Statistical NLP
Fall 2016

Lecture 13:
Sentiment Analysis, Summarization
Class Summary and Outlook

Slav Petrov – Google

Deep Compositional Model
- Induce representation and activation along parse tree backbone (note: have label at every node)

<table>
<thead>
<tr>
<th>Model</th>
<th>Fine-grained All</th>
<th>Fine-grained Root</th>
<th>Positive/Negative All</th>
<th>Positive/Negative Root</th>
</tr>
</thead>
<tbody>
<tr>
<td>NB</td>
<td>67.2</td>
<td>41.0</td>
<td>82.6</td>
<td>81.8</td>
</tr>
<tr>
<td>SVM</td>
<td>64.3</td>
<td>40.7</td>
<td>84.6</td>
<td>79.4</td>
</tr>
<tr>
<td>BiNB</td>
<td>71.0</td>
<td>41.9</td>
<td>82.7</td>
<td>83.1</td>
</tr>
<tr>
<td>VecAvg</td>
<td>73.3</td>
<td>32.7</td>
<td>85.1</td>
<td>80.1</td>
</tr>
<tr>
<td>RNN</td>
<td>79.0</td>
<td>43.2</td>
<td>86.1</td>
<td>82.4</td>
</tr>
<tr>
<td>MV-RNN</td>
<td>78.7</td>
<td>44.4</td>
<td>86.8</td>
<td>82.9</td>
</tr>
<tr>
<td>RNTN</td>
<td><strong>80.7</strong></td>
<td><strong>45.7</strong></td>
<td><strong>87.6</strong></td>
<td><strong>85.4</strong></td>
</tr>
</tbody>
</table>

Sentiment Treebank
- Crowdsourced fine-grained sentiment annotations

Subjectivity vs. Sentiment
- Subjective sentences express private states, i.e. internal mental or emotional states
  - speculations, beliefs, emotions, evaluations, goals, opinions, judgments, …
    - Jill said, “I hate Bill.”
    - Jack thought he won the race.
    - John liked the book.
- Sentiment expressions are a type of subjective expression
  - expressions of positive and negative emotions, judgments, evaluations, …
Why Sentiment Analysis?

- Many webpages with opinion out there
- Opinions matter for decision making:
  - 60% of US residents have done online research on a product at least once, and 15% do so on a typical day.
  - 73%-87% of US readers of online reviews of services reported that the reviews had a significant influence on their purchase.
  - 30% of online US residents have posted an online comment or review, as have 18% of online US senior citizens.
- But, 58% of US internet users report that online information was missing, impossible to find, confusing, and/or overwhelming.
- Also (more later): Review spam is becoming a serious problem

Sentiment Classification

- Treat Document as bag-of-words. Extract features:
  - Unigrams, Bigrams
  - TF/IDF
  - Feed into your favorite classifier
- Limitations?
- Apply to IMDB reviews

Pang et al., EMNLP 2002
Turney, ACL 2002
Turney&Littman, TOIS 2003

Sentiment Classification Results

- A lot harder than text classification!

<table>
<thead>
<tr>
<th>Features</th>
<th># of features</th>
<th>frequency or presence?</th>
<th>NB</th>
<th>ME</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) unigrams</td>
<td>16165</td>
<td>freq.</td>
<td>78.7</td>
<td>N/A</td>
<td>72.8</td>
</tr>
<tr>
<td>(2) unigrams</td>
<td></td>
<td>press.</td>
<td>81.0</td>
<td>80.4</td>
<td>82.9</td>
</tr>
<tr>
<td>(3) unigrams+bigrams</td>
<td>32330</td>
<td>press.</td>
<td>80.6</td>
<td>80.8</td>
<td>82.7</td>
</tr>
<tr>
<td>(4) bigrams</td>
<td>16165</td>
<td>press.</td>
<td>77.3</td>
<td>77.4</td>
<td>77.1</td>
</tr>
<tr>
<td>(5) unigrams+POS</td>
<td>16605</td>
<td>press.</td>
<td>81.5</td>
<td>80.4</td>
<td>81.9</td>
</tr>
<tr>
<td>(6) adjectives</td>
<td>2633</td>
<td>press.</td>
<td>77.0</td>
<td>77.7</td>
<td>75.1</td>
</tr>
<tr>
<td>(7) top 2633 unigrams</td>
<td>2633</td>
<td>press.</td>
<td>80.3</td>
<td>81.0</td>
<td>81.4</td>
</tr>
<tr>
<td>(8) unigrams+position</td>
<td>22430</td>
<td>press.</td>
<td>81.0</td>
<td>80.1</td>
<td>81.6</td>
</tr>
</tbody>
</table>

