Amputation or Accident?
Making the Computer Understand Urban Legend Types (And Making Humans Understand the Computer)

Myrthe Reuver,
ReMa student Linguistics Radboud University
Research Intern Meertens Institute, Amsterdam

Introduction

Folktale Database or “Volksverhalenbank” (Prof. Dr. Theo Meder)

- Thousands legends, fairytales, and jokes and other Dutch folktales
- One genre: 3.000 urban legends (*broodjeaapverhalen*)
- Stories come from interviews, news articles, questionnaires..

First things first: What are Urban Legends?

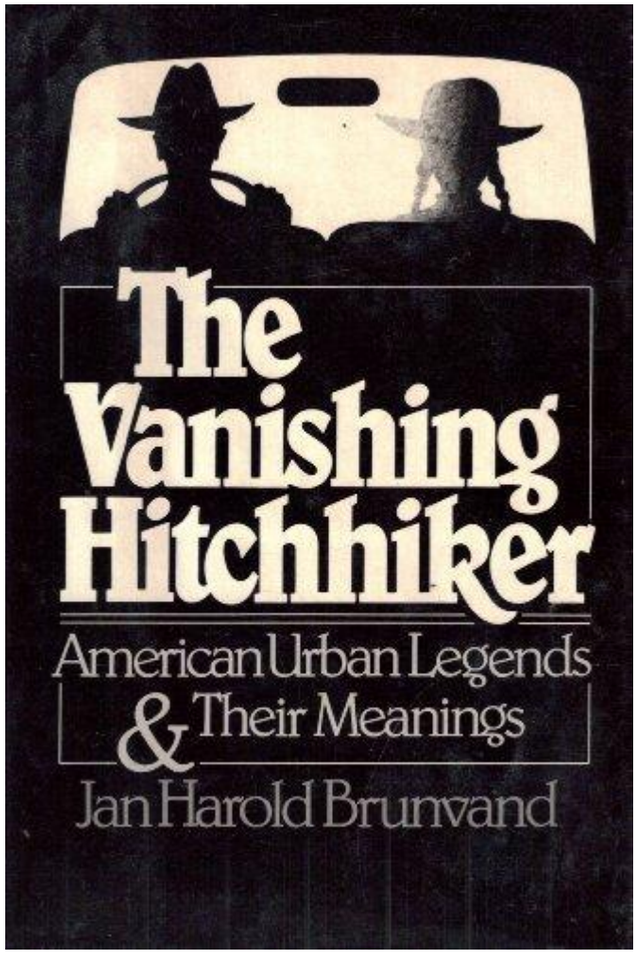
- Example: *"The Babysitter and the Man Upstairs"*

- Urban Legends interest almost everyone:
  - Psychologists when is it "sharable"?;
  - Folklorists "folk stories", "community identity", "performance";
  - Literary scholars narratives, motifs;
  - Linguists the language and genre of urban legends;
  - And computer scientists urban legend detection & ***classification***

- Working definition:
  - "Extensively *shared* narratives with emotional or sensational content, believed to be true, about *common modern anxieties* ("stranger danger", processed food), with an *underlying moral* and/or community message."

Urban Legends typology

Dr. Brunvand: “mr. Urban Legend”



Brunvand’s typology:

- Main category (*HORROR*)
- Sub-category (*Amputation*)
- Storytype (“plot” or “motif”) □ “girl finds severed arm on top of car”

One story-type can have many versions: where the journey goes to, whether the girl is on a date, in a taxi..

