Natural Language Processing: Interpretation, Reasoning and Machine Learning

Roberto Basili
(Università di Roma, Tor Vergata)
dblp: http://dblp.uni-trier.de/pers/hd/b/Basili:Roberto.html
Google scholar: https://scholar.google.com/citations?user=U1A22fYAAAAJ&hl=it&oi=sra

Politecnico di Milano, 4-5 Maggio 2017
NLP & ML: Selected Topics

• Natural Language Processing - linguistic background: Tasks, Models and Methods

• The Machine Learning View on NLP
  • Natural Language as an inductive process
  • Statistical Language Processing: from bayesian modeling to NLU

• Statistical NLP: tasks
  • Statistical Parsing
  • Semantic Role Labeling

• Applications of Statistical NLP
  • Web-based Opinion Mining Systems, Market Watch & Brand Reputation Management
  • Human Robotic Voice Interaction
Semantics, Open Data and Natural Language

Web contents, characterized by rich multimedia information, are mostly opaque from a semantic standpoint.
Who is Hu Jintao?

Chinese President Hu Jintao (R) shakes hands with Honorary Chairman of the Chinese Kuomintang (KMT) Lien Chan, in Honolulu, Hawaii, the U.S., Nov. 11, 2011.

(Xinhua/Huang Jingwen)

HONOLULU, United States, Nov. 11 (Xinhua) -- Hu Jintao, general secretary of the Central Committee of the Communist Party of China and Chinese President, meets with Lien Chan, honorary chairman of the Chinese Nationalist Party (KMT), here on Tuesday.

(Xinhua/Huang Jingwen)
Search results for "Hu Jintao" on Google.
Content Semantics and Natural Language

• Human languages are the main carrier of the information involved in processes such as retrieval, publication and exchange of knowledge as it is associated to the open Web contents.

• Words and NL syntactic structures express concepts, activities, events, abstractions and conceptual relations we usually share through data.

• “Language is parasitic to knowledge representation languages but the viceversa is not true” (Wilks, 2001)

• From Learning to Read to Knowledge Distillation as a (integrated pool of) Semantic interpretation Task(s).
Semantics, Natural Language & Learning: tasks

• From Learning to Read to Knowledge Distillation as a (integrated pool of)

Semantic Interpretation Task(s)

• **Information Extraction**
  • Entity Recognition and Classification
  • Relation Extraction
  • Semantic Role Labeling (Shallow Semantic Parsing)

• **Estimation of Text Similarity**
  • Structured Text Similarity/Textual Entailment Recognition
  • Sense disambiguation

• **Semantic Search, Question Classification and Answer Ranking**

• **Knowledge Acquisition**, e.g. ontology learning

• **Social Network Analysis, Opinion Mining**
Overview

- Artificial Intelligence, Natural Language & Speech
  - Information, Representation, (re)current challenges, success(and unsuccess)ful stories

- Natural Language Processing: linguistic background

- break

- Natural Language Processing: Tasks, Models and Methods

- The Role of Machine Learning Technologies
  - Lexicon Acquisition : Automatic Development of Dictionaries, Semantic Lexicons and Ontologies
  - Statistical Language Processing: Semantic Role Labeling

- break

- Natural Language Processing : Results & Applications
  - Semantic Document Management
  - Web-based Opinion Mining Systems, Market Watch & Brand Reputation Management
  - Human Robotic Voice Interaction
Which Knowledge?

• HAL 9000, da “2001: A Space Odyssey”

• Dave: Open the pod bay doors, Hal.

• HAL: I’m sorry Dave, I’m afraid I can’t do that.
What’s HAL knowledge?

• Recognition & Synthesis of spoken language
  • Dictionaries (spelling)
  • Phonetics (how to produce/recognize sound)

• Understanding
  • Lexical Knowledge
    • What do the words mean?
    • How they combine (`pod bay door’)
  • Knowledge about the syntagmatic structure of sentences
    • I’m I do, Sorry that afraid Dave I’m can’t
What’s HAL knowledge?

- Dialogue & pragmatics
  - “open the door” is a request (and not a declaration or a search query)
  - Replying is a type of action that imply kindness (even if a planning to kill is in progress ...)
  - It is useful to behave cooperatively (I’m afraid, I can’t...)
  - What about `that’ in `I can’t do that’?
Language Processing as a (semantic) interpretation process

• Processing a text corresponds to the understanding of a number of aspects related to its meaning
  • Thematic Domain (e.g. science/economics/sport)
  • Operational Objectives (e.g. e-mail spam)
  • Involved Entities, such as people or locations
  • Potential events described (e.g. facts told by news)
  • Communicative Objectives (e.g. dialogue, orders/declarations/planning)

• Outcome: an explicit representation of the text meaning ...

• able to trigger different inferences
  (e.g. IR relevance, planning, knowledge updates, ....)
Some Reflections

- Understanding linguistic information requires specific knowledge about:
  - The natural language itself (e.g. grammar)
  - The world (e.g. bay door, Dave or opening)
  - How language make reference to the world

- NLP applications deals with texts by exploiting the specific context:
  - Application purposes, e.g. document search
  - The domain and the operational context of an application
  - The distinction between language producer (speaker/writer) and consumer (hearer/reader)
Major Challenges

• Linguistic Accuracy in approximating the human-level of performance
• Robustness (errors/noise/incompleteness)
• Scale
  • Coverage of the phenomena (Lexicons/Grammars)
• Expressivity
  • Dictionaries, Lexicons and Thesauri
  • World Models and types of inference
• Flexibility
  • Adequate performance across linguistic variability (e.g. producer vs. consumer)
• Naturalness
NLP: the standard processing chain

Lexical Analysis

Tokens+
features

Syntactic Analysis

Parse
tree

Semantic Analysis

Logic
Form

Pragmatics/ Application

Interpretation/Plan

Lexicons

Grammar(s)

World
Model

Task
Model
Grammatical Analysis

Mortgage approvals fall back to January level
Mortgage approvals fell sharply in June, lending yet more weight to the theory of a dip in the UK housing market as the Nationwide index showed UK house prices starting to fall in July.

- Halifax index shows 0.6% fall in house prices
- In depth: UK house prices
- House prices rise at slowing rate

Default retirement age to be scrapped
Move delights pressure groups but dismays business organisations, which warn that the measure is being introduced too quickly.

