What’s _new_?

different
next
happening

Up to 50% relative improvement in generating context-sensitive lexical substitutes

Our _new_ results are based on a _new_ way to model contexts

latest
recent
current

Try it yourselves with our software toolkit
Empty Category Detection With Joint Context-Label Embedding

To the MT Community:
Bothered by word alignments?
How can a word aligned with a “NULL”?

Using Empty Category Detection to make it easier.

Before               After
吃了吗？ ー＞       你吃了晚饭了吗？
食べた？ ー＞あなたは晩ご飯をたべた？
Did you have supper?

Xun Wang, Katsuhito Sudoh, Masaaki Nagata
NTT, Japan
a Novel Approach to a Semantically-Aware Representation of Items

Transforming vectors to a semantic space of concepts

- Inherent disambiguation & dimensionality reduction

Lexical specificity
- consistent improvements over tf-idf

SoA performance on five datasets in two tasks:

- word similarity
- sense clustering

José Camacho-Collados, Mohammad Taher Pilehvar, and Roberto Navigli
Multi-target Machine Translation with Multi-synchronous Context-free Grammars
Graham Neubig, Kevin Duh, Philip Arthur (NAIST)

Can you pick the correct translation?

 arabic: اتفاقية كيوتو

 Kyoto Treaty (a)
 Kyoto sanctions (b)
 Kyoto Agreement (c)

At the poster, find out more about:

• **Multi-synchronous CFGs**, which can generate strings in multiple languages

• **How to use Spanish to improve English-French translation**
Using Zero-Resource Spoken Term Discovery for Ranked Retrieval
Jerome White, Douglas W. Oard, Jiaul Paik, Rashmi Sankepally, Aren Jansen

What do you mean there’s no ASR for my language?

How am I supposed to retrieve anything?

But these aren’t even words!

Is this possible?

Zero resource term discovery!

Session 1B. Poster 178.
Sign constraints on feature weights improve a joint model of word segmentation and phonology
Mark Johnson, Joe Pater, Robert Staubs and Emmanuel Dupoux

- Task: jointly segment utterances into words and learn phonological alternations
  
  ju wɑn tu si θə bʊk
  “you want to see the book”

- Our model combines:
  » a MaxEnt model of word segmentation (Berg-Kirkpatrick et al), and
  » a Harmony Theory/MaxEnt model of phonology (Smolensky, Goldwater et al)

- Harmony Theory says that certain feature weights should have specific signs

- Our results are significantly better when we constrain the weights as Harmony theory suggests
A supervised SVM-based *Word Sense Disambiguation* system is improved by incorporating *word embeddings*.

Moreover, we use a feed-forward neural network to enrich pre-trained embeddings with *task-specific discriminative* information.
Model Invertibility Regularization: Sequence Alignment With and Without Parallel Data

$T_1 \times T_2 \approx I$

Alignment models trained in reverse directions should be inverses of each other.

Works even without parallel data!

Reviewer 1
“*Ingenious* and *useful* trick”

Reviewer 2
“*Easier to understand*” than other methods

Reviewer 3
“The idea is very *intuitive*”

We can also use *transitivity*! For more information see Poster #122
Continuous Space Representations of Linguistic Typology and their Application to Phylogenetic Inference

Yugo Murawaki, Kyushu University

How related to other languages?
How did the ancestor look like?
How have features changed?

Word order: SOV
Affixation: Strongly suffixing
Adposition: Postpositions

Mundari (Munda)

Word order: SVO
Affixation: Little affixation
Adposition: Prepositions

Khmer (Mon-Khmer)

Proto-Austroasian
Interpreting Compound Noun Phrases Using Web Search Queries

Marius Pașca - Google Inc. - Mountain View, California

- Goal: Collect normalized **lexical interpretations** of the semantic roles played by **modifiers** relative to **heads** within noun phrases.
  - *victorinox knives* $\rightarrow$ *knives made by victorinox*
  - *kitchen knives* $\rightarrow$ *knives used in the kitchen* $\rightarrow$ *knives*
  - *metal knives* $\rightarrow$ *knives made of metal*

- Status: From queries, collect candidate **lexical interpretations** of at most one **modifier** relative to **heads** within noun phrases.

<table>
<thead>
<tr>
<th>Query</th>
<th>Interpretations</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009 movies</td>
<td>movies of 2009, movies released in 2009, movies in 2009, movies from 2009, movies for 2009, ...</td>
</tr>
<tr>
<td>ayurvedic medicinal plants</td>
<td>medicinal plants in ayurveda, medicinal plants used in ayurveda, medicinal plants of ayurveda, medicinal plants used in ayurvedic medicines, medicinal plants used in ayurvedic products, ...</td>
</tr>
<tr>
<td>tsunami charities</td>
<td>charities for tsunami, charities for the tsunami, charities that are helping the tsunami, charities involved in tsunami, charities that have donated to the tsunami, ...</td>
</tr>
</tbody>
</table>
Diamonds in the Rough: Event Extraction from Imperfect Microblog Data

Ander Intxaurrondo, Eneko Agirre, Oier Lopez de Lacalle, Mihai Surdeanu

University of the Basque Country & University of Arizona

- Twitter to extract information about earthquakes.
- Distant supervision for event extraction.
- New evaluation metric.
- Two simple strategies that address inaccurate and ambiguous information.
- Public dataset containing a knowledge base of earthquakes and corresponding tweets.
A Nonparanormal Approach to Predicting and Generating Popular Meme Descriptions

William Y. Wang (CMU)
Unsupervised Dependency Parsing: Let’s Use Supervised Parsers

Phong Le & Willem Zuidema — ILLC, University of Amsterdam — {p.le,zuidema}@uva.nl

Supervised parsing develops fast with
- rich models
- word embedding integration
- efficient learning algorithms

Why is it difficult to reuse supervised parsers?
- supervised parsers are designed for being trained on manual annotated data
- using EM to train supervised parsing models is very expensive

Our proposal: Iterated Reranking (a variant of self-training)
A Transition-based Algorithm for AMR Parsing

Chuan Wang, Nianwen Xue, Sameer Pradhan

There are many similarities between AMR and dependency structure.