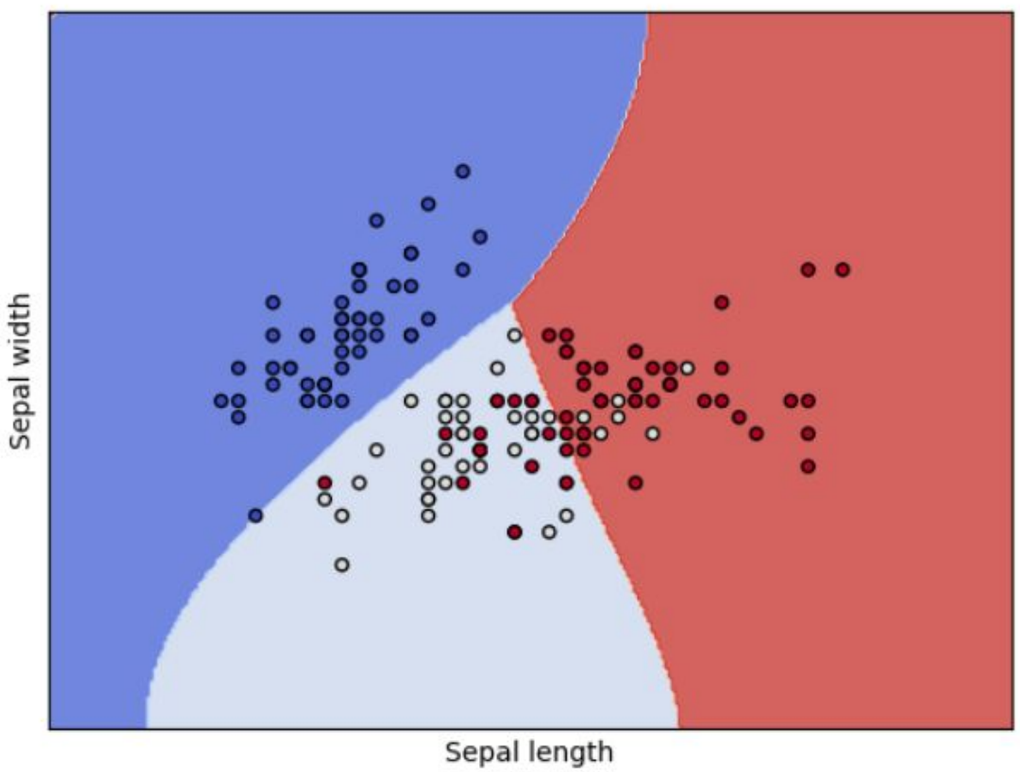
Solution: classification

Classification: **automatically** finding labels that belong to individual texts.

- Now: employees of Meertens need to manually **read** & assess each tekst

However: computers “read” and classify texts differently than humans.

Classic examples: learning the values for “roses”, “petunias”, and “daisies”:



Solution: Machine Learning, specifically: classification.

How does it work?

- An algorithms gets many examples with their **class labels** and infers rules from these examples (**the training phase**)

- Usually, these examples get **features** (abstractions) so it is easier to generalize.

- With the extracted rules, the algorithm can classify a new, unseen example or text (**the test phase**)

A computer “learning” text:

- **Input text:** “She has slightly long hair wearing a black shirt with the light of new knowledge on it”
- **Transform to sparse matrix (every word is a number with a weight):**

  (example of first 3 words, all words + bigrams + trigrams have weight 1):

  (0, 9740) 1

  (0, 8062) 1

  (0, 12217) 1
- **TF-IDF (Inverse term frequency) weighting on this vector:**

  (0, 15062) 0.16410407328747026

  (0, 15059) 0.1574035312175246

  (0, 15054) 0.1071220560481581

Problem 1: Unimportant Words

Remember: all urban legends have *sources*:

- Interviews
- Questionnaire answers
- News articles
- Email-chains (!)