- Can get 60% with a list of 5-10 positive and negative sentiment words

Human vs. Machine Learned

<table>
<thead>
<tr>
<th>Proposed word lists</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human 1 Positive: dazzling, brilliant, phenomenal, excellent, fantastic Negative: suck, terrible, awful, unwatchable, hideous</td>
<td>58%</td>
</tr>
<tr>
<td>Human 2 Positive: gripping, mesmerizing, riveting, spectacular, cool, awesome, thrilling, badass, excellent, moving, exciting Negative: bad, cliched, sucks, boring, stupid, slow</td>
<td>84%</td>
</tr>
<tr>
<td>Statistics-based Positive: love, wonderful, best, great, superb, beautiful, still Negative: bad, worst, stupid, waste, boring, ?, !</td>
<td>69%</td>
</tr>
</tbody>
</table>
What’s the Problem?

- Sentiment of words is context dependent:
  - This laptop is a great deal.
  - A great deal of media attention surrounded the release of the new laptop model.
  - If you think this laptop is a great deal, I’ve got a nice bridge for you to buy.

- …The protagonist tries to protect her good name…
- “Read the book.”
- Unpredictable

Structural Correspondence Learning

1. Choose most salient sentiment words that appear in both domains: Pivot
2. Project other features based on correlations

SCL Results

Polarity Lexicon

[Blitzer, Dredze & Pereira]
Positive Phrases

<table>
<thead>
<tr>
<th>POSITIVE PHRASES</th>
<th>Typical</th>
<th>Multword expressions</th>
<th>Spelling variations</th>
</tr>
</thead>
<tbody>
<tr>
<td>cute</td>
<td>once in a life time</td>
<td>loveable</td>
<td></td>
</tr>
<tr>
<td>fabuloust</td>
<td>state - of - the - art</td>
<td>nicee</td>
<td></td>
</tr>
<tr>
<td>cuddly</td>
<td>fail - safe operation</td>
<td>nice</td>
<td></td>
</tr>
<tr>
<td>plucky</td>
<td>just what the doctor ordered</td>
<td>cooool</td>
<td></td>
</tr>
<tr>
<td>revnishing</td>
<td>out of this world</td>
<td>cooool</td>
<td></td>
</tr>
<tr>
<td>spankly</td>
<td>top of the line</td>
<td>kool</td>
<td></td>
</tr>
<tr>
<td>enchanting</td>
<td>melt in your mouth</td>
<td>kewl</td>
<td></td>
</tr>
<tr>
<td>precious</td>
<td>snug as a bug</td>
<td>cozy</td>
<td></td>
</tr>
<tr>
<td>charming</td>
<td>out of the box</td>
<td>cosy</td>
<td></td>
</tr>
<tr>
<td>stupendous</td>
<td>more good than bad</td>
<td>sikk</td>
<td></td>
</tr>
</tbody>
</table>