Global Insight: Cameron needs to be more subtle
David Cameron has led the largest official delegation to India since its independence from Britain 63 years ago. By doing so, he has tested Britain’s place in the world, and how far it has traveled since 1947.

Gifts lose lustre for overseas investors
Flight from eurozone risk to UK government bonds is moderating.
Syntax and Semantics in textual data

Compositionality

• The meaning of a complex expression is solely determined by the meanings of its constituent expressions and the rules used to combine them.

• "I will consider a language to be a set (finite or infinite) of sentences, each finite in length and constructed out of a finite set of elements. All natural languages are languages in this sense. Similarly, the set of "sentences" of some formalized system of mathematics can be considered a language" Chomsky 1957
Syntax

• In linguistics, **syntax** is the study of the rules that govern the structure of sentences, and which determine their relative grammaticality.

• Such rules govern a number of language phenomena as systems for phonology, morphology, syntax as well as discourse.
Parse Trees

- The representation of the parsing result is a structure that expresses:
  - The **order of constituent elements** in the sentence
  - The **grammatical type** of constituents
  - The **hierarchical organization** of constituents

- The structures able to express these properties are the derivation trees also called **parse trees**
Grammars and Trees

“The firm holds some stakes”

- $V_n = \{S, NP, VP, Det, N\}$, Axiom: $S$

- Productions: \{$S \rightarrow NP \ VP$, $VP \rightarrow V \ NP$, $NP \rightarrow Det \ N$\}

- Derivation:
  - $S \rightarrow NP \ VP \rightarrow Det \ N \ VP \rightarrow The \ N \ VP \rightarrow The \ firm \ VP \rightarrow The \ firm \ V \ NP \rightarrow The \ firm \ holds \ NP \rightarrow The \ firm \ holds \ Det \ N \rightarrow The \ firm \ holds \ some \ N \rightarrow The \ firm \ holds \ some \ stakes$
Constituency-based Parsing (marked Head)

S
  /   
 NP   VP
  /    
 N    NP
   /    VP
  N    PP
   /    
 N    N
    /    P
     N    NP
      /    
     N    June
      /    
      approvals fell sharply in
Dependency Parsing

Basic Dependencies:

President Xi Jinping of China, on his first state visit to the United States, showed off his familiarity with American history and pop culture on Tuesday night.
Grammatical Relation Centered Tree (GRCT)
Challenges for Parsing

- Huge complexity as for the ambiguity in the morphosyntactic descriptions of words
  - E.g. La vecchia porta la sbarra,

  Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo

- Interdependency with semantic information
  - Most ambiguity cannot be solved only at the grammatical level
  - Lexical Semantic information is crucial as grammatical structures are constrained by word senses
    - Operating in a market vs. Operating a patient

Bison from Buffalo, New York who are intimidated by other bison in their community also happen to intimidate other bison in their community
FT (July, 29): Mortgage approvals fell sharply in June.
President Xi Jinping of China, on his first state visit to the United States, showed off his familiarity with American history and pop culture on Tuesday night.
Semantics

• What is the meaning of the sentence

   John saw Kim?

• Desirable Properties:
  • It should be derivable as a function of the individual constituents, i.e. the meanings of constituents such as Kim, John and see
  • Independent from syntactic phenomena, e.g. Kim was seen by John is a paraphrasis
  • It must be directly used to trigger some inferences:
    • Who was seen by John? Kim!
    • John saw Kim. He started running to her.
A Truth conditional semantics

John saw Kim
NL Interpretation as compositional processing through *lambda* expressions
Three Linguistic Perspectives on Meaning

- **Lexical Semantics**
  - The meanings of individual words

- **Formal Semantics** (or Compositional Semantics or Sentential Semantics)
  - How those meanings combine to make meanings for individual sentences or utterances

- **Discourse or Pragmatics**
  - How those meanings combine with each other and with other facts about various kinds of context to make meanings for a text or discourse
  - Dialog or Conversation is often lumped together with Discourse
Lexical Semantic: Relationships between word meanings

- Homonymy
- Polysemy
- Synonymy
- Antonymy
- Hypernymy
- Hyponymy
- Meronymy
Homonymy

- **Homonymy:**
  - Lexemes that share a form
    - Phonological, orthographic or both
  - But have unrelated, distinct meanings
  - Clear example:
    - Bat (wooden stick-like thing) vs Bat (flying scary mammal thing)
    - Or bank (financial institution) versus bank (riverside)
  - Can be also homophones, homographs, or both:
    - Homophones:
      - Write and right
      - Piece and peace
Polysemy

- The **bank** is constructed from red brick. I withdrew the money from the **bank**
- Are those the same sense?
- Or consider the following WSJ example
  - While some banks furnish sperm only to married women, others are less restrictive
- Which sense of bank is this?
  - Is it distinct from (homonymous with) the river bank sense?
  - How about the savings bank sense?
Synonyms

- Word that have the same meaning in some or all contexts.
  - filbert / hazelnut
  - couch / sofa
  - big / large
  - automobile / car
  - vomit / throw up
  - Water / $H_2O$

- Two lexemes are synonyms if they can be successfully substituted for each other in all situations.
  - If so they have the same **propositional meaning**
Synonyms

• But there are few (or no) examples of perfect synonymy.
  • Why should that be?
  • Even if many aspects of meaning are identical still may not preserve the acceptability based on notions of politeness, slang, register, genre, etc.

• Example:
  • Water and $\text{H}_2\text{O}$
  • I would not say

  *I like fresh $\text{H}_2\text{O}$ after the tennis*
Some terminology

• Lemmas and wordforms
  • A **lexeme** is an abstract pairing of meaning and form
  • A **lemma** or **citation form** is the grammatical form that is used to represent a **lexeme**.
    • **Carpet** is the lemma for *carpets*, **Dormir** is the lemma for *duermes*.
  • Specific surface forms *carpets*, *sung*, *duermes* are called **wordforms**

• The lemma **bank** has two **senses**:
  • Instead, a **bank** can hold the investments in a custodial account in the client’s name
  • But as agriculture burgeons on the east **bank**, the river will shrink even more.

