What’s Hot in Human Language Technology:
Highlights from NAACL HLT 2015

Joyce Y. Chai, Michigan State University
Anoop Sarkar, Simon Fraser University
Rada Mihalcea, University of Michigan
A Quick Glance

• 14\textsuperscript{th} Conference of the North American chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT)

• New records in NAACL history
  – Received 714 submissions (overall 26\% acceptance rate)
  – Closed to 1000 registered participants

• Eighteen topic areas
Topic Areas

- Language and Vision
- Information Retrieval
- Dialogue and Interactive Systems
- Linguistic and Psycholinguistic
- NLP-enabled Technology
- Discourse and Pragmatics
- Spoken Language Processing
- Phonology and Morphology and Word...
- Sentiment Analysis and Opinion Mining
- Text Categorization and Topic Models
- Language Resources and Evaluation
- Generation and Summarization
- Machine Learning for NLP
- Tagging and Chunking and Syntax and...
- Information Extraction and Question...
- Machine Translation
- NLP for Web and Social Media and...
- Semantics
Invited speaker: Fei-Fei Li, Stanford University
  – Title: *A Quest for Visual Intelligence in Computers*
Main problems:
  – Alignment between language instructions and video segments (Malmuaud et al., 2015; Naim et al., 2015)
  – Video/Image to text generation (Ortiz et al., 2015; Venugopalan et al., 2015).
Discriminative Unsupervised Alignment of Natural Language Instructions with Corresponding Video Segments

Iftekhar Naim, Young Chol Song, Qiguang Liu, Liang Huang, Henry Kautz, Jiebo Luo, and Daniel Gildea

NAACL, 2015

Credit: I. Naim
Unsupervised Alignment

- Unsupervised Grounded Language Learning
  - No manual annotation
  - Intuition: Objects and actions in videos usually appear in the same sequential order as their text mentions.

Latent alignment: (1) Sentences to Video and (2) Nouns to Blobs

Credit: I. Naim
Unsupervised Alignment Models

- Generative Alignment (Naim et al. AAAI [2014])
  - Hidden Markov Model (HMM) + IBM Model 1
- Discriminative Alignment
  - Latent CRF, Latent Perceptron, and Latent SVM

<table>
<thead>
<tr>
<th>System</th>
<th>Alignment Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Manual Tracking</td>
</tr>
<tr>
<td>HMM</td>
<td>75.50</td>
</tr>
<tr>
<td>Latent HMM</td>
<td>75.58</td>
</tr>
<tr>
<td>LSP-H</td>
<td>80.41</td>
</tr>
<tr>
<td>LSSVM-H</td>
<td>80.41</td>
</tr>
<tr>
<td>Latent CRF</td>
<td>85.09</td>
</tr>
</tbody>
</table>
Translating Videos to Natural Language Using Deep Recurrent Neural Networks

S. Venugopalan, H. Xu, M. Rohrbach, J. Donahue, R. Mooney, K. Saenko

A monkey pulls a dog’s tail and is chased by the dog.

Motivation: Given a video clip generate a sentence to describe the event.

Translating Videos to Natural Language using Deep RNNs (NAACL-HLT 15)
S. Venugopalan, H. Xu, M. Rohrbach, J. Donahue, R. Mooney, K. Saenko

Credit: S. Venugopalan

Decoding a visual vector to a sequence using Recurrent Neural Networks (RNNs)

Key Insight:
Generate feature representation of the video and “decode” it to a sentence

Translating Videos to Natural Language using Deep RNNs (NAACL-HLT 15)
S. Venugopalan, H. Xu, M. Rohrbach, J. Donahue, R. Mooney, K. Saenko


Credit: S. Venugopalan
CNN+LSTM network for video description.

Train on image-caption datasets and tune on videos.

Translating Videos to Natural Language using Deep RNNs (NAACL-HLT 15)
S. Venugopalan, H. Xu, M. Rohrbach, J. Donahue, R. Mooney, K. Saenko


Credit: S. Venugopalan
Invited speaker: Lillian Lee, Cornell University

- Title: “Big data pragmatics!”, or “Putting the ACL in Computational Social Science”, or, if you think these title alternatives could turn people on, turn people off, or otherwise have an effect, this talk might be for you.

Some Topics:
- Framing (Baumer et al., 2015; Guerini et al., 2015)
- Social structure and dynamics (Krishnan and Eisenstein 2015).
Research question: how does *language* reflect and define *social relationships*?

Setting: network structure and *address terms* are observed.

Address terms like “Dude” and “Mr.” play an important role in setting the nature of social relationships.

Edge labels are *latent variables*.

Credit: J. Eisenstein
Probabilistic model

\[ P(y, x \mid G; \Theta, \beta, \eta) = P(x \mid y; \Theta)P(y \mid G; \beta, \eta) \]

The **likelihood** explains how social factors determine linguistic features.

This term factors over the dyads in the social network.

The **prior** explains which network label configurations are likely.

This term factors over the **triads** in the network.

All parameters are estimated using **variational inference**. Noise-contrastive estimation lets us avoid summing over all label configurations.