Negative Phrases

<table>
<thead>
<tr>
<th>NEGATIVE PHRASES</th>
<th>Typical</th>
<th>Multword expressions</th>
<th>Vulgarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>dirty</td>
<td>run of the mill</td>
<td>fucking stupid</td>
<td></td>
</tr>
<tr>
<td>repulsive</td>
<td>out of touch</td>
<td>fucked up</td>
<td></td>
</tr>
<tr>
<td>crappy</td>
<td>over the hill</td>
<td>complete bullshit</td>
<td></td>
</tr>
<tr>
<td>sucky</td>
<td>flash in the pan</td>
<td>shitty</td>
<td></td>
</tr>
<tr>
<td>subpar</td>
<td>bumps in the road</td>
<td>half assed</td>
<td></td>
</tr>
<tr>
<td>horridous</td>
<td>foaming at the mouth</td>
<td>jackass</td>
<td></td>
</tr>
<tr>
<td>miserable</td>
<td>dime a dozen</td>
<td>piece of shit</td>
<td></td>
</tr>
<tr>
<td>lousy</td>
<td>pie - in - the - sky</td>
<td>son of a pitch</td>
<td></td>
</tr>
<tr>
<td>abysmal</td>
<td>sick to my stomach</td>
<td>sonofabitch</td>
<td></td>
</tr>
<tr>
<td>wretched</td>
<td>pain in my ass</td>
<td>sonuvabitch</td>
<td></td>
</tr>
</tbody>
</table>

Fine-Grained Sentiment

Sometimes reviews contain multiple aspects and we want to model those:

Nikos’ Fine Dining

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Rating</th>
<th>Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food</td>
<td>4/5</td>
<td>“Best fish in the city”, “Excellent appetizers”</td>
</tr>
<tr>
<td>Decor</td>
<td>3/5</td>
<td>“Cozy with an old world feel”, “Too dark”</td>
</tr>
<tr>
<td>Service</td>
<td>1/5</td>
<td>“Our waitress was rude”, “Awful service”</td>
</tr>
<tr>
<td>Value</td>
<td>5/5</td>
<td>“Good Greek food for the $”, “Great price!”</td>
</tr>
</tbody>
</table>

Fine-Grained Extraction

Food: 5; Decor: 5; Service: 5; Value: 5
The chicken was great. On top of that our service was excellent and the price was right. Can’t wait to go back!
Food: 2; Decor: 1; Service: 3; Value: 2
We went there for our anniversary. My soup was cold and expensive plus it felt like they hadn’t painted since 1980.
Food: 3; Decor: 5; Service: 4; Value: 5
The food is only mediocre, but well worth the cost. Wait staff was friendly. Lot’s of fan decorations.

Food: “The chicken was great”, “My soup was cold”, “The food is only mediocre”
Decor: “It felt like they hadn’t painted since 1980”, “Lots of fan decorations”
Service: “Service was excellent”, “Wait staff was friendly”
Value: “The price was right”, “My soup was cold and expensive”, “Well worth the cost”
Detour: Graphical Models

- Nodes are random variables
- Edges denote possible dependence
- Observed variables are shaded
- Plates denote replicated structure

Example Topics

- human
dna
gene
sequence
sequencing
map
information
genetics
project
sequences

- evolution
- species
- origin
- phylogenetic
- control
- diversity
- group
- new
- common

- disease
- bacteria
- system
- strains
- infectious
- malaria
- parasites
- united
- tuberculosis

- computer
- models
- data
- system
- networks
- parasites
- software
- new

- graph
- models
- diseases
- data
- systems
- networks

Topic Models

Latent Dirichlet Allocation

[Blei & Jordan '03]

Each piece of the structure is a random variable.

Seeking Life's Bare (Genetic) Necessities

COLD SPRING HARBOR, NEW YORK—How near are we to passing 800 genes to the job—but that millions of copies of the 800 genes are all that are more than not.

Although the numbers billion, much precisely, those produced

* Genome Mapping and Sequencing, Cold Spring Harbor, New York, May 8 to 10, 1996

Stripping dawn, Computer analysis yields an estimate of the minimum modern and ancient genomes.
Multi-Grained LDA

MG-LDA Example Topics

<table>
<thead>
<tr>
<th>MG-LDA</th>
<th>Key topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local</td>
<td>sound quality headphones volume base memorable good settings ear rock excellent</td>
</tr>
<tr>
<td>Global</td>
<td>sound quality headphones volume base memorable good settings ear rock excellent</td>
</tr>
<tr>
<td>[all topics]</td>
<td>sound quality headphones volume base memorable good settings ear rock excellent</td>
</tr>
</tbody>
</table>

MG-LDA Results

MG-LDA Results

MG-LDA Results

MG-LDA Results

Figure 4: (a) Aspect service. (b) Aspect location. (c) Aspect rooms.

