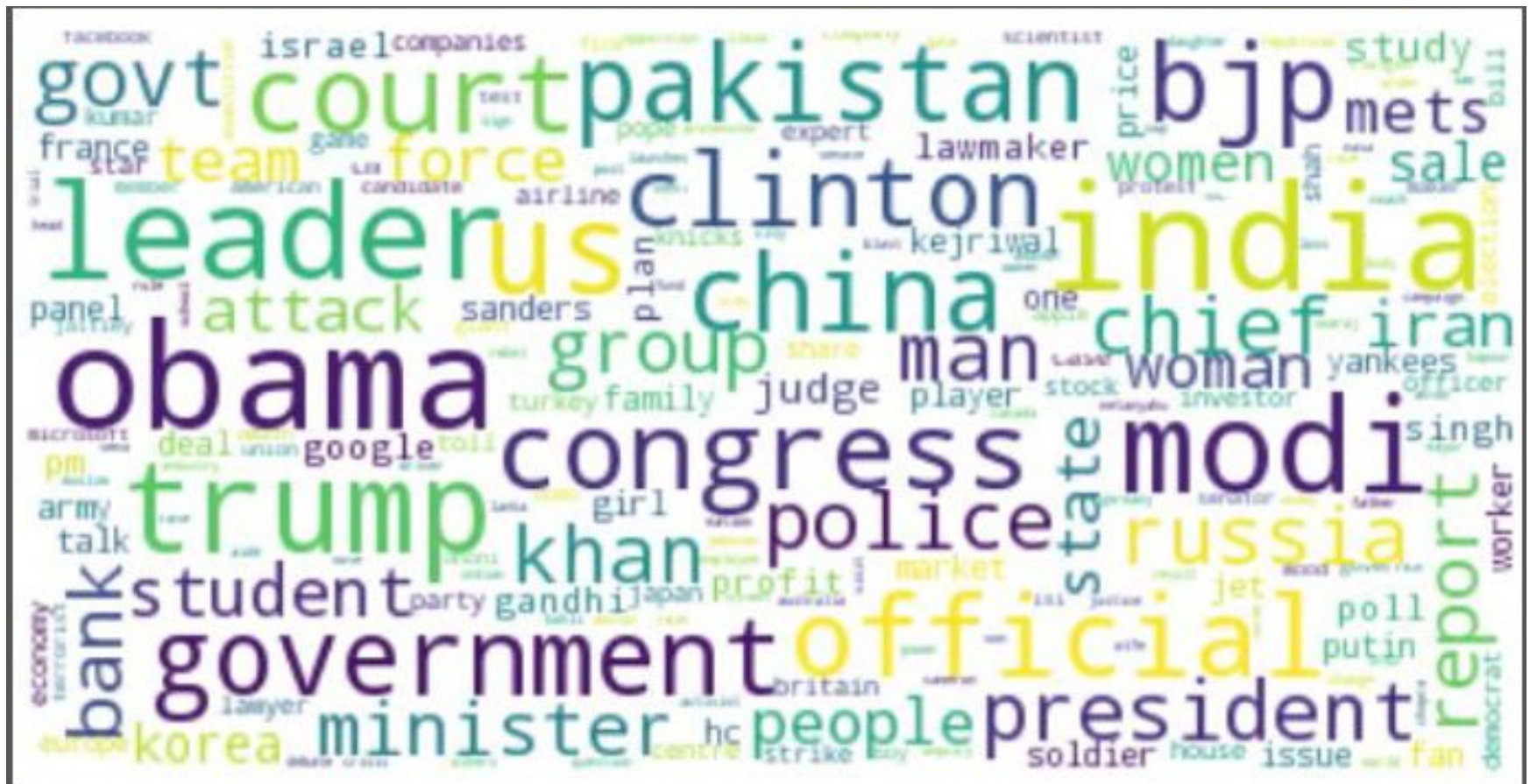
Structural and Punctuation Features

- Length of the headline, #stop words, quotes, ellipses.

Linguistic Features

- Frequency of popular subjects
- Presence of pronouns and possessives
- Acronyms and Abbreviations.

