Discovering Sound Concepts and Acoustic Relations in Text

Anurag Kumar, Bhiksha Raj, Ndapandula Nakashole

Language Technologies Institute
School of Computer Science, Carnegie Mellon University

Abstract

- **Cataloging, Understanding, Interpreting and Relating Sounds**
- Natural Language Understanding for sounds
- **Audible Phrases** - Phrases we use to describe sounds or sound events or sound concepts
- Understanding and interpreting sounds - Higher level semantic information

Machine Perception of Sounds

Machines should

- know about or able to find various sounds - **catalog sounds**
- know or be able to find relationships between them - **understanding sounds**
- be able to recognize and detect them in audio - **Audio Event Detection**

Cataloging Sounds

- **Is there a large list of sounds ?**
  - Few hand crafted taxonomies for soundscapes
  - Too small, Too subjective to be of any major use
- **Identifying "Audible Phrases"**
- **Sounds are result of action on interaction between objects**
  - Same source different actions, Same action different sources
- *Car, Jackhammer, Garage door, washing dishes* – used to denote sounds
- A variety of ways to describe sounds

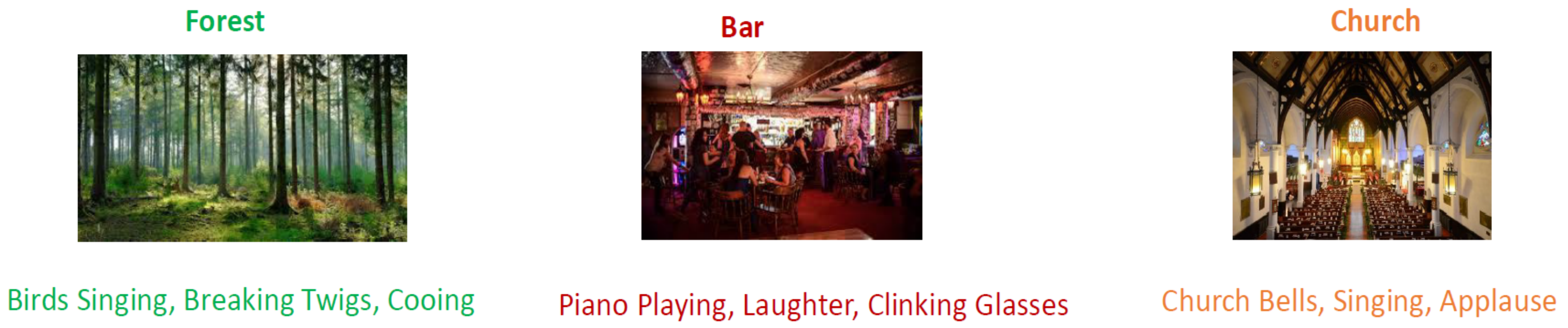
Forest
Bar
Church
Birds Singing, Breaking Twigs, Cooing
Piano Playing, Laughter, Clinking Glasses
Church Bells, Singing, Applause

Figure 1: **Examples of sounds found for a few scenes**

Cataloging Sounds

- Discover potential sound concepts and filter
- **Start with a simple pattern E.g <sound of Y>**
- Sound of gunshots, sound of man yelling
- Unsupervised Filtering - Generalize phrases by Parts of Speech Tag
- Sound of man yelling - sound of NN VBG
- 6 Patterns which expresses sound
- Supervised Filtering - Label and Train a classifier

Cataloging Sounds - Results

- Clueweb corpus - 500 million webpages
- Final List - ***116,729 sound concepts***

Pattern		**Example Concept**
P1	$<X>$of (DT) VBG NN(S)	honking cars
P2	$<X>$of VBG	yelling
P3	$<X>$of (DT) NN(S) VBG	dogs barking
P4	$<X>$of (DT) NN(S)	gunshots
P5	$<X>$of (DT) NN NN(S)	string quartet
P6	$<X>$of (DT) JJ NN(S)	classical music

Table 1: Patterns for discovered sound concepts in text. VBG is the part of speech tag for verbs, NN for nouns, DT for determiners, and JJ for adjectives.

- Manual Inspection - 100 most frequent phrase from each pattern
- **Overall positve hit rate - 77**%
- For 4 patterns - Average around **88**%

	Pattern	+ in 100 Most Freq.
P1	$<X>$of (DT) VBG NN(S)	**98**
P2	$<X>$of VBG	**71**
P3	$<X>$of (DT) NN(S) VBG	**91**
P4	$<X>$of (DT) NN(S)	59
P5	$<X>$of (DT) NN NN(S)	**93**
P6	$<X>$of (DT) JJ NN(S)	49

Table 2: + Hit Rate - 100 Most Frequent

- Supervised - Word Embedding + Linear Classifier on ∼ 6000 phrases
- Result - Accuracy of around **90**%

Understanding, Interpreting and Relating Sounds

- DCASE 2016 Challenge - Dishes, Object Banging
- What does Dishes , Object Banging, Screaming represent ?
- The catalog carries a lot of information on its own
- Source-Sound, Scene-sound relations

Scene - Sound Relations

- **What type of sounds can be found in an environment ?**
- Commonsense knowledge for humans
  - Park - Children Laughing, Birds Chirping
  - Construction Site - Hammering, Jackhammers, Blasting
- **A relation classification task**
- Sentences where a scene and at least one of sound concept occur
- Relate scene and sound concept through *dependency paths*
- Label most frequent dependency paths as positive or negative
- Train a classifier on the labeled examples
- **Unusual cases**
  - Library - Chirping Birds
  - Church - Rifle Shots

Conclusions

- A first step towards NLU for sounds
- **Largest vocabulary of sound events**
- Higher level semantic information using sounds

Additional Info

- Visit webpage *http://www.cs.cmu.edu/~alnu/SOExpt.htm* for full sound catalog and more scene-sound relations

Contact Information

- Anurag Kumar
- alnu@andrew.cmu.edu
- www.cs.cmu.edu/~alnu