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 be 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

Cataloging Sounds - Results

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

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

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

<table>
<thead>
<tr>
<th>Pattern</th>
<th>+ in 100 Most Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1.&lt;X&gt;.of (DT) VBG NN(S)</td>
<td>98</td>
</tr>
<tr>
<td>P2.&lt;X&gt;.of VBG</td>
<td>91</td>
</tr>
<tr>
<td>P3.&lt;X&gt;.of (DT) NN(S) VBG</td>
<td>91</td>
</tr>
<tr>
<td>P4.&lt;X&gt;.of (DT) NN(S) VBG</td>
<td>91</td>
</tr>
<tr>
<td>P5.&lt;X&gt;.of (DT) NN NN(S)</td>
<td>93</td>
</tr>
<tr>
<td>P6.&lt;X&gt;.of (DT) JJ NN(S)</td>
<td>49</td>
</tr>
</tbody>
</table>

Table 2: + Hit Rate - 100 Most Frequent

Understanding, Interpreting and Relating Sounds

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

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/SOEExpt.htm for full sound catalog and more scene-sound relations

Contact Information

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

Figure 1: Examples of sounds found for a few scenes

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.