Frequently occurring subjects in conventional news

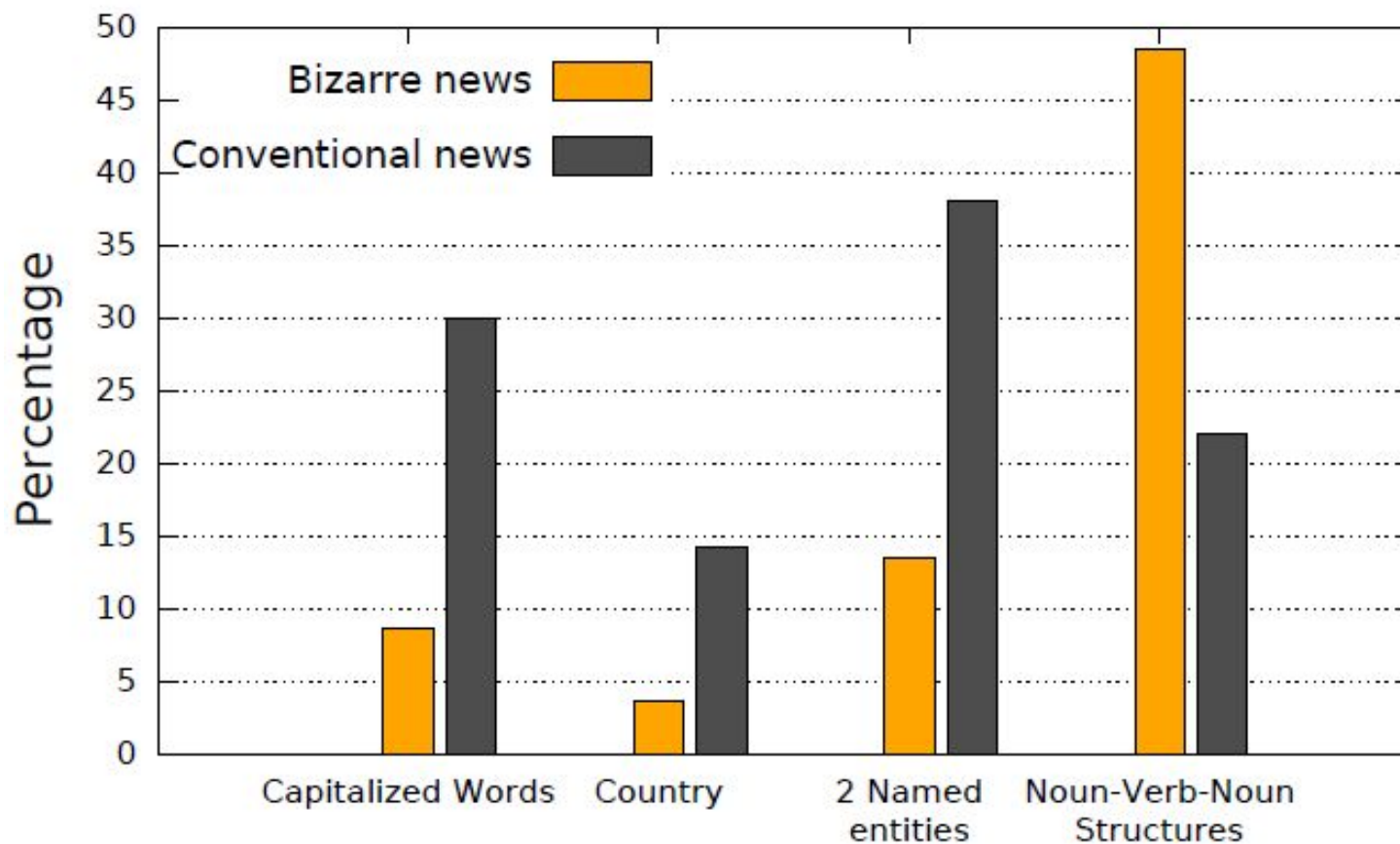# Frequently occurring subjects in bizarre news

Named Entity and POS Features

- Persons, Locations, Organizations, Country Names
- POS counts
- Animals and Plants
- Vehicles
- Noun – Verb – Noun structure