Sentiment Analysis

<table>
<thead>
<tr>
<th>Product Name</th>
<th>Rating</th>
<th>Reviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>GoodFellas [Blu-ray]</td>
<td>3 stars</td>
<td>1,197 reviews</td>
</tr>
</tbody>
</table>

What people are saying:
- "Great acting from an excellent cast."
- "A modern classic by a great director.
- "Great performances and characters."
- "Great, great, great!"
- "Very close to the Southern Italian theme"
Deceptive Reviews

- Since reviews are so important for businesses, there is an incentive to produce fake reviews.
- Businesses can (and do):
  - bribe customers to leave positive reviews
  - force employees to write reviews
  - crowdsource (=pay for) positive reviews
- How can we identify spam reviews?

Detecting Fake Reviews Results

- Reasonable in-domain accuracy can be achieved with simple models and features
- Models don’t generalize well across domains

Negation Detection

What are the things (nouns) and properties of them (adjectives)?

How are they connected?

Still lots of work left to do

Table 4: Classifier performance in cross-domain adaptation.

<table>
<thead>
<tr>
<th>Model</th>
<th>Features</th>
<th>Domain</th>
<th>A</th>
<th>P</th>
<th>R</th>
<th>F1</th>
<th>Domain</th>
<th>A</th>
<th>P</th>
<th>R</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>Unigram</td>
<td>Restaurant</td>
<td>0.785</td>
<td>0.813</td>
<td>0.742</td>
<td>0.778</td>
<td>Doctor</td>
<td>0.550</td>
<td>0.537</td>
<td>0.725</td>
<td>0.617</td>
</tr>
<tr>
<td></td>
<td>LIWC</td>
<td>Restaurant</td>
<td>0.745</td>
<td>0.692</td>
<td>0.840</td>
<td>0.759</td>
<td>Doctor</td>
<td>0.521</td>
<td>0.512</td>
<td>0.965</td>
<td>0.669</td>
</tr>
<tr>
<td></td>
<td>POS</td>
<td>Restaurant</td>
<td>0.735</td>
<td>0.697</td>
<td>0.815</td>
<td>0.751</td>
<td>Doctor</td>
<td>0.540</td>
<td>0.521</td>
<td>0.975</td>
<td>0.679</td>
</tr>
<tr>
<td>SAGE</td>
<td>Unigram</td>
<td>Restaurant</td>
<td>0.770</td>
<td>0.793</td>
<td>0.750</td>
<td>0.784</td>
<td>Doctor</td>
<td>0.520</td>
<td>0.547</td>
<td>0.705</td>
<td>0.616</td>
</tr>
<tr>
<td></td>
<td>LIWC</td>
<td>Restaurant</td>
<td>0.742</td>
<td>0.728</td>
<td>0.749</td>
<td>0.738</td>
<td>Doctor</td>
<td>0.647</td>
<td>0.650</td>
<td>0.608</td>
<td>0.628</td>
</tr>
<tr>
<td></td>
<td>POS</td>
<td>Restaurant</td>
<td>0.746</td>
<td>0.732</td>
<td>0.687</td>
<td>0.701</td>
<td>Doctor</td>
<td>0.834</td>
<td>0.823</td>
<td>0.682</td>
<td>0.651</td>
</tr>
</tbody>
</table>

[Li et al. '04]
Summarization

‘Genres’ of Summary?

- Indicative vs. informative
  - used for quick categorization vs. content processing.
- Extract vs. abstract
  - lists fragments of text vs. re-phrases content coherently.
- Generic vs. query-oriented
  - provides author’s view vs. reflects user’s interest.
- Background vs. just-the-news
  - assumes reader’s prior knowledge is poor vs. up-to-date.
- Single-document vs. multi-document source
  - based on one text vs. fuses together many texts.

Summly

He Has Millions and a New Job at Yahoo. Soon, He’ll Be 18.

CrunchBase

Simple, intuitive and elegant. Summly redefines news for the mobile world with algorithmically generated summaries from thousands of sources. Innovative gestures, animations and great summaries make reading the news fun, easy to use, easy to scan, easy to read, share and construe.

