Rishabh Nigam 10598

Cognitive Sciences

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへぐ

Abstract: Collected data from audio, vedio, kinect and RFID tags to augment raw data with annotations for actions performed. The action in this case is making a cup of tea.

Abstract: Collected data from audio, vedio, kinect and RFID tags to augment raw data with annotations for actions performed. The action in this case is making a cup of tea.

Goal: Cognitive Assistance for everyday's task

- Abstract: Collected data from audio, vedio, kinect and RFID tags to augment raw data with annotations for actions performed. The action in this case is making a cup of tea.
- Goal: Cognitive Assistance for everyday's task
- Related work: the CMU Multi-Modal Activity Database (2009) is a corpus of recorded and annotated video, audio and motion capture data of subjects cooking recipes in a kitchen.[1]

- Abstract: Collected data from audio, vedio, kinect and RFID tags to augment raw data with annotations for actions performed. The action in this case is making a cup of tea.
- Goal: Cognitive Assistance for everyday's task
- Related work: the CMU Multi-Modal Activity Database (2009) is a corpus of recorded and annotated video, audio and motion capture data of subjects cooking recipes in a kitchen.[1]
- Difference: Here we also include 3-d data using Kinect, the subject verbally describes what he is doing and there are attached anotations to each action performed.

 Audio – 3 microphones – to capture what the subject is using to describe the task he/she is performing.

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 臣 のへぐ

 Audio – 3 microphones – to capture what the subject is using to describe the task he/she is performing.

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 臣 のへぐ

Vedio – HD vedios

 Audio – 3 microphones – to capture what the subject is using to describe the task he/she is performing.

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへぐ

- Vedio HD vedios
- Kinect RGB + depth data

- Audio 3 microphones to capture what the subject is using to describe the task he/she is performing.
- Vedio HD vedios
- Kinect RGB + depth data
- RFID tags : The subject was supposed to wear an RFID sensing iBracelet which records the RFID tag closest to the wrist at any time. sensors attached to Kitchen appliances to give better data on which instrument is used.

- Audio 3 microphones to capture what the subject is using to describe the task he/she is performing.
- Vedio HD vedios
- Kinect RGB + depth data
- RFID tags : The subject was supposed to wear an RFID sensing iBracelet which records the RFID tag closest to the wrist at any time. sensors attached to Kitchen appliances to give better data on which instrument is used.
- Power Consumption: use of electric kettle and we determine using power consumption whether the kettle is on or not.

The Audio data was transcribed and transcription was segmented to utterances. Breaks as one speaks where used to mark the end of sentences. If an utterance were not complete sentences, longer pause where looked at.

- The Audio data was transcribed and transcription was segmented to utterances. Breaks as one speaks where used to mark the end of sentences. If an utterance were not complete sentences, longer pause where looked at.
- Then it uses a parser using semantic lexicons to create the logical form, the semantic representation of the language

- The Audio data was transcribed and transcription was segmented to utterances. Breaks as one speaks where used to mark the end of sentences. If an utterance were not complete sentences, longer pause where looked at.
- Then it uses a parser using semantic lexicons to create the logical form, the semantic representation of the language
- IM (Interpretation manager) was used to extract a concise event description from each clause, derived from each main verb and its arguments.

e.g. Place tea bag in the cup => PUT THE TEA BAG INTO THE CUP.

- The Audio data was transcribed and transcription was segmented to utterances. Breaks as one speaks where used to mark the end of sentences. If an utterance were not complete sentences, longer pause where looked at.
- Then it uses a parser using semantic lexicons to create the logical form, the semantic representation of the language
- IM (Interpretation manager) was used to extract a concise event description from each clause, derived from each main verb and its arguments.

e.g. Place tea bag in the cup => PUT THE TEA BAG INTO THE CUP.

To learn the name of the given IDs that the audio description has, we gather the nouns mentioned by the subject, convert them into ontological concepts using parse data and determine the concept with the highest probability of being mentioned when that ID is detected

- The Audio data was transcribed and transcription was segmented to utterances. Breaks as one speaks where used to mark the end of sentences. If an utterance were not complete sentences, longer pause where looked at.
- Then it uses a parser using semantic lexicons to create the logical form, the semantic representation of the language
- IM (Interpretation manager) was used to extract a concise event description from each clause, derived from each main verb and its arguments.

e.g. Place tea bag in the cup => PUT THE TEA BAG INTO THE CUP.

To learn the name of the given IDs that the audio description has, we gather the nouns mentioned by the subject, convert them into ontological concepts using parse data and determine the concept with the highest probability of being mentioned when that ID is detected

Results

While they only have a small amount of data, the labels generated by the algorithm agreed with a human anno-tator, who used the video to determine the mappings, for six out of the eight tags.

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 臣 の�?

References

[1] http://kitchen.cs.cmu.edu/

[2] Mary Swift, George Ferguson, Lucian Galescu, Yi Chu, Craig Harman, Hyuckchul Jung, Ian Perera, Young Chol Song, James Allen, Henry Kautz "A multimodal corpus for integrated language and action", Department of Computer Science, University of Rochester, Rochester, NY 14627