• A **sense** is a discrete representation of one aspect of the meaning of a word
Synonymy is a relation between senses rather than words

- Consider the words *big* and *large*
- Are they synonyms?
  - How *big* is that plane?
  - Would I be flying on a *large* or small plane?
- How about here:
  - Miss Nelson, for instance, became a kind of *big* sister to Benjamin.
  - Miss Nelson, for instance, became a kind of *large* sister to Benjamin.
- Why?
  - *big* has a sense that means being older, or grown up
  - *large* lacks this sense
Antonyms

- Senses that are opposites with respect to one feature of their meaning
- Otherwise, they are very similar!
  - dark / light
  - short / long
  - hot / cold
  - up / down
  - in / out
- More formally: antonyms can
  - define a binary opposition or opposite ends of a scale (long/short, fast/slow)
  - Be **reversives**: rise/fall, up/down
II. WordNet

- A hierarchically organized **lexical** database
- On-line thesaurus + aspects of a dictionary
  - Versions for other languages are under development

<table>
<thead>
<tr>
<th>Category</th>
<th>Unique Forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun</td>
<td>117,097</td>
</tr>
<tr>
<td>Verb</td>
<td>11,488</td>
</tr>
<tr>
<td>Adjective</td>
<td>22,141</td>
</tr>
<tr>
<td>Adverb</td>
<td>4,601</td>
</tr>
</tbody>
</table>
WordNet Search - 3.1
- WordNet home page - Glossary - Help

Word to search for: meaning  Search WordNet

Display Options: (Select option to change)  Change

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
Display options for sense: (gloss) "an example sentence"

Noun

- **S:** (n) meaning, significance, signification, import (the message that is intended or expressed or signified) "what is the meaning of this sentence"; "the significance of a red traffic light"; "the signification of Chinese characters"; "the import of his announcement was ambiguous"
  - direct hyponym / full hyponym
  - direct hypernym / inherited hypernym / sister term
  - derivationally related form

- **S:** (n) meaning, substance (the idea that is intended) "What is the meaning of this proverb?"
Logic, Predicates and Arguments

- The syntax-semantic *mapping*

Different Annotation schemes: PropBank vs. FrameNet
Linking syntax to semantics: see later slides on Semantic Role Labeling

Mario arrestò il baro per truffa

Mario arrestò il baro per truffa
ML in NLP ... a prologue

• The syntax-semantic mapping

• Different semantic theories
  (e.g. PropBank vs. FrameNet)
Police arrested the man for shoplifting
### A tabular vision

<table>
<thead>
<tr>
<th>Word</th>
<th>Predicate</th>
<th>Semantic Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Police</td>
<td>-</td>
<td>Authority</td>
</tr>
<tr>
<td>arrested</td>
<td>Target</td>
<td>Arrest</td>
</tr>
<tr>
<td>the</td>
<td>-</td>
<td>SUSPECT</td>
</tr>
<tr>
<td>man</td>
<td>-</td>
<td>SUSPECT</td>
</tr>
<tr>
<td>for</td>
<td>-</td>
<td>OFFENSE</td>
</tr>
<tr>
<td>Shoplifting</td>
<td>-</td>
<td>OFFENSE</td>
</tr>
</tbody>
</table>

![Diagram showing the semantic roles and structural components of the sentence](image-url)
Using Framenet/PropBank

SRL Pipeline

1. Syntactic Parse
   - S
   - NP₁
   - VP
   - V
   - PP
   - P
   - in
   - the park

2. Prune Constituents
   - NP₁
   - VP
   - V
   - PP
   - NP₂

3. Argument Identification
   - Argument Classification
     - NP₁ Agent/Patient
     - V Predicate
     - PP Location/Patient
     - NP₁ Agent/Patient
     - V Predicate
     - PP Location/Patient

4. Structural Inference
   - Semantic Roles

5. Argument Classification
   - NP₁ Yes
   - VP No
   - V given
   - PP Yes
   - NP₂ No

6. Candidates

NLP: linguistic levels

- Speech
- Text
- Phonetics
- Orthography
- Phonology

"shallower"

- Morphology
- Syntax
- Semantics

"deeper"

- Pragmatics
- Discourse
Applications: Target Semantic Phenomena

- **Entities.** Entities cited in texts (people, locations, organizations, date, numerical or monetary expressions)

- **Relations.** Relationships / Associations among entities

- **Facts.** Facts and Events

- **Topics.** Discussion topics / Context / Domain
NLP & ML: Selected Topics

• Natural Language Processing - linguistic background: Tasks, Models and Methods

• The Machine Learning View on NLP
  • Natural Language as an inductive process
  • Statistical Language Processing: from bayesian modeling to NLU

• Statistical NLP: tasks
  • Statistical Parsing
  • Semantic Role Labeling

• Applications of Statistical NLP
  • Web-based Opinion Mining Systems, Market Watch & Brand Reputation Management
  • Human Robotic Voice Interaction
Language as a system of rules

... comincia qui la mia disperazione di scrittore. Ogni linguaggio è un alfabeto di simboli il cui uso presuppone un passato che gli interlocutori condividono; come trasmettere agli altri l’infinito Aleph che la mia timorosa memoria a stento abbraccia?

• Meaning is acquired and recognized within the daily practices related to its usage

  • The meaning of a word is to be defined by the rules for its use, not by the feeling that attaches to the words

• Recognizing one meaning consists in the ability of mapping a linguistic expression to an experience (praxis) through mechanisms such as analogy or approximating equivalence functions or through the minimization of the risks of being wrong/inappropriate/obscure

• The interpretation process can be obtained through the induction of one (or more) decision function(s) from experience
The inductive process

Annotazione Fenomeni → Esempi → Osservazione_1

Osservazione_2 → Osservazione_3 → Osservazione_n

Esempi

Modello

Learning Machine
The inductive process
The inductive process

Annotazione

Fenomeni

Testi

Kernel_{Parole}

Kernel_{Sintagmi}

Kernel_{Tree}

Kernel_{FattiNoti}

Riconoscimento

Modello

SVM Learning

Annotazioni
The inductive process
The inductive process

- Annotazione
- Fenomeni
- Testi
- Citazioni
- Riconoscimento
- Modello
- SVM Learning
- $\text{Kernel}_{\text{Parole}}$
- $\text{Kernel}_{\text{Sintagmi}}$
- $\text{Kernel}_{\text{Tree}}$
- $\text{Kernel}_{\text{FattiNoti}}$
The inductive process

Annotazione Fenomeni → Testi → Citazioni → Riconoscimento → Modello → SVM Learning

Kernel_{Parole} → Kernel_{Sintagmi} → Kernel_{Tree} → Kernel_{FattiNoti}
The inductive process
IBM Watson: between Intelligence and Data

- IBM’s Watson

Jeopardy!
HEDGEHOGS ARE COVERED WITH QUILLS OR SPINES, WHICH ARE HOLLOW HAIRS MADE STIFF BY THIS PROTEIN.
Watson: a DeepQA architecture
Semantic Inference in Watson QA

In May 1498 Portugal celebrated the 400th anniversary of this explorer’s arrival in India.