(a) Dependency tree

(b) AMR graph

AMR parsing as transitions from dependency tree to AMR Graph.
Why do we struggle with feature-rich MT?

Convex Geometry: A fresh perspective on the training of linear models (MERT, PRO, MIRA, etc.)

Optimisers are constrained at low feature dimensions

Polynomial-time multidimensional MERT
Unsupervised Multi-Domain Adaptation with Feature Embeddings
Yi Yang and Jacob Eisenstein (Georgia Tech)

• Improve over pivot based approaches — leverage the full feature co-occurrence matrix

• Beyond single source and target domains setting — unsupervised multi-domain adaptation
We have **Big DATA** to train SMT systems.

- Thanks to Europarl, UN, Common Crawl, ...

Wait ...

- Data come from **very different domains**.
- **How does this affect the word alignment accuracy?**

Bigger data ≠ producing better alignment quality
Cost Optimization For Crowdsourcing Translation
Mingkun Gao, Wei Xu, Chris Callison-Burch
(University of Pennsylvania)
An In-depth Analysis of the Effect of Text Normalization in Social Media
Tyler Baldwin and Yunyao Li

Past
- TTS Normalization
- Parser Normalization
- NER Normalization
- MT Normalization

Present
@someGuy idk y u think that tho

Future

… ...
Multitask Learning for Adaptive Quality Estimation of Automatically Transcribed Utterances

José G. C. de Souza, Hamed Zamani, Matteo Negri, Marco Turchi, Daniele Falavigna

PROBLEM:
Predict ASR output quality without references nor access to confidence information...

![Diagram showing the process of input signal going through an ASR system to produce a transcription with quality 0.03 or "BAD".]

...by learning from data coming from multiple, heterogeneous domains

![Examples of different domains: Medical, IT, Legal.]

SOLUTION:
Come to poster session 1B poster 460!
A Dynamic Programming Algorithm for Tree Trimming-based Text Summarization

Masaaki Nishino (NTT)

A tree trimming problem

Input

A dependency tree (sentence/document)

Output

An optimal subtree within a length limit \( L \) (summary)

\( O(NL \log N) \) exact algorithms!
We don’t seem to be doing well in the foreign beverage market. However, due to a mistranslation of our slogan we’ve become the leading international provider of embalming fluid.

Sentiment after Translation:
A Case-Study on Arabic Social Media Posts

Mohammad Salameh
University of Alberta

Saif M. Mohammad
National Research Council Canada

Svetlana Kiritchenko
National Research Council Canada

Dialectal Arabic:
مش غلط يكون اقلها مرتب (Positive)

Manual Translation:
it is not wrong to at least be neat (Neutral)

Automatic Translation:
it's not a mistake be less of their salary (Negative)

Questions:
• What is the accuracy of a state-of-the-art Arabic sentiment analysis (SA) system?
• What accuracy is expected when the posts are translated (manually or automatically) and run through an English SA system?
  • What is the impact of bad translation?
• How difficult is it for humans to determine sentiment of automatically translated text?
Corpus-based discovery of semantic intensity scales

They are synonyms!

Really?

Chomsky

Adjective scale

Word2vec + clustering

Derive clusters

Adverb scales

Applications

Sentiment Analysis

BONUS!

THE OHIO STATE UNIVERSITY
Dialogues with forms of the pronoun "he".

I heard you all went to the movies.

Yes, ___ went yesterday.

What is he doing?

___ eating a sandwich.

I think I'm going to head out.

No, ___ should stay!

___ went yesterday.

___ eating a sandwich.

___ should stay!
One fundamental difficulty of coreference resolution is to resolve instances that require background knowledge.

- The *bird* perched on the *limb* and [it] *bent*.  
- The *bird* perched on the *limb* and [it] *sang*.  
- *Bill* was robbed by *John*, so the officer *arrested* [him].  
- *Bill* was robbed by *John*, so the officer *helped* [him].

Inject Knowledge into Coreference

- We build a state-of-the-art coreference system that at the same time also handles hard instances at close to 90% precision.

Demo: http://cogcomp.cs.illinois.edu/page/demo_view/Coref
A bomb in a Hebrew University cafeteria killed five Americans and four Israelis.

A bombing at Hebrew University in Jerusalem killed nine people, including five Americans.

Ryan has 72 marbles and 17 blocks. If he shares the marbles among 9 friends, how many marbles does each friend get?

Each friend gets $72/9 = 8$ marbles. The number of blocks is irrelevant.
Unsupervised Declarative Knowledge Induction for Constraint-Based Learning of Information Structure in Scientific Documents

Yufan Guo, Roi Reichart, Anna Korhonen

• **Declarative Knowledge**: $0.9 \leq p(\text{CON} | \text{suggest}) \leq 1$

• **Automatic Induction**: $? \leq p(?) | (?) \leq ?$