Nick D’Aloisio launched Summly in December 2011 as a tech summarization prototype that garnered significant interest worldwide. With backing from Horizon Ventures, and help from many NLP and AI experts around the world, Nick and the Summly team have been able to further develop summarization technology, the first result being the new Summly mobile news app.

Sentence Extraction

\[ D \rightarrow S \]
Selection

mid-'90s

• Maximum Marginal Relevance
  [Carbonell and Goldstein, 1998]

Greedy search over sentences

Maximize similarity to the query
Minimize redundancy

\[ MMR = \arg\max_{D_i \in \mathcal{P}} \left( \lambda \text{Sim}(D_i, Q) - (1 - \lambda) \max_{D_j \in \mathcal{P}} \text{Sim}(D_i, D_j) \right) \]

present

Selection

mid-'90s

• Maximum Marginal Relevance
  • Graph algorithms [Mihalcea 05++]

present

Selection

mid-'90s

• Maximum Marginal Relevance
  • Graph algorithms

Nodes are sentences

Selection

mid-'90s

• Maximum Marginal Relevance
  • Graph algorithms

Nodes are sentences

Edges are similarities
Selection

- Maximum Marginal Relevance
- Graph algorithms

Nodes are sentences
Edges are similarities
Stationary distribution represents node centrality

Selection

- Maximum Marginal Relevance
- Graph algorithms
- Word distribution models

Input document distribution
Summary distribution

Selection

- Maximum Marginal Relevance
- Graph algorithms
- Word distribution models

SumBasic [Nenkova and Vanderwende, 2005]

Value(w_i) = P_D(w_i)
Value(s_i) = sum of its word values
Choose s_i with largest value
Adjust P_D(w)
Repeat until length constraint

Selection

- Maximum Marginal Relevance
- Graph algorithms
- Word distribution models
- Regression models

F(x)
frequency is just one of many features

Input document distribution
Summary distribution

- Obama 0.017
- speech 0.024
- health 0.009
- Montana 0.002
Selection

- Maximum Marginal Relevance
- Graph algorithms
- Word distribution models
- Regression models
- Topic model-based
  [Haghighi and Vanderwende, 2009]

Summarization Criterion

\[ S^* = \min_{S: \text{words}(S) \leq L} KL(P_C \| P_S) \]

\[ P_C(.) \]
- Barack Obama: 0.15
- Serve America Act: 0.13
- signed: 0.12

\[ P_S(.) \]
- Barack Obama: 0.18
- Serve America Act: 0.16
- signed: 0.10

Raw Count Content Model

President Barack Obama received the Serve America Act after congress’ vote. The ailing senator was instrumental in its passage.

[Haghighi & Vanderwende, NAACL '09]
President Barack Obama received the Serve America Act after congress’ vote...

The bill is named after Massachusetts Senator Ted Kennedy who was present at its signing. The ailing senator was instrumental...

The legislation would greatly expand the ranks of Ameri-Corps, which was created by President Bill Clinton in 1993...
Selection

- Maximum Marginal Relevance
- Graph algorithms
- Word distribution models
- Regression models
- Topic models
- Globally optimal search

[McDonald, 2007]

[McDonald, 2007]

Optimal search using MMR

Integer Linear Program

Maximize: \( \sum_i \text{Rel}_i s_i - \sum_j \text{Rel}_j s_{ij} \)

Subject to: \( \sum_j l_j x_j \leq L \)

\( s_{ij} \leq s_i \quad s_{ij} \leq s_j \quad \forall i, j \)

\( s_i + s_j - s_{ij} \leq 1 \quad \forall i, j \)

\( s_i \in \{0, 1\} \quad \forall i \)

\( s_{ij} \in \{0, 1\} \quad \forall i, j \)

Concept - Value

[Gillick and Favre, 2008]

The healthcare bill is a major test for the Obama administration.

<table>
<thead>
<tr>
<th>concept</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>obama</td>
<td>3</td>
</tr>
</tbody>
</table>

Universal health care is a divisive issue.