We tried predicting **source labels:** “newspaper”, “interview”

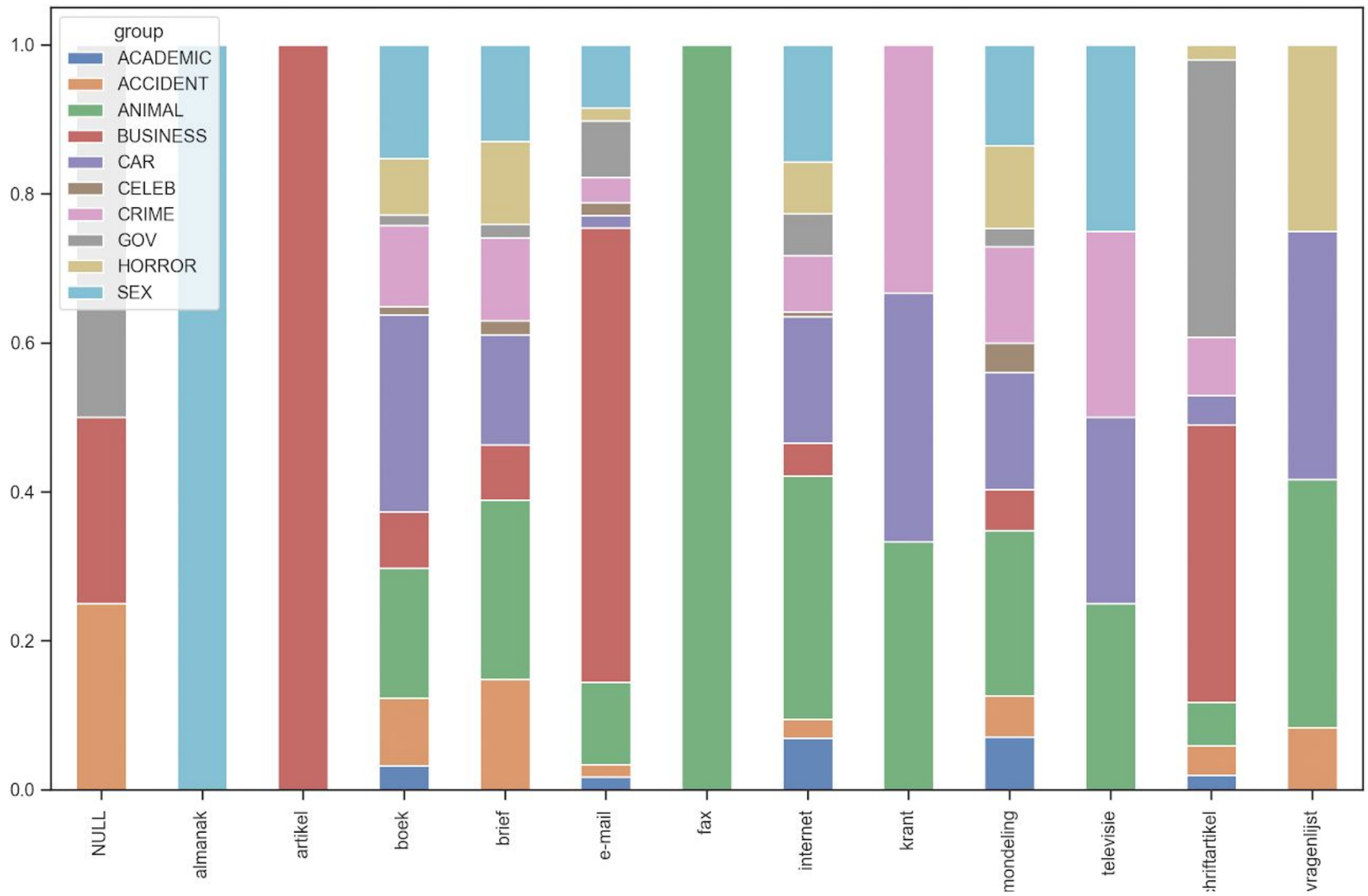
And also predicting **type labels**: “HORROR”, “

 sometimes, most predictive words were the same (i.e. a *confound*):

