Language and Classification

Sravana Reddy

Nov 9, 2015
Why isn't the future here already?
Where is my evil robot space travel companion?
How close are we to this?

https://www.youtube.com/watch?v=n1AjIAlJe3o
Machine conversations: Where do we begin?

1. Convert voice input into words
Machine conversations: Where do we begin?

2. Convert words into meaning

Merhaba!
Machine conversations:
Where do we begin?

3. Use world knowledge to respond
Machine conversations:
Where do we begin?

4. Convert response into words
Machine conversations:
Where do we begin?

5. Convert words into natural-sounding speech
Machine conversations: Where do we begin?

1. Convert voice input into words  speech recognition
2. Convert language into meaning  language understanding
3. Use world knowledge to respond  "core AI", search, logic
4. Convert response into words  language generation
5. Convert words into natural-sounding speech  speech synthesis
But that's just the minimum

• Judge *emotion* from voice and word usage

• Identify the general *topic* of conversation

• Identify the speaker

• Identify the speaker's *gender* or *ethnicity*
E-mail Classification

Dear Faculty/Staff,

The following is information regarding self-scheduled and fixed exam policies and deadlines. To see schedules of fixed exams and self-scheduled exams and other important exam information, go to the Registrar's website at: http://wellesley.edu/registrar/exams

Classroom locations for fixed exams are being finalized and will be posted when confirmed.

Hi Sravana,

Hope you're still doing well at Wellesley. One more reminder letter for my NSF application is due on November 5 (Thursday) can still finish it in time.

Hello Sravana, Paul,

I hope your semester is going well!

Paul, is there still interest at

3rd INTERNATIONAL CONFERENCE ON STATISTICAL LANGUAGE AND SPEECH
SLSP 2015
Budapest, Hungary
November 24-26, 2015
Organised by:
E-mail Classification

Dear Faculty/Staff,

The following is information regarding self-scheduled and fixed exam policies and deadlines. To see schedules of fixed exams and self-scheduled exams and other important exam information, go to the Registrar’s website at http://wellesley.edu/registrar/exams

Classroom locations for fixed exams are being finalized and will be posted when confirmed.

Final Exams

- Students may not take midterms at home.
- Final exam period (Dec 14, 15, 16, 17) is mandatory. No make-up exams are scheduled.

Hi Sravana,

Hope you're still doing well at Wellesley. One more reminder letter for my NSF application is due on November 5 (Thursday) and can still finish it in time.

Hello Sravana, Paul,

I hope your semester is going well!

Paul, is there still interest at the conference?

3rd INTERNATIONAL CONFERENCE ON STATISTICAL LANGUAGE AND SPEECH
SLSP 2015
Budapest, Hungary
November 24-26, 2015
Organised by:
E-mail Classification

Dear Faculty/Staff,

The following is information regarding self-scheduled and fixed exam policies and deadlines. To see schedules of fixed exams and self-scheduled exams and other important exam information, go to the Registrar’s website at: http://wellesley.edu/registrar/exams

Classroom locations for fixed exams are being finalized and will be posted when confirmed.

Final Exams

- Students may not take exams during the final exam period (Dec 14, 15, 17).

Hi Sravana,

Hope you’re still doing well at Wellesley. One more reminder letter for my NSF application is due on November 5 (Thursday) can still finish it in time.

Hello Sravana, Paul,

I hope your semester is going well!

Paul, is there still interest at...

---

3rd INTERNATIONAL CONFERENCE ON STATISTICAL LANGUAGE AND SPEECH

SLSP 2015

Budapest, Hungary

November 24-26, 2015

Organised by:
Building Classifiers

- Labeled Examples
  - Model
  - training

- Unlabeled Example
  - Label
  - application

- Unlabeled Example
  - Label
Classification Algorithms

• Naïve Bayes
• Logistic Regression
• Perceptrons
• Support Vector Machines
• Decision Trees
• Gaussian Mixture Models
• K Nearest Neighbors
• Neural Networks
• ...
Sentiment Analysis

https://www.csc.ncsu.edu/faculty/healey/tweet_viz/tweet_app/

http://alt.qcri.org/semeval2016/task4/
Gender (or other demographic) Prediction
Solving Authorship Mysteries!

James Madison

Alexander Hamilton

The Federalist Papers
Other Features?

• Fine-grained language:
  – Word sequences
  – Grammar
  – Word meanings

• User information

• Social network
Classification for more complex tasks?

1. Convert voice input into words  
   *speech recognition*

2. Convert language into meaning  
   *language understanding*

3. Use world knowledge to respond  
   *"core AI", search, logic*

4. Convert response into words  
   *language generation*

5. Convert words into natural-sounding speech  
   *speech synthesis*
Classifiers with Structured Outputs

Labeled Examples

Model

Unlabeled Example

Unlabeled Example

Data: Audio. Label: sentence
Data: Spanish sentence. Label: English sentence
Data: Meaning. Label: sentence

Sentence

application

training