On the 27th of May 1498, Vasco da Gama landed in Kappad Beach.
... Intelligence in Watson
Ready for Jeopardy!
Strongly positive aspects
- Adaptivity of the overall workflow
- Significant exploitation of available data
- Huge volumes of knowledge involved

Criticalities
- The encyclopedic knowledge needed for Jeopardy is quite different in nature from the domain expertise required in many applications
- Wason is based on Factoid Questions strongly rooted on objective facts, that are explicit and non subjective
- Formalizing the input knowledge, as it is done a priori for Watson, is very difficult to achieve in cost-effective manner: sometimes such knowledge is even absent in an enterprise
- For many natural languages the amount of information and resources is not available, so that a purely data-driven approach is not applicable
Machine Learning: the weapons

- Rule and Pattern learning from Data
  - Frequent Pattern Mining (Basket analysis)
- Probabilistic Extensions of Grammars
  - Probabilistic CFGs
  - Stochastic Grammars
- Discriminative learning in neural networks
- SVM: perceptrons
  - Kernel functions in implicit semantic spaces
- Bayesian Models & Graphical Models
NLP & ML: Selected Topics

• **Natural Language Processing - linguistic background: Tasks, Models and Methods**

• **The Machine Learning View on NLP**
  - Natural Language as an inductive process
  - Statistical Language Processing: from bayesian modeling to NLU

• **Statistical NLP: tasks**
  - POS tagging, Statistical Parsing
  - Semantic Role Labeling

• **Applications of Statistical NLP**
  - Web-based Opinion Mining Systems, Market Watch & Brand Reputation Management
  - Human Robotic Voice Interaction
• POS tagging (Church, 1989)
• Probabilistic Context-Free Grammars (Pereira & Schabes, 1991)
• Data Oriented Parsing (Scha, 1990)
• Stochastic Grammars (Abney, 1993)
• Lessicalizzati Modelli (C. Manning, 1995)

Weighted Grammars, tra Sintassi & Statistica

Figure 13.2 Two parse trees for an ambiguous sentence. The transitive parse (a) corresponds to the sentence “I bought a book at the dinner flight.” The intransitive parse (b) corresponds to the sentence “I bought a book at the dinner.”
**Hidden Markov Models (HMM)**

- States = Categories/Classes
- Observations
- Emissions
- Transitions

\[ p(X_1,\ldots,T, Y_1,\ldots,T) = p(X_1)p(Y_1|X_1) \prod_{t=2}^{T} [p(X_t|X_{t-1})p(Y_t|X_t)] \]

- Applications:
  - Speech Recognition
  - Sequence Labeling (e.g. POS tagging)
The task of POS tagging

POS tagging
Given a sequence of morphemes $w_1, ..., w_n$ with ambiguous syntactic descriptions (i.e., part-of-speech tags) $t_j$, compute the sequence of $n$ POS tags $t_{j_1}, ..., t_{j_n}$ that characterize correspondingly all the words $w_i$.

Examples:
- *Secretariat is expected to race tomorrow*
  - $\Rightarrow$ NNP VBZ VBN TO VB NR
  - $\Rightarrow$ NNP VBZ VBN TO NN NR
**The task of POS tagging**

**An example**

![Diagram](image)

<table>
<thead>
<tr>
<th>Emission probabilities</th>
<th>. the this cat kid eats runs fish fresh little big</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;FF&gt;</td>
<td>1.0</td>
</tr>
<tr>
<td>Dt</td>
<td>0.6 0.4</td>
</tr>
<tr>
<td>N</td>
<td>0.6 0.1 0.3</td>
</tr>
<tr>
<td>V</td>
<td>0.7 0.3</td>
</tr>
<tr>
<td>Adj</td>
<td>0.3 0.3 0.4</td>
</tr>
</tbody>
</table>
How to map a POS tagging problem into a HMM

The above problem

\[(t_1, \ldots, t_n) = \arg\max_{pos_1, \ldots, pos_n} P(pos_1, \ldots, pos_n | w_1, \ldots, w_n)\]

can be also written (Bayes law) as:

\[(t_1, \ldots, t_n) = \arg\max_{pos_1, \ldots, pos_n} P(w_1, \ldots, w_n | pos_1, \ldots, pos_n)P(pos_1, \ldots, pos_n)\]
The final equation is thus:

$$(t_1, \ldots, t_n) = \arg\max_{t_1, \ldots, t_n} P(t_1, \ldots, t_n | w_1, \ldots, w_n) = \arg\max_{t_1, \ldots, t_n} \prod_{i=1}^{n} P(w_i | t_i) P(t_i | t_{i-1})$$
The forward algorithm: estimation

Figure 6.6 The forward trellis for computing the total observation likelihood for the ice-cream events 3 1 3. Hidden states are in circles, observations in squares. White (unfilled) circles indicate illegal transitions. The figure shows the computation of $\alpha_t(j)$ for two states at two time steps. The computation in each cell follows Eq 6.11: $\alpha_t(j) = \sum_{i=1}^{N-1} \alpha_{t-1}(i) a_{ij} b_j(o_t)$. The resulting probability expressed in each cell is Eq 6.10: $\alpha_t(j) = P(o_1, o_2, \ldots, o_t, q_t = j|\lambda)$. 
HMM Decoding: Viterbi Algorithm

Intuition:

- $7.6 \times 10^{-6}$ students/V
- $0.00725$ students/N
- $1.3 \times 10^{-5}$ need/N
- $0.0002$ need/P
- $7.2 \times 10^{-5}$ another/ART
- $2.6 \times 10^{-9}$ break/V
- $4.3 \times 10^{-6}$ break/N
- $0$ another/P
Viterbi decoding

Figure 6.9 The Viterbi trellis for computing the best path through the hidden state space for the ice-cream eating events 3 1 3. Hidden states are in circles, observations in squares. White (unfilled) circles indicate illegal transitions. The figure shows the computation of $v_t(j)$ for two states at two time steps. The computation in each cell follows Eq. 6.10: $v_t(j) = \max_{1 \leq i < N-1} v_{t-1}(i) \alpha_t b_j(\alpha_i)$ The resulting probability expressed in each cell is Eq. 6.16: $v_t(j) = P(q_0, q_1, \ldots, q_{t-1}, o_1, o_2, \ldots, o_t, q_t = j | \lambda)$.
NLP and HMM decoding

• The HMM sequence labeling approach can be applied to a variety of linguistic subtasks:
  • Tokenization
  • MWE recognition
  • POS tagging
  • Named Entity Recognition
  • Predicate Argument Structure Recognition
  • SRL: Shallow Semantic Parsing
Considerable asset for doctors who have to see as many patients as those in a community setting do. Watson can present cancer care teams with reports ranking the most effective therapies and treatment options.