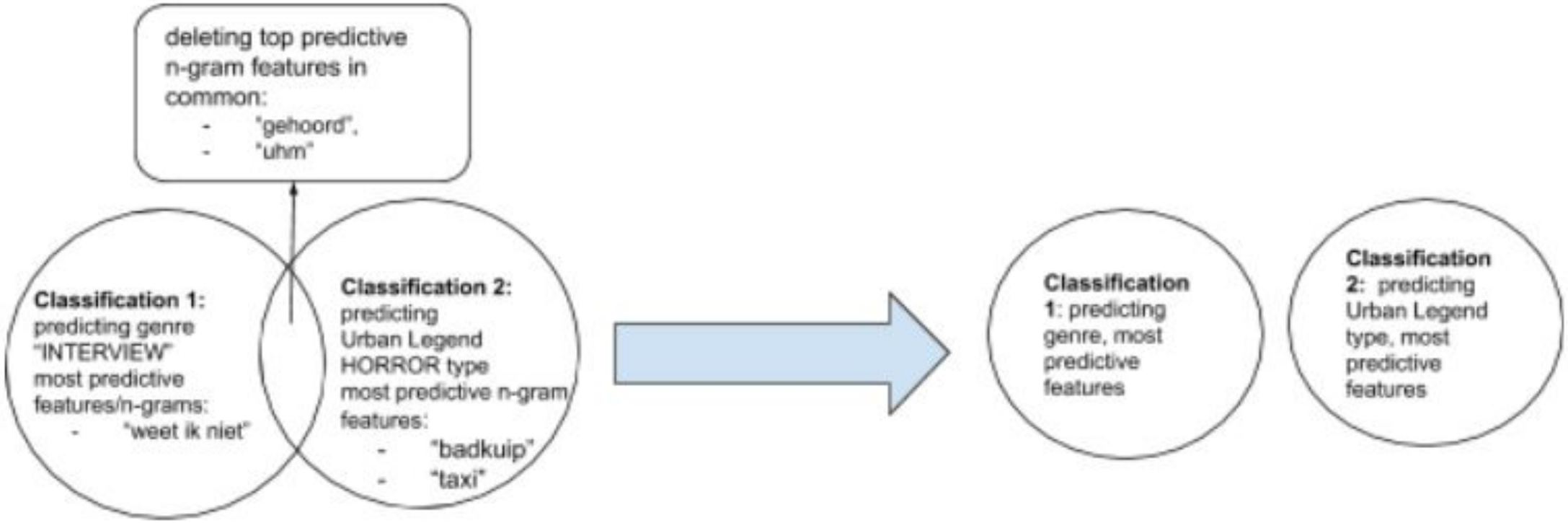
“Published on”, “kuch/hoest/nou”, “verzonden”

In other words: sometimes the model predicted *source*, not urban legend, especially when legend type and source were closely connected!

Visualization of class imbalance and source confound



Cleaning the source confound with Quoll



Cleaning the source confound

Old:

[11.37]RH: Het merkwaardige was, toen ik, toen men daarover praatte kwamen
dr meteen andere verhalen over **uhm** over over kunstgebitten los.
Daar was **uhm** een prachtig **eh** verhaal van een van de deelnemers die zei in Princenhage hier ging men regelmatig in de zomer enkele malen vissen op zee, gingen ze zeevissen,
en wanneer dan **eh** de hengels uitgeworpen waren en die stonden vast **[…]**

New:

Het merkwaardige was, toen ik, toen men daarover praatte kwamen
dr meteen andere verhalen over over over kunstgebitten los.
Daar was een prachtig verhaal van een van de deelnemers die zei in Princenhage hier ging men regelmatig in de zomer enkele malen vissen op zee, gingen ze zeevissen,
en wanneer dan de hengels uitgeworpen waren en die stonden vast in hun standard [..]