Distribution of features

Most frequently occurring animals in bizarre news

Bizarre news titles with animals

- 9-foot **crocodile** visits U.S. Navy base in Florida

- Injured **dog** limps to find help for injured owner

- Dell offers replacements for laptops smelling of **cat** urine

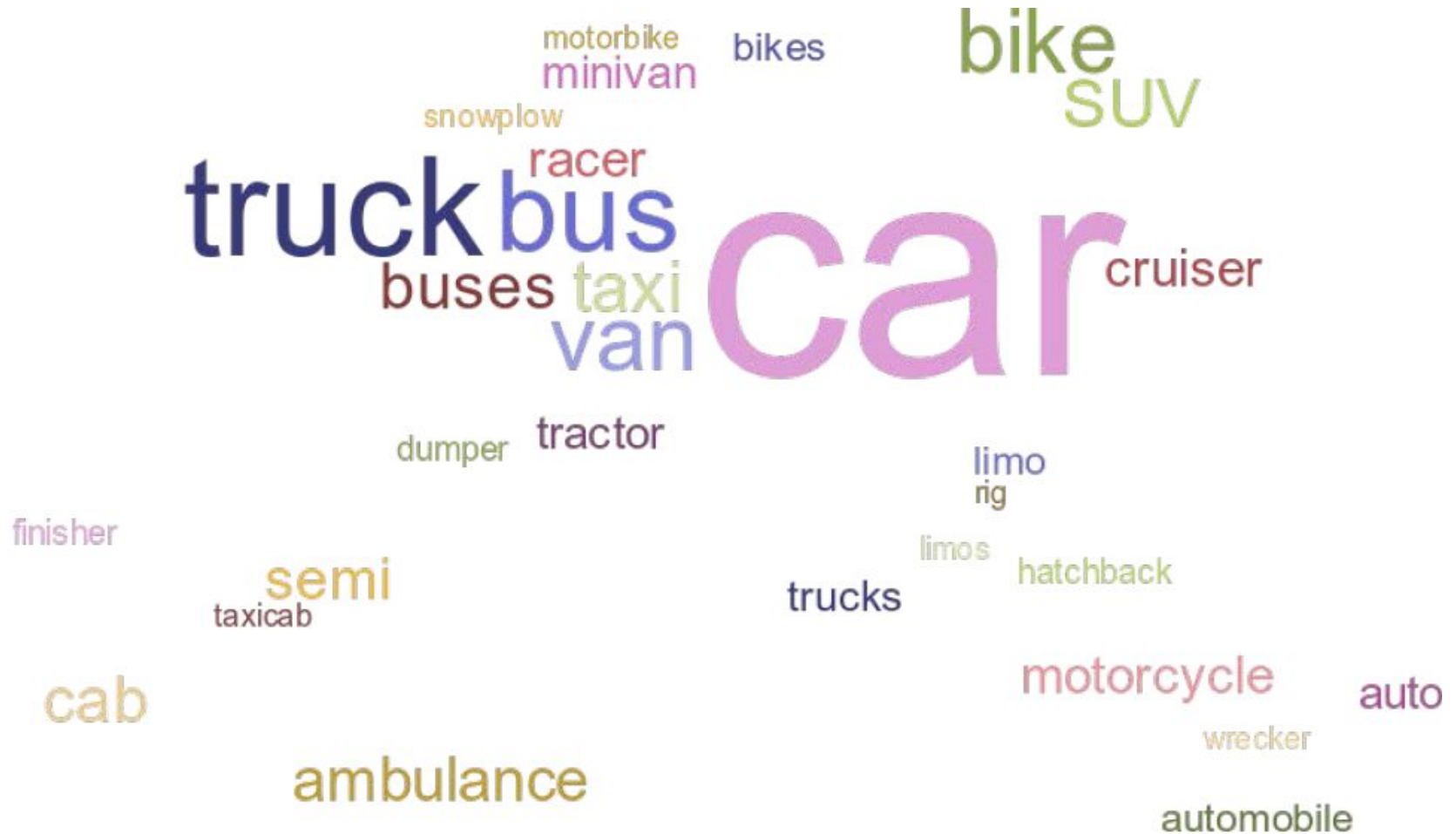- Dead **snake** found in case of Bud Light

Most frequently occurring body parts in bizarre news

Bizarre news titles with body parts

- Grandmother accused of forcing girl to hold Bible over **head** to exorcise demons.

- Obama shakes Marine's **hand** after failing to salute.