"Identifying the right course of treatment for cancer patients has always been challenging but today’s rapid pace of discovery creates new dilemmas in oncology clinical decision support," notes Frost & Sullivan digital health analyst Nancy Fabozzi in a statement. "Keeping up with the pace of change is difficult enough for oncologists at the most sophisticated medical centers—and can be near impossible for those practicing in community settings with fewer resources. Watson for Oncology is fundamentally reshaping how oncologists derive insights that enable the best possible decision making and highest quality patient care."

IBM has already struck numerous health care-focused collaborations, such as
Multiword Expressions

he was willing to budge a little on

O O O O B b i l

the price which means a lot to me.

O O O B i l l l O

a little; means a lot to me; budge... on

See: “Discriminative lexical semantic segmentation with gaps: running the MWE gamut,” Schneider et al. (2014).
With Commander Chris Ferguson at the helm,

Atlantis touched down at Kennedy Space Center.
Part-of-Speech Tagging

<table>
<thead>
<tr>
<th>Word</th>
<th>POS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ikr</td>
<td>interjection</td>
</tr>
<tr>
<td>smh</td>
<td>acronym</td>
</tr>
<tr>
<td>he</td>
<td>pronoun</td>
</tr>
<tr>
<td>asked</td>
<td>verb</td>
</tr>
<tr>
<td>fir</td>
<td>prep.</td>
</tr>
<tr>
<td>yo</td>
<td>det.</td>
</tr>
<tr>
<td>last</td>
<td>adj.</td>
</tr>
<tr>
<td>name</td>
<td>noun</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Word</th>
<th>POS</th>
</tr>
</thead>
<tbody>
<tr>
<td>so</td>
<td>preposition</td>
</tr>
<tr>
<td>he</td>
<td>proper noun</td>
</tr>
<tr>
<td>can</td>
<td></td>
</tr>
<tr>
<td>add</td>
<td></td>
</tr>
<tr>
<td>u</td>
<td></td>
</tr>
<tr>
<td>on</td>
<td></td>
</tr>
<tr>
<td>fb</td>
<td></td>
</tr>
<tr>
<td>lololol</td>
<td></td>
</tr>
</tbody>
</table>

! | GO | V | PD | AN |
Supersense Tagging

ikr    smh    he    asked    fir    yo    last    name
----------    communication    ----------    cognition

so    he    can    add    u    on    fb    lololol
----------    stative    ----------    group    

NLP & Structured Prediction

- HMM Decoding is an example of a large class of structured prediction task

- **Key elements:**
  - Transform a NLP task into a sequence of classification problem.
  - Transform into a sequence labeling problem and use a variant of the Viterbi algorithm.
  - Design a representation (e.g. features and metrics), a prediction algorithm, and a learning algorithm for your particular problem.
 Discriminative Learning

- Characterizes neural networks since the early Cybernetics (Minsky & Papert, 1956)

- Strongly rooted in the notion of
  - Inner product that in turns characterizes the norms thus the distances in the space
  - Use a vector space in $\mathbb{R}^n$ as a input representation space

- (not so) Recent Achievements
  - Statistical Learning Theory and **Support Vector Machines** (Vapnik, 1987)
  - Deep Learning (Bengio et al., 2001)
    - Deep NNs, CNNs, RNNS
Linear Classification (1)

In the hyperplane equation:

\[ f(\vec{x}) = \vec{x} \cdot \vec{w} + b, \quad \vec{x}, \vec{w} \in \mathbb{R}^n, b \in \mathbb{R} \]

\( \vec{x} \) is the vector describing the targeted input example

\( \vec{w} \) is the gradient of the hyperplane

Classification Inference: \( h(x) = \text{sign}(f(x)) \)
If a specific function called kernel is available such that \( k(x_i,x_j)=\phi(x_i) \cdot \phi(x_j) \), there is no need to project the individual examples through the projection function \( \phi \) (Cristianini et al., 2002).

A structured paradigm is applied such that

- It is trained against more complex structures
- It moves the machine learning focus onto the representation \( (\phi(x_j)) \)

\( k(.,.) \) expresses a similarity (metrics) that can account for linguistic aspects and depend on the lexicon, syntax and/or semantics.

\[
h(x) = \text{sgn}(\vec{w} \cdot \varphi(\vec{x}) + b) = \text{sgn}\left( \sum_{j=1}^{\ell} \alpha_j y_j \varphi(x_j) \cdot \varphi(x) + b \right) = \\
= \text{sgn}\left( \sum_{i=1}^{\ell} \alpha_j y_j k(x_j, x) + b \right)
\]
Examples of Kernels sensitive to syntactic structures

- Given a tree we can see it as the occurrence of a joint event ....
Kernels & Syntactic structures: a collective view of the joint event

The tree can be see it as the joint occurrence of all the following subtrees:
Tree Kernels: the implicit metric space

- The function $\phi$ in a tree kernel define a vector representing ALL subtrees of the input tree $T$. It naturally (i.e. without feature engineering) emphasizes:
  - Lexical information (magazine)
  - Coarse grain grammatical information (POS tags such as VBZ)
  - Syntactic information (frammenti complessi)

- The inner product in the space of all subtrees is proportional to the number of subtrees shared between two sentences

- The learning algorithm (e.g. SVM) will select discriminating examples in (infinite dimensional) space
NLP & ML: Selected Topics

• Natural Language Processing - linguistic background: Tasks, Models and Methods

• The Machine Learning View on NLP
  • Natural Language as an inductive process
  • Statistical Language Processing: from bayesian modeling to NLU

• Statistical NLP: tasks
  • POS tagging, Statistical Parsing
  • Semantic Role Labeling

• Applications of Statistical NLP
  • Web-based Opinion Mining Systems, Market Watch & Brand Reputation Management
  • Human Robotic Voice Interaction
Frame Semantics