Problem 2: 176 labels

Labels for urban legends:

10x MAIN labels

SUBTYPE labels

TYPE labels

HORROR

MEDICAL HORROR

BRUN 03155: Accidental Cannibalism

Problem 2: 176 labels

- New idea: predicting 176 labels is hard, so why not start with 10 labels and make a **hierarchical model?**

- **Learner: Support Vector Machine with 10fold CV from scikitlearn**
- **Features: simple word counts with TF-IDF weight on cleaned texts**
- **10% test set (“unseen” examples)**

Classification: results

level 1:
accuracy 10CV = .59 // F1 score macro = .57 // F1 score micro = .67 (10 labels)

level 2:
accuracy 10CV = .53 // F1 score macro = .34 // F1 score micro = .46
(43 labels)

level 3:
accuracy 10CV = .32 // F1 score macro = .23 // F1 score micro = .36 (173 labels

- Accuracy: how many texts are correctly classified
- F1 score: harmonic mean between precision (how many of the items I identify are really the right class?) and recall (how many items do I identify of the total items of the class?
- Earlier work (Nguyen, 2012) obtained higher metrics (.70), but likely did not account for source confound

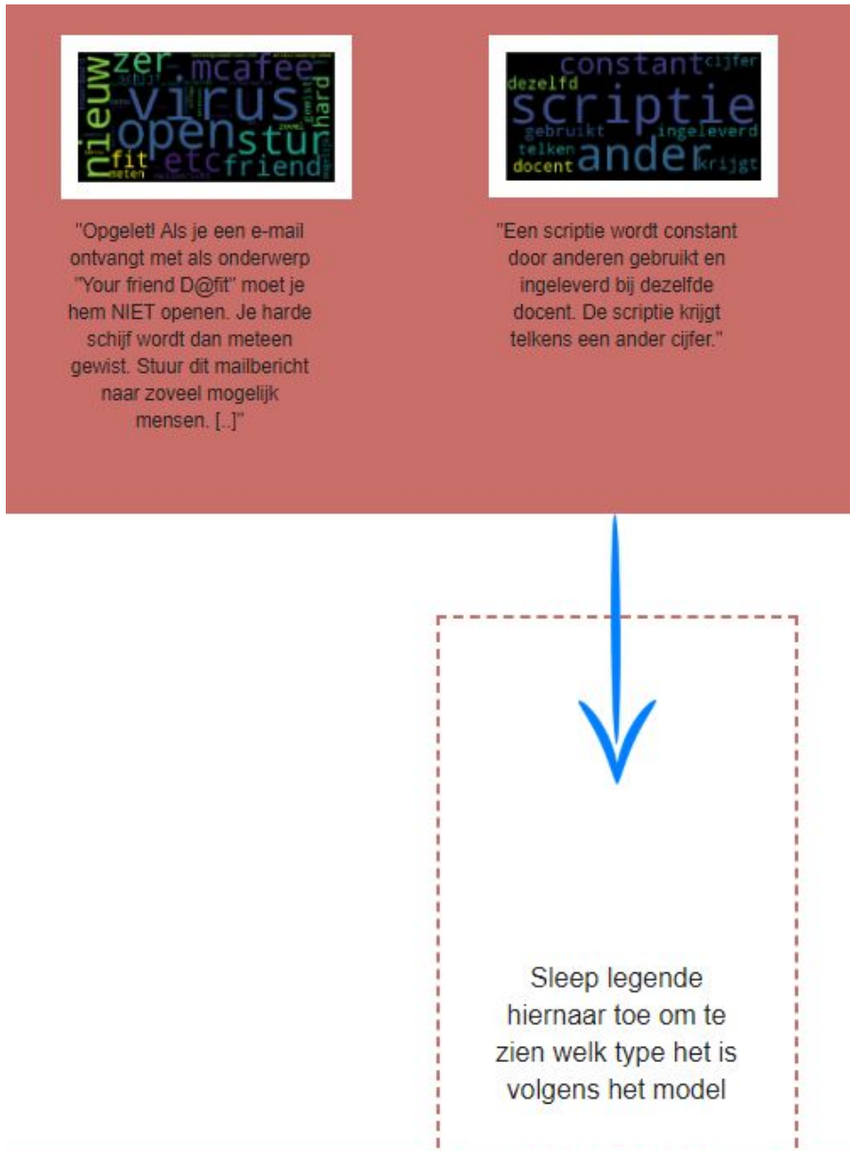
Interesting find: Not All Classes Are Created Equal

- There is a large class inequality: some types have 50 examples, some only two.