<table>
<thead>
<tr>
<th>concept</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>health</td>
<td>2</td>
</tr>
</tbody>
</table>

President Obama remained calm.

<table>
<thead>
<tr>
<th>concept</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obama</td>
<td>3</td>
</tr>
</tbody>
</table>

Obama addressed the House on Tuesday.

<table>
<thead>
<tr>
<th>concept</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>health</td>
<td>2</td>
</tr>
</tbody>
</table>

Universal health care is a divisive issue.

<table>
<thead>
<tr>
<th>concept</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>health</td>
<td>2</td>
</tr>
</tbody>
</table>

President Obama remained calm.

<table>
<thead>
<tr>
<th>concept</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obama</td>
<td>3</td>
</tr>
</tbody>
</table>

Obama addressed the House on Tuesday.
**Universal Health Care**

The health care bill is a major test for the Obama administration.

**President Obama remained calm.**

The health care bill is a major test for the Obama administration.

**Obama addressed the House on Tuesday.**

**Importance of Search**

The health care bill is a major test for the Obama administration.

Universal health care is a divisive issue.

President Obama remained calm.

Obama addressed the House on Tuesday.

**Results**

This ILP is tractable for reasonable problems.

**Concept - Value**

- The health care bill is a major test for the Obama administration.
- Universal health care is a divisive issue.
- President Obama remained calm.
- Obama addressed the House on Tuesday.

**Integer Linear Program**

Integer Linear Program for the maximum coverage model

\[
\text{Maximize: } \sum_i w_i c_i \rightarrow \text{total concept value}
\]

\[
\text{Subject to: } \sum_j s_{ij} \leq L \rightarrow \text{summary length limit}
\]

\[
\begin{align*}
& s_{ij} \cdot \text{Occ}_{ij} \leq c_i, \quad \forall i, j \\
& \sum_j s_{ij} \cdot \text{Occ}_{ij} \geq c_i, \quad \forall i \\
& c_i \in \{0, 1\}, \quad \forall i \\
& s_{ij} \in \{0, 1\}, \quad \forall j
\end{align*}
\]

this ILP is tractable for reasonable problems.
**Sentence Position**

First sentences are unique

How to include sentence position?

**Incorporating Sentence Position**

Only allow first sentences in the summary

Up-weight concepts appearing in first sentences

Identify more sentences that look like first sentences

surprisingly strong baseline

included in TAC 2009 system

first sentence classifier is not reliable enough yet

How to include sentence position?

**Sentence Ordering**

Some interesting work on sentence ordering

[Barzilay et. al., 1997; 2002]

But choosing independent sentences is easier

• First sentences usually stand alone well
• Sentences without unresolved pronouns
• Classifier trained on OntoNotes: <10% error rate

Baseline ordering module (chronological) is not obviously worse than anything fancier

**Results [G & F, 2009]**

Overall Quality

- Rating scale: 1-10
- Humans in [8.3, 9.3]

Pyramid

- Rating scale: 0-1
- Humans in [0.62, 0.77]

Linguistic Quality

- Rating scale: 1-10
- Humans in [8.5, 9.3]

ROUGE-2

- Rating scale: 0-1
- Humans in [0.11, 0.15]
Error Breakdown

[Gillick and Favre, 2008]

<table>
<thead>
<tr>
<th>Culprits in low-scoring summaries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad ordering</td>
</tr>
<tr>
<td>Broken quote</td>
</tr>
<tr>
<td>Unclear reference</td>
</tr>
<tr>
<td>No Verb</td>
</tr>
<tr>
<td>Sentence segmentation failed</td>
</tr>
<tr>
<td>Redundancy</td>
</tr>
<tr>
<td>Relative date</td>
</tr>
<tr>
<td>0.0000   0.15   0.30   0.45   0.60</td>
</tr>
</tbody>
</table>

Selection

- Maximum Marginal Relevance
- Graph algorithms
- Word distribution models
- Regression models
- Topic models
- Globally optimal search
- LSMTs (with Attention)

[Rush et al. 2015]

Sentence Compression

Social activist Medha Patkar on Monday extended her “complete” support to Arvind Kejriwal-led Aam Aadmi Party in Maharashtra.