• Research in Empirical Semantics suggests that words represent categories of experience (situations)

• A frame is a cognitive structuring device (i.e. a kind of prototype) indexed by words and used to support understanding (Fillmore, 1975)
  • Lexical Units evoke a Frame in a sentence

• Frames are made of elements that express participants to the situation (Frame Elements)

• During communication LU's evoke the frames
<table>
<thead>
<tr>
<th>Frame Elements</th>
<th>Predicates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>KILLER</strong></td>
<td>annihilate.v, annihilation.n, asphyxiate.v, assassin.n, assassinate.v, assassination.n, behead.v, beheading.n, blood-bath.n, butcher.v, butchery.n, carnage.n, crucifixion.n, crucify.v, deadly.a, decapitate.v, decapitation.n, destroy.v, dispatch.v, drown.v, eliminate.v, euthanasia.n, euthanize.v, ...</td>
</tr>
<tr>
<td><strong>VICTIM</strong></td>
<td>John drowned Martha.</td>
</tr>
<tr>
<td><strong>MEANS</strong></td>
<td>The flood exterminated the rats by cutting off access to food.</td>
</tr>
<tr>
<td><strong>CAUSE</strong></td>
<td>The rockslide killed nearly half of the climbers.</td>
</tr>
<tr>
<td><strong>INSTRUMENT</strong></td>
<td>It’s difficult to suicide with only a pocketknife.</td>
</tr>
</tbody>
</table>
Frame Semantics

- Lexical descriptions are expected to define the indexed frame and the frame elements with their realization at the syntactic level:
  - *John bought a computer from Janice for 1000 $*

- Mapping into syntactic arguments
  - the *buyer* is (usually) in the subject position

- Obligatory vs. optional arguments

- Selectional preferences
  - *The seller* and *the buyer* are usually “humans” or “social groups”
An example from Babel (SAG)

- Example
- A law enforcement official told CNN that the FBI was investigating.
  vs
  - CNN was told that the FBI was investigating by a law enforcement official
  vs
  - CNN was told by a law enforcement official that the FBI was investigating
Babel output:
The FrameNet project

• The aims
  • Create a lexical resource by describing a significant portion of English in terms of precise and rich frame semantics

• The output
  • Frame Database: a structured system of Frames and Fes
  • Lexical database: syntactic and semantic descriptions of frame-evoking words (N,V,A)
  • Annotated Corpus: wide coverage examples
Committing crime

Definition:
A perpetrator (generally intentionally) commits a crime, i.e. does something not permitted by the laws of society.

They PERPETRATED a crime by substituting a lie for negotiations.
The suspect had allegedly COMMITTED the crime to gain the attention of a female celebrity.

FEs:

Core:

Crime (Cr) An act, generally intentional, that has been formally forbidden by law.
How can the COMMIT crime against the King of England in a foreign country, if he is not English?

He PERPETRATED treachery against mother nature.

Perpetrator (Pero) The individual that commits a crime.
How can the COMMIT treason against the King of England in a foreign country, if he is not English?

He PERPETRATED a crime against mother nature.

Non-Core:

Frequency (Frec) The frequency with which a crime is committed.
The average serial killer COMMIT a crime every five years.

Instrument (Inst) The instrument used in committing the crime.
Most crimes are COMMITTED with a firearm.
Killing

FEs:

Non-Core:

Beneficiary [ben]

This extra-thematic FE applies to participants that derive a benefit from the occurrence of the event specified by the target predicate.

Circumstances [ ]

Circumstances describe the state of the world (at a particular time and place) which is specifically independent of the event itself and any of its participants.

Semantic Type: Physical_entity
Excludes: Cause

Intrin

Killer [Kill]
Excludes: Cause

The person or sentient entity that causes the death of the Victim.

Exclud

Killer [Kill] Means [ ]
Excludes: Cause

The method or action that the Killer or Cause performs resulting in the death of the Victim.

Semant

Killer [Kill] Exclud State_of_affairs
Excludes: Cause

The flood EXTERMINATED the rats by cutting off access to food.

Semantic Type: Sentient

Victim [ ]

The living entity that dies as a result of the killing.

Non-Core:

Beneficiary [ben]

This extra-thematic FE applies to participants that derive a benefit from the occurrence of the event specified by the target predicate.
The FrameNet Hierarchy

Reciprocity

Commercial_transaction

6 children total

Transfer

Commerce_goods-transfer

Commerce_money-transfer

Ordering Relation:

Parent frame → Child frame

Parent → Child Relation Types:

- Inheritance
- Subframe
- Perspective On
- Using
- Cause Of
- Inchoative Of
- See Also

Ordering Relation:

- Precedes
Framenet - Data

• Methodology of constructing FrameNet
  • Define/discover/describe frames
  • Decide the participants (frame elements)
  • List lexical units that evoke the frame
  • Find example sentences in the BNC and annotate them

• Corpora
  • FrameNet I - British National Corpus only
  • FrameNet II - LDC North American Newswire corpora

• Size
  • >10,000 lexical units, >825 frames, >135,000 sentences

• http://framenet.icsi.berkeley.edu
Using Framenet/PropBank

SRL Pipeline

- Syntactic Parse
  - S
  - NP₁
  - VP
  - V
  - PP
  - P
  - He
  - Walked
  - NP₂
  - the park

- Prune Constituents
  - NP₁
  - VP
  - V
  - PP
  - NP₂

- Argument Identification
  - NP₁
  - VP
  - V
  - PP
  - NP₂
  - V
  - PP
  - NP₂

- Argument Classification
  - NP₁
  - Yes
  - VP
  - No
  - V
  - given
  - PP
  - Yes
  - NP₂
  - No

- Structural Inference
  - NP₁ Agent
  - V
  - Predicate
  - PP
  - Location
  - Semantic Roles

- Candidates
  - NP₁ Agent/Patient
  - V
  - Predicate
  - PP
  - Location/Patient
Application of distributional lexicons for Semantic Role Labeling @ UTV

- An important application of tree-kernel based SVMs is Semantic Role labeling wrt Framenet
- In the UTV system, a cascade of classification steps is applied:
  - Predicate detection
  - Boundary recognition
  - Argument categorization (Local models)
  - Reranking (Joint models)
- Input: a sentence and its parse trees
- Adopted kernel: the combination of lexical (e.g. bow) and tree kernel (that is still a kernel)
Linking syntax to semantics