- However, not all “large” classes are recognized well:
  - 37 examples of BRUN 05515, "Masturbating Into Food”, in the training set, but is fairly badly recognized (F1 = .33). Why? Likely because all concern different foods and different settings. The same counts for “Kidney Heist”, BRUN 06305.
  - Other with many examples (“Big Cat on the Loose”, 46 in the training set”) are fairly well recognized (F1 = .78), likely because (nearly) all are about “puma on the Veluwe”.

- The same counts for “small” classes, not all do badly:
  - “Poodle in the microwave” has 16 examples in the training set, but a high F-score (.86), “Pregnancy through the bathtub” is similar (F-score = 1).
  - The category “Tourist Horror Stories” gets F1 = 0, likely because this is a rest category.

Humans and computers do not read the same, but can they work together?

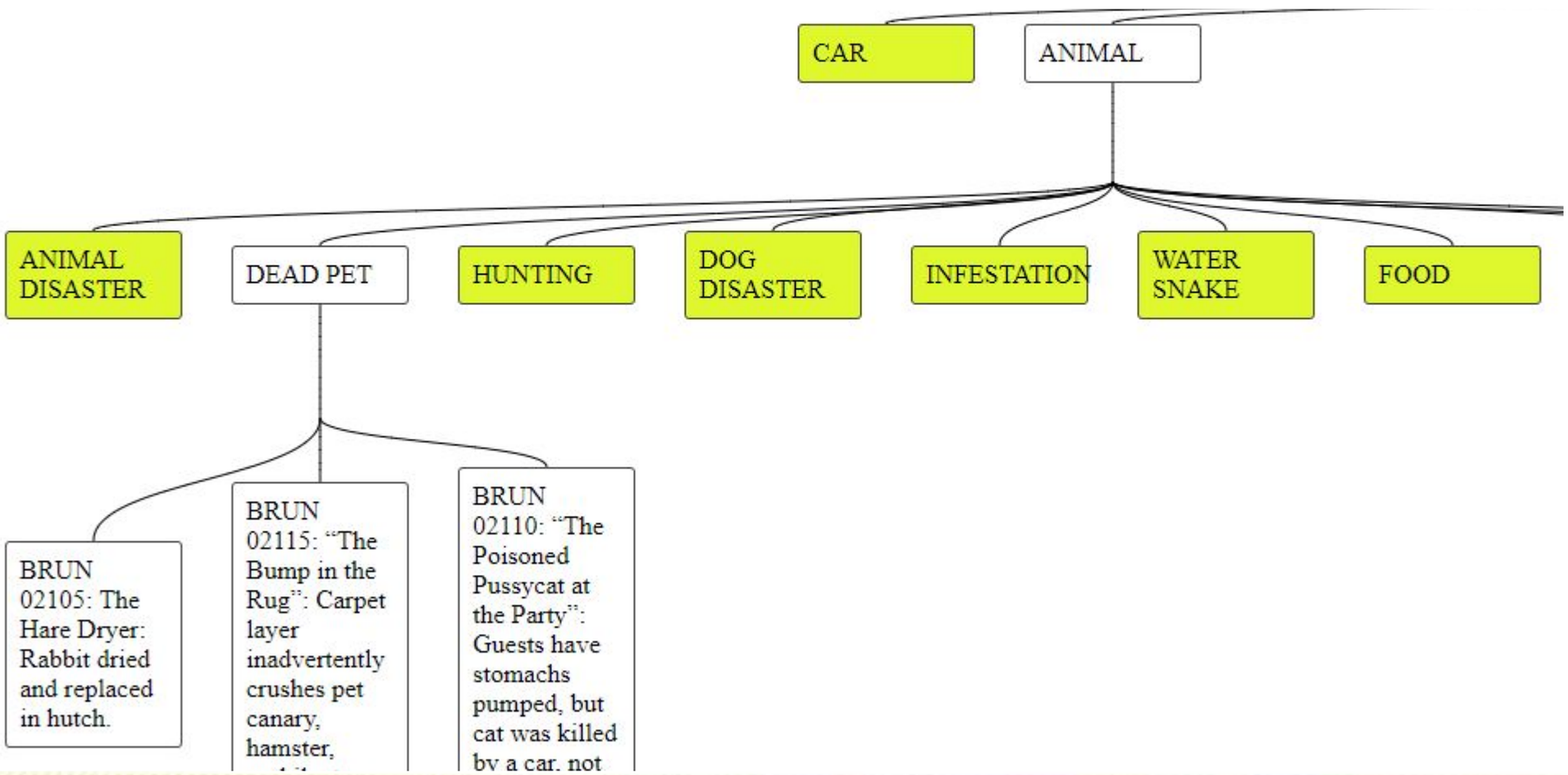


Demo: https://myrthereuver.github.io/UrbanLegendCategorizer/

Conclusion

- The computer can “read” urban legends and classify them, but is easily confused by specific features of urban legends: **knowing your data** helps.

- Annotator and human could work together to find the answer to urban legend classification



References and Sources

- Best, J. and Horiuchi, G. 1985. The Razor Blade in the Apple: The Social Construction of Urban Legends. Social Problems 32(5), 488-499.
- Brunvand, J.H. 2002. Encyclopedia of Urban Legends. W.W. Norton & Company.
- Fine, G. 1980. The Kentucky fried rat. Journal of the Folklore Institute 17, 222-43.
- Fine, G. 1985. The Goliath effect. Journal of American Folklore 98, 63-84.
- Meder, T., Karsdorp, F., Nguyen, D., Theune, M.,Trieschnigg, D. and Muiser, I. 2016. Automatic Enrichment and Classification of Folktales. Journal of American Folklore 129, 76–94.