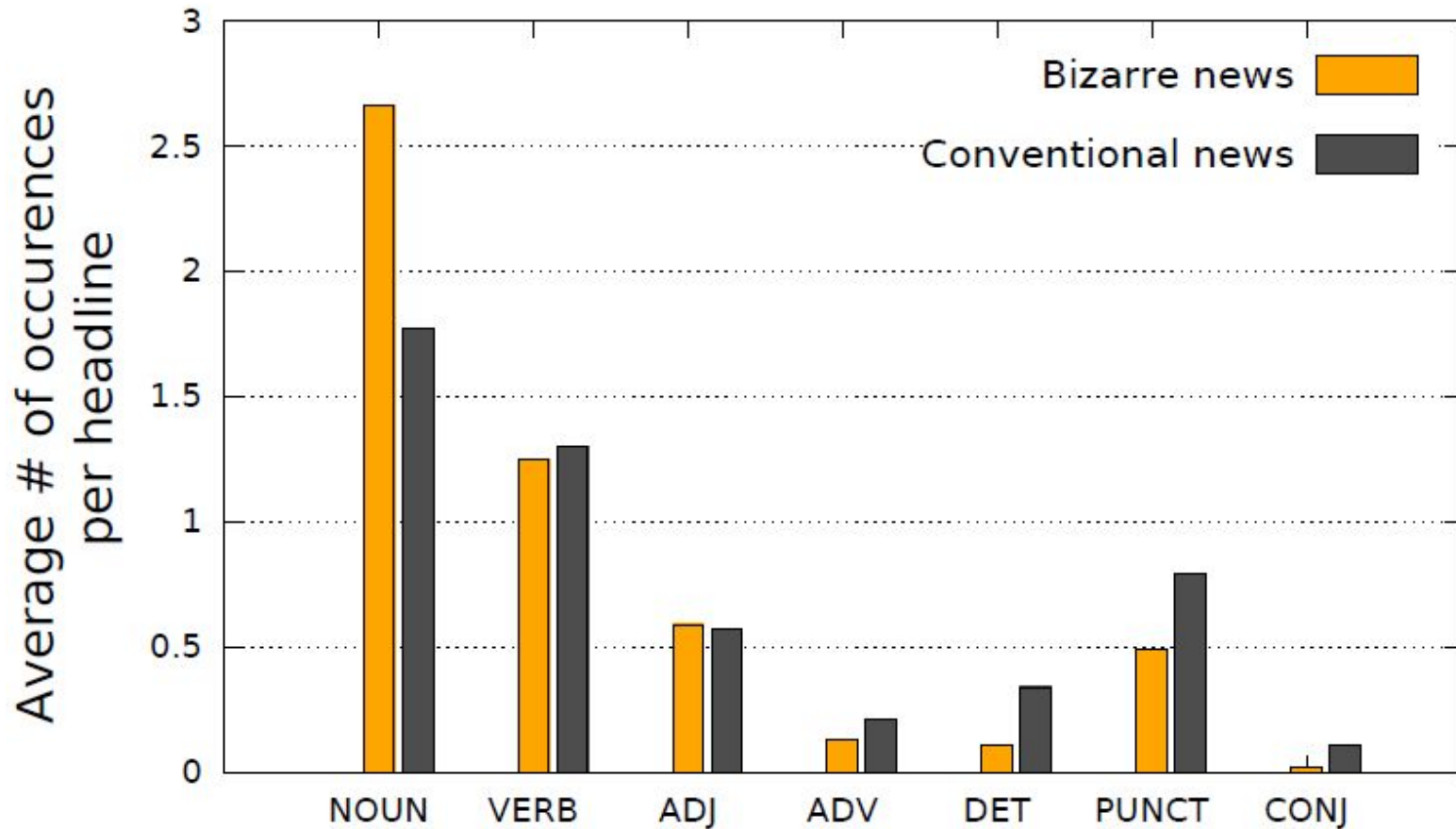- Woman whose **face** was used for Healthcare.gov fighting back against cyberbullying

Most frequently occurring vehicles in bizarre news

Bizarre news titles with vehicles

- Handcuffed suspect steals police **cruiser**
- Ice cream **truck** music riles neighborhoods
- Prison for men who stole **van** full of corpses
- Hungry **bus** driver pulls over for sandwich
- Man builds **car** for pet parrot
- Patient steals **ambulance** in Atlanta, takes paramedics on wild ride

Distribution of Parts of Speech

Distributed Word Embeddings

- Used pretrained GloVe vectors.
- 300 dimensions, Trained on Wikipedia 2014 + Gigaword 5.
- Used in 2 approaches
  - Deep learning with Word Embeddings only
  - Concatenating Word Embeddings with hand crafted features

Classification Methods

- Neural Network

- SVM

- Random Forests

- Logistic Regression

- XGBoost

- Deep Learning

Deep Learning Results

- Used CNN for classification.

Features	Precision	Recall	F1
Satic GLoVe Embeddings	0.90	0.90	0.90
Learned Embeddings	0.88	0.88	0.88

Classification Results

- All cells are F1 scores

	NN	SVM	RF	LR	XGBoost
(1) Structural and Punct.	0.83	0.83	0.84	0.83	0.84
(2) Linguistic features	0.52	0.52	0.52	0.52	0.52
(3) NE and POS features	0.78	0.77	0.80	0.77	0.81
(1)+(2)+(3)	0.87	0.87	0.88	0.87	0.88
(4) Distributed Embeddings	0.88	0.88	0.89	0.88	0.88
(1)+(2)+(3)+(4)	0.92	0.90	0.90	0.92	0.89

Top 10 most important hand crafted features

Rank	Feature description
1	Number of words in headline
2	Number of nouns in headline
3	Average word length in the headline
4	Number of stop words in the headline
5	Number of named entities in the headline
6	Presence of country in the headline
7	Presence of N-V-N structure in the headline
8	Number of verbs in the headline
9	Number of punctuation characters in the headline
10	Presence of most frequently used noun in the headline

Future work

- Rating the "weirdness" of a bizarre news.
- "Understanding" why a news is bizarre.

Key Learnings

- Carefully handcrafted features with simple ML techniques can outperform deep learning techniques.

- Augmenting pre-trained vectors to hand crafted features can improve the classification performance.

Thank you

- Questions ?