Credit: J. Eisenstein
# Movie script evaluation

<table>
<thead>
<tr>
<th>V-cluster</th>
<th>T-cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>sir</td>
<td>firstName</td>
</tr>
<tr>
<td>mr+lastName</td>
<td>man</td>
</tr>
<tr>
<td>mr+firstName</td>
<td>baby</td>
</tr>
<tr>
<td>mr</td>
<td>honey</td>
</tr>
<tr>
<td>miss+lastName</td>
<td>darling</td>
</tr>
<tr>
<td>son</td>
<td>sweetheart</td>
</tr>
<tr>
<td>mister+firstName</td>
<td>buddy</td>
</tr>
<tr>
<td>mrs</td>
<td>sweetie</td>
</tr>
</tbody>
</table>

1. Address terms naturally cluster by formality

2. The model makes reasonable inferences about individual films.

3. We can estimate the stability of signed triads.

<table>
<thead>
<tr>
<th>Triad</th>
<th>Stability</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>+3.73</td>
</tr>
<tr>
<td>t</td>
<td>-6.48</td>
</tr>
<tr>
<td>t</td>
<td>-1.05</td>
</tr>
<tr>
<td>v</td>
<td>+1.23</td>
</tr>
</tbody>
</table>
Trends in Methodology

Distributional Semantic Models

- *Distributional Hypothesis*: words that are used in the same context tend to have similar meanings

- Vector space representation for words – word vectors

- Semantic compositionality (Fyshe et al. 2015), verbs with multiple senses (Greenberg et al. 2015), incorporating visual features (Lazaridou et al. 2015), incorporating semantic resources (Faruqui et al., 2015)
Retrofitting Word Vectors to Semantic Lexicons, NAACL 2015

Credit: M. Faruqi
Retrofitting

\[ \Psi(Q) = \sum_{i=1}^{n} \left[ \alpha_i \| q_i - r_i \|^2 + \sum_{(i,j) \in E} \beta_{ij} \| q_i - q_j \|^2 \right] \]

Iterative Updates:

\[ q_i = \frac{\sum_{j:(i,j) \in E} \beta_{ij} q_j + \alpha_i r_i}{\sum_{j:(i,j) \in E} \beta_{ij} + \alpha_i} \]

Credit: M. Faruqi
Sentiment Analysis

Credit: M. Faruqi
Unsupervised Morphology Induction using Word Embeddings

Radu Soricut, Franz J. Och*
NAACL 2015

*now at Human Longevity Inc.
Unsupervised Morphology Induction

Q: What do we want?

A: We want morphology-based transformations that can accurately analyze words (even ones unseen at training time)

prefix: un: ε
unabated unable unabridged...
unaware unbalance unbeaten...
undoing undone undoubted...
untrusted untrustedworthy...

\[ \text{rank}(\text{unaware} \rightarrow \text{aware}) = 0 \]
\[ \text{rank}(\text{undone} \rightarrow \text{done}) = 129 \]

\[ \text{rank}(\text{undone} + \uparrow\text{un-} \rightarrow \text{done}) = 4 \]

Query against embedding space: morphology shifts meaning consistently

Credit: R. Soricut
Unsupervised Morphology Induction: Algorithm

Output: labeled, weighted, acyclic, directed graph $D^\text{V}_{\text{Morph}}$

- words are nodes, morphological mappings are weighted edges

Credit: R. Soricu
Unsupervised Morphology Induction: Evaluation

Evaluation on similarity datasets (RG-DE, RW-EN)

| Language | Train Set | | Tokens | | V | \(|G_{Morph}^{V}|\) | \(|D_{Morph}^{V}|\) |
|----------|-----------|----------------|--------|----------------|--------|
| EN       | Wiki-EN   | 1.1b           | 1.2m   | 780k           | 75,823 |
| DE       | WMT-DE    | 1.2b           | 2.9m   | 3.7m           | 169,017|
| EN       | News-EN   | 120b           | 1.0m   | 2.9m           | 98,268 |
| DE       | News-DE   | 20b            | 1.8m   | 6.7m           | 351,980|

RW-EN Testset

<table>
<thead>
<tr>
<th></th>
<th>[Unembedded]</th>
<th>Spearman (\rho)</th>
</tr>
</thead>
<tbody>
<tr>
<td>System</td>
<td>Wiki-EN</td>
<td>News-EN</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wikipedia EN</td>
</tr>
<tr>
<td>SkipGram</td>
<td>80</td>
<td>177</td>
</tr>
<tr>
<td>SkipGram+Morph</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

RG-DE Testset

<table>
<thead>
<tr>
<th></th>
<th>[Unembedded]</th>
<th>Spearman (\rho)</th>
</tr>
</thead>
<tbody>
<tr>
<td>System</td>
<td>WMT-DE</td>
<td>News-DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wikipedia DE</td>
</tr>
<tr>
<td>SkipGram</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>SkipGram+Morph</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Credit: R. Soricut
Trends in Methodology

Deep Neural Network

– Document categorization (Johnson and Zhang 2015)
– Word sense disambiguation (Taghipour and Ng 2015)
– Paraphrase identification (Yin and Schutze 2015; Zhao et al. 2015)
– Response generation (Sordoni et al. 2015)
– Machine Translation (Xing et al. 2015; de Gispert et al. 2015)
– Video-to-text generation (Venugopalan et al. 2015)
Summary

• NLP has grown into one of the most exciting and diverse community

• Conference website:

• ACL Anthology: an open-access repository
  – http://aclweb.org/anthology/
  – Journal, conference, and workshop papers since 1965