- Mullen, P. 1972. Modern Legends and Rumor Theory. Journal of the Folklore Institute, 92(3), 95-10
- Nguyen, Dong, Trieschnigg, Dolf and Theune, M. 2013. Folktale classification using learning to rank. In 35th European Conference on IR Research, ECIR 2013, 195-206.

Related Definitions

genre	intention	aim	truth value	creator	medium	spread
fake news	disinformation and intentional misleading/convincing	convincing	not all untrue, but more framing/misleading	one person or organization	written; narrative	social media; **extensivity = debated**
clickbait/ junknews	wanting clicks and attention	earning money	low journalistic standards	one person or company	written; narrative	social media; **extensive**
Rumours	social bonds creation, relieving anxiety/uncertainty	earning trust and finding meaning	**usually untrue by definition**	community	oral; non-narrative	very local
Urban Legends	moral component: entertainment but also community forming and (emotional) sense-making	finding meaning; community forming; "performance"	not by definition untrue, and believed to be true	community	oral; traditional form and function; **motifs**	very widely

All:

- **sensationalist**
- **emotional content to get people to share?**
- **<u>weak factual basis</u>**
- **usually people (teller, recipient, or both) believe them to be true**
- **being 'shared', going 'viral', spreading through network**
  - **"underbelly" ?**

Earlier research on Urban Legend classification

Nguyen et. al. 2013 worked with the Meertens Urban Legends corpus in 2012, but only with 700 texts.

Accuracy around .70, but: **might not have accounted for genre confound.**