- Police arrested the man for shoplifting
Using Framenet/PropBank
Semantic Role Labeling via SVM Learning

- Two steps:
  - Boundary Detection
    - One binary classifier applied to the parse tree nodes
  - Argument Type Classification
    - Multi-classification problem, where n binary classifiers are applied, one for each argument class (i.e. frame element)
    - They are combined in a ONE-vs-ALL scheme, i.e. the argument type that is categorized by an SVM with the maximum score is selected
## SRL in FrameNet: Results (2009)

<table>
<thead>
<tr>
<th>Eval Setting</th>
<th>Tree Kernels</th>
<th>Tree Kernels + PK</th>
<th>PK alone</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
<td>F₁</td>
</tr>
<tr>
<td>BD</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BD Proj.</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BD+RC</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BD+RC Proj.</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>TK</th>
<th>TK + PK</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
</tr>
<tr>
<td>BD</td>
<td>.949</td>
<td>.652</td>
</tr>
<tr>
<td>BD Proj.</td>
<td>.919</td>
<td>.631</td>
</tr>
<tr>
<td>BD+RC</td>
<td>.697</td>
<td>.479</td>
</tr>
<tr>
<td>BD+RC Proj.</td>
<td>.672</td>
<td>.462</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>TKL</th>
<th>TKL + PK</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
</tr>
<tr>
<td>BD</td>
<td>.938</td>
<td>.659</td>
</tr>
<tr>
<td>BD Proj.</td>
<td>.906</td>
<td>.636</td>
</tr>
<tr>
<td>BD+RC</td>
<td>.689</td>
<td>.484</td>
</tr>
<tr>
<td>BD+RC Proj.</td>
<td>.663</td>
<td>.466</td>
</tr>
</tbody>
</table>

Table 4.1: Results on FrameNet dataset. The table shows Precision, Recall, and F-measure achieved by the Polynomial Kernel (PK) and two different Tree Kernels (TK and TKL). Also, results for their combinations are shown. All experiments exploit 2% training data for Boundary Detection, and 90% for Role Classification.
Framenet SRL: best results

- Best system [Erk&Pado, 2006]
  - 0.855 Precision, 0.669 Recall
  - 0.751 F1
- Trento (+RTV) system (Coppola, PhD2009)

<table>
<thead>
<tr>
<th>Enhanced PK+TK</th>
<th>P</th>
<th>R</th>
<th>F₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eval Setting</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BD (nodes)</td>
<td>1.0</td>
<td>.732</td>
<td>.847</td>
</tr>
<tr>
<td>BD (words)</td>
<td>.963</td>
<td>.702</td>
<td>.813</td>
</tr>
<tr>
<td>BD+RC (nodes)</td>
<td>.784</td>
<td>.571</td>
<td>.661</td>
</tr>
<tr>
<td>BD+RC (words)</td>
<td>.747</td>
<td>.545</td>
<td>.630</td>
</tr>
</tbody>
</table>

Table 4.2: Results on the FrameNet dataset. Best configuration from Table 4.1, raised to 90% of training data for BD and RC.
**SRL: recent results for Italian (2014)**

### Table 6
Evalita 2011 - Boundary Detection Results

<table>
<thead>
<tr>
<th>System</th>
<th>Argument-Based P</th>
<th>Argument-Based R</th>
<th>Argument-Based F1</th>
<th>Token-Based P</th>
<th>Token-Based R</th>
<th>Token-Based F1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First Run</td>
<td>SVM-SPTK</td>
<td>66.67%</td>
<td>72.50%</td>
<td>69.46%</td>
<td>81.99%</td>
<td>83.15%</td>
</tr>
<tr>
<td></td>
<td>BABEL</td>
<td>50.70%</td>
<td>51.43%</td>
<td>51.06%</td>
<td>68.02%</td>
<td>77.18%</td>
</tr>
<tr>
<td>Second Run</td>
<td>SVM-SPTK</td>
<td>66.67%</td>
<td>72.50%</td>
<td>69.46%</td>
<td>81.99%</td>
<td>83.15%</td>
</tr>
<tr>
<td></td>
<td>BABEL</td>
<td>49.91%</td>
<td>50.36%</td>
<td>50.13%</td>
<td>68.14%</td>
<td>76.69%</td>
</tr>
</tbody>
</table>

### Table 7
Evalita 2011 - Argument Classification Results

<table>
<thead>
<tr>
<th>System</th>
<th>Argument-Based P</th>
<th>Argument-Based R</th>
<th>Argument-Based F1</th>
<th>Token-Based P</th>
<th>Token-Based R</th>
<th>Token-Based F1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First Run</td>
<td>SVM-SPTK</td>
<td>48.44%</td>
<td>52.68%</td>
<td>50.47%</td>
<td>62.58%</td>
<td>63.47%</td>
</tr>
<tr>
<td></td>
<td>BABEL</td>
<td>33.10%</td>
<td>33.57%</td>
<td>33.33%</td>
<td>46.77%</td>
<td>53.06%</td>
</tr>
<tr>
<td>Second Run</td>
<td>SVM-SPTK</td>
<td>51.23%</td>
<td>55.71%</td>
<td>53.38%</td>
<td>69.01%</td>
<td>70.99%</td>
</tr>
<tr>
<td></td>
<td>BABEL</td>
<td>37.52%</td>
<td>37.86%</td>
<td>37.69%</td>
<td>54.63%</td>
<td>61.48%</td>
</tr>
<tr>
<td>Third Run</td>
<td>SVM-SPTK</td>
<td>70.36%</td>
<td>70.36%</td>
<td>70.36%</td>
<td>78.35%</td>
<td>78.35%</td>
</tr>
<tr>
<td></td>
<td>BABEL</td>
<td>66.67%</td>
<td>65.36%</td>
<td>66.01%</td>
<td>77.71%</td>
<td>77.46%</td>
</tr>
</tbody>
</table>
Argument Classification in English
(Croce et al., 2013)

- UTV experimented with a FrameNet SRL classification (gold standard boundaries)
- We used the FrameNet version 1.3: 648 frames are considered
  - Training set: 271,560 arguments (90%)
  - Test set: 30,173 arguments (10%)

\[ \text{Bootleggers}_{\text{CREATOR}}, \text{then copy } [\text{the film}]_{\text{ORIGINAL}} [\text{onto hundreds of VHS tapes}]_{\text{GOAL}} \]