Medha Patkar extended her support to Aam Aadmi Party.

<table>
<thead>
<tr>
<th>Model</th>
<th>FI</th>
<th>Acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIRA</td>
<td>0.75</td>
<td>0.21</td>
</tr>
<tr>
<td>LSTM</td>
<td>0.80</td>
<td>0.30</td>
</tr>
<tr>
<td>LSTM+PAR</td>
<td>0.81</td>
<td>0.31</td>
</tr>
<tr>
<td>LSTM+PAR+PRES</td>
<td>0.82</td>
<td>0.34</td>
</tr>
</tbody>
</table>

Fillipova et al. 2015

Beyond Extraction?

Sentence extraction is limiting

... and boring!

But abstractive summaries are much harder to generate...

The Adventures of Huckleberry Finn
Mark Twain

in 25 words?
But what about XOR?

- Minsky and the AI Winter

<table>
<thead>
<tr>
<th>or</th>
<th>and</th>
<th>xor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>-</td>
<td>1</td>
</tr>
</tbody>
</table>

**Yep** | **Yep** | **Nope**
The Deep Learning Tsunami

- Chris Manning: “Deep Learning waves have lapped at the shores of computational linguistics for several years now, but 2015 seems like the year when the full force of the tsunami hit the major Natural Language Processing (NLP) conferences.”

- Neil Lawrence: “NLP is kind of like a rabbit in the headlights of the Deep Learning machine, waiting to be flattened.”
Some of these ideas have been around…

**Deep Learning Recipe**

Deep Learning = Lots of training data + Parallel Computation + Scalable, smart algorithms

![Image of Deep Learning “Computer Vision Recipe”]

- Big Data: ImageNet
- Deep Convolutional Neural Network
- Backprop on GPU
- Learned Weights

**NIPS vs ACL**

**Grammar as a foreign language**

Authors: Oriol Vinyals, Lukasz Kaiser, T sunny Koo, Slav Petrov, Ilya Sutskever, Geoffrey Hinton

Publication date: 2015

Conference: Advances in Neural Information Processing Systems

Pages: 2775-2781

Description: Abstract: Syntactic constituency parsing is a fundamental problem in natural language processing which has been the subject of intense research and engineering for decades. As a result, the most accurate parsers are domain-specific, complex, and inefficient. In this paper we show that the domain-agnostic attention-enhanced sequence-to-sequence model achieves state-of-the-art results on the most widely used syntactic constituency parsing dataset, when trained on a large synthetic corpora that was annotated using existing...

**Structured Training for Neural Network Transition-Based Parsing**

Authors: David Weise, Chris Alberti, Michael Collins, Slav Petrov

Publication date: 2015

Description: Abstract: We present structured perception training for neural network transition-based dependency parsing. We learn the neural network representation using a gold corpus augmented by a large number of automatically parsed sentences. Given this fixed network representation, we learn a final layer using the structured perception with beam-search algorithm. This allows us to model dependencies spanning multiple words, and gain a 20% unlabeled F1 increase over the exact beam search method. We also show that 100% unlabeled attachment accuracy, which to our knowledge is the best accuracy on Stanford...

**What’s different this time?**

And why we have found the solution:
- More Data
- OpenSource
- More Compute
- More People

And why there is more work to do:
- Not Enough Data for Most Tasks
- Multi-Task and Transfer Learning don’t work yet
- Not always the right type of Compute
My Parting Thoughts

Classic NLP Pipeline of the Future? Shared representations between tasks, multi-task learning with low-level tasks reduces amount of training data needed.

Final Project

- Talk to me if you are unsure
- Presentations:
  - Everybody should talk:
    - individual projects: 4+1 min
    - 2 people projects: 6+2 min
    - 3 people projects: 9+2 min
    - 4 people projects: 12+3 min
  - Explain:
    - (1) the problem
    - (2) a baseline
    - (3) your approach
      (OK not to have final results)
  - I will bring pizza

Stay in touch

slav@petrovi.de