<table>
<thead>
<tr>
<th>Kernel</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRCT</td>
<td>87.60%</td>
</tr>
<tr>
<td>GRCT\textsubscript{LSA}</td>
<td>88.61%</td>
</tr>
<tr>
<td>LCT</td>
<td>87.61%</td>
</tr>
<tr>
<td>LCT\textsubscript{LSA}</td>
<td>88.74%</td>
</tr>
<tr>
<td>GRCT+LCT</td>
<td>87.99%</td>
</tr>
<tr>
<td>GRCT\textsubscript{LSA}+LCT\textsubscript{LSA}</td>
<td>88.91%</td>
</tr>
</tbody>
</table>
Semantics, Natural Language & Learning

- From **Learning to Read** to **Knowledge Distillation** as a (integrated pool of) Semantic interpretation Task(s)
  - **Information Extraction**
    - Entity Recognition and Classification
    - Relation Extraction
    - Semantic Role Labeling (Shallow Semantic Parsing)
  - **Estimation of Text Similarity**
    - Structured Text Similarity/Textual Entailment Recognition
    - Sense disambiguation
  - **Semantic Search, Question Classification and Answer Ranking**
  - **Knowledge Acquisition**, e.g. ontology learning
  - **Social Network Analysis, Opinion Mining**
NLP & ML: Selected Topics

- **Natural Language Processing - linguistic background:** Tasks, Models and Methods
- **The Machine Learning View on NLP**
  - Natural Language as an inductive process
  - Statistical Language Processing: from bayesian modeling to NLU
- **Statistical NLP: tasks**
  - POS tagging, Statistical Parsing
  - Semantic Role Labeling
- **Applications of Statistical NLP**
  - Web search, Question Answering, Opinion Mining
  - Human Robotic Voice Interaction
Web queries: Actions vs. Intents

User Intents and Goals
- plan vacation
- get in shape

Query
- hilton orlando reviews
- sea world location
- how to lose weight

Query Intent
- Informational
- Navigational
- Transactional

Finer-grained Intents
- Advice
- Locate
- Download
- Obtain
- Interact

Actions on Entities
- get address
- add to Netflix queue
- read reviews
- buy

[Broder, 2002]
[Rose and Levinson, 2004]
Learning actions from web usage logs

- Three months of us-en web logs
- Annotate with Freebase entities
- Keep queries with an entity in set of 21 types
- Filter out navigational queries
- Filter out clicked hosts that weren’t clicked at least 100 times

<table>
<thead>
<tr>
<th>(query, host) pairs</th>
<th>entities</th>
<th>types</th>
</tr>
</thead>
<tbody>
<tr>
<td>over 3 months</td>
<td>2,164,579</td>
<td>235,385</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>contexts</th>
<th>hosts</th>
</tr>
</thead>
<tbody>
<tr>
<td>129,088</td>
<td>58,123</td>
</tr>
</tbody>
</table>
Entity disambiguation and linking

- Key requirement is that entities get identified
  - Named entity recognition (e.g., Stanford NER!)
- and disambiguated
  - Entity linking (or sometimes “Wikification”)
    - e.g., Michael Jordan the basketballer or the ML guy
Sergio talked to Ennio about Eli's role in the Ecstasy scene. This sequence on the graveyard was a highlight in Sergio’s trilogy of western films.
• and linked to a canonical reference
  • Freebase, dbPedia, Yago2, (WordNet)
3 approaches to Question Answering: Knowledge-based approaches (Siri)

- Build a semantic representation of the query
  - Times, dates, locations, entities, numeric quantities
- Map from this semantics to query structured data or resources (SQL or SparQL)
  - Geospatial databases
  - Ontologies (Wikipedia infoboxes, dbPedia, WordNet, Yago)
  - Restaurant review sources and reservation services
  - Scientific databases
  - Wolfram Alpha
Text-based (mainly factoid) QA

- **QUESTION PROCESSING**
  - Detect question type, answer type, focus, relations
  - Formulate queries to send to a search engine

- **PASSAGE RETRIEVAL**
  - Retrieve ranked documents
  - Break into suitable passages and rerank

- **ANSWER PROCESSING**
  - Extract candidate answers (as named entities)
  - Rank candidates
    - using evidence from relations in the text and external sources
Hybrid approaches (IBM Watson)

- Build a shallow semantic representation of the query
- Generate answer candidates using IR methods
  - Augmented with ontologies and semi-structured data
- Score each candidate using richer knowledge sources
  - Geospatial databases
  - Temporal reasoning
  - Taxonomical classification
Texts are Knowledge
Task – Answer Sentence Selection

- Given a factoid question, find the sentence that
  - Contains the answer
  - Can sufficiently support the answer

Q: Who won the best actor Oscar in 1973?
S1: Jack Lemmon was awarded the Best Actor Oscar for Save the Tiger (1973).
S2: Academy award winner Kevin Spacey said that Jack Lemmon is remembered as always making time for others.

Scott Wen-tau Yih (ACL 2013) paper
Lemmon was awarded the Best Supporting Actor Oscar in 1956 for *Mister Roberts* (1955) and the Best Actor Oscar for *Save the Tiger* (1973), becoming the first actor to achieve this rare double…

*Source: Jack Lemmon -- Wikipedia*
Word Alignment for Question Answering
TREC QA (1999-2005)

What is the fastest car in the world?

The Jaguar XJ220 is the dearest, fastest and most sought after car on the planet.

[Harabagiu & Moldovan, 2001]

Assume that there is an underlying alignment
• Describes which words in and can be associated
See if the (syntactic/semantic) relations support the answer
Full NLP QA: LCC (Harabagiu/Moldovan)
[below is the architecture of LCC’s QA system circa 2003]
NLP & ML: Selected Topics

• Natural Language Processing - linguistic background: Tasks, Models and Methods

• The Machine Learning View on NLP
  • Natural Language as an inductive process
  • Statistical Language Processing: from bayesian modeling to NLU

• Statistical NLP: tasks
  • POS tagging, Statistical Parsing
  • Semantic Role Labeling

• Applications of Statistical NLP
  • Web-based Opinion Mining Systems, Market Watch & Brand Reputation Management
  • Human Robotic Voice Interaction (see separated slides)
The LU4R project

- See Lu4R slides in a separated file
References


• NLP & ML:

• URLs:
  • SAG, Univ. Roma Tor Vergata: http://sag.art.uniroma2.it/
  • Reveal s.r.l.: http://www.revealsrl.it/