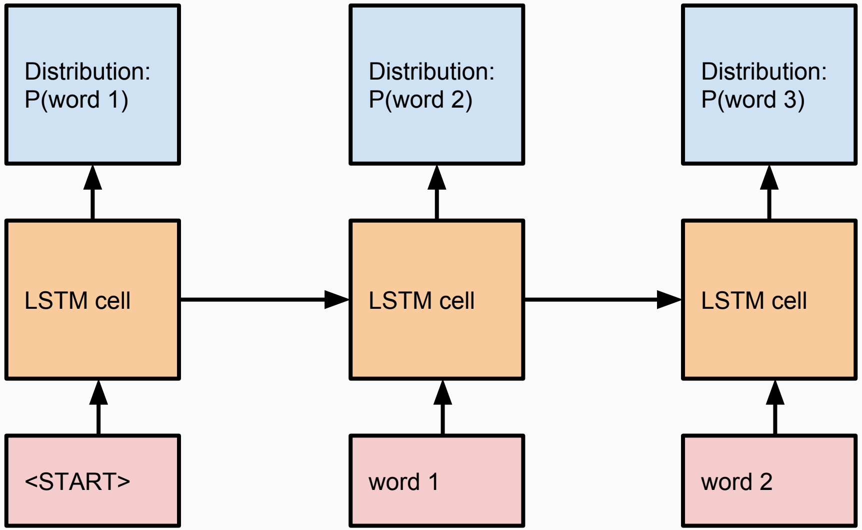
Reverse-order neural language models

Sam Bowman
sbowman@stanford.edu
Ling/NLP PhD Student

Background

Recurrent neural network language models (RNNLMs)

- Neural network models that assign probabilities to sequences of words.
- In some applications, it could be easier to train/test them backwards sentences.



Do RNNLMs produce different distributions when they're trained and tested backwards?

- Simple bigram/trigram/… language models behave exactly the same in both cases.
- But neural networks are all about learning generalizations, and reversing the input forces the model to learn generalizations about left-contexts instead of right-contexts.
- One of these could be easier to do well.
- Which one could depend on specific facts about the syntax of a language.

Project

Theory half:

- Under what circumstances would order make a difference?

Experimental half:

- Find existing RNNLM code and train it on both forward and reversed data from a few different languages.
- Look for patterns in the results.

Training F1 Directly

Gabor Angeli <angeli@stanford.edu>

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Problem: NLP tasks are not class balanced

- Named entity recognition
- Relation extraction
- Sentiment
- …

But, we train against **likelihood** on often **balanced training sets**

Solution? Train against F1 directly

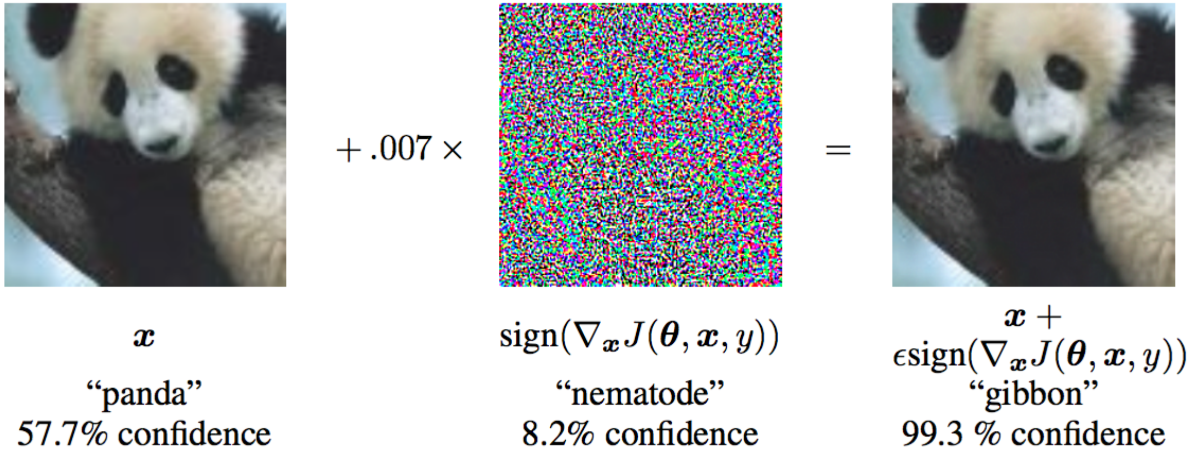
Challenge:

- Nasty objective.
- => develop useful approximations; approximate training algorithms
- **=> Improve real-world performance!**

224N project idea:

Adversarial examples for NLP applications

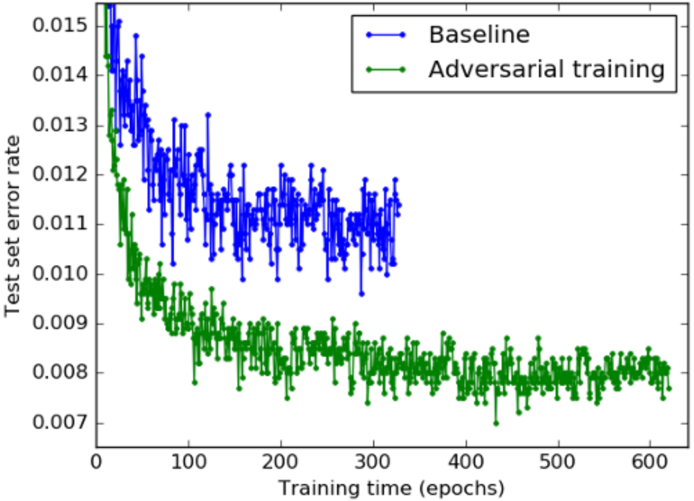
Jon Gauthier
jgauthie@stanford.edu



[Goodfellow et al. 2014]

EXPLAINING AND HARNESSING ADVERSARIAL EXAMPLES

Ian J. Goodfellow, Jonathon Shlens & Christian Szegedy
Google Inc., Mountain View, CA
{goodfellow,shlens,szegedy}@google.com



Say hi!

Jon Gauthier
jgauthie@stanford.edu

Edits on Kickstarter

Rob Voigt (robvoigt@stanford.edu)

- Some cool papers on language in crowdfunding!
  - Althoff et al. 2014, Mitra and Gilbert 2014, Xu et al. 2014, Althoff and Leskovec 2015
- But there's a potential confound!
  - Data (usually) only from the final day!
  - But project creators often make revealing edits!
  - That all sounds a little post-hoc, dawg!

Edits on Kickstarter

Rob Voigt (robvoigt@stanford.edu)

- So I collected a big dataset!
  - Over more than a year: ~30k projects in full, snapshots on each day of the campaign, metadata!
- So we can ask better questions!
  - How well does language help prediction *on day one*?
  - Does editing behavior matter? What kinds of edits tend to precede or follow big jumps in funding?
- Could well be publishable!

Multipersonality Multilinguals

Rob Voigt (robvoigt@stanford.edu)

- Susan Ervin-Tripp has shown that bilinguals may talk quite differently in each language:

1. WHEN MY WISHES CONFLICT WITH MY FAMILY. . .
   (Japanese) it is a time of great unhappiness.
   (English) I do what I want.
2. I WILL PROBABLY BECOME . . .
   (Japanese) a housewife.
   (English) a teacher.
3. REAL FRIENDS SHOULD . . .
   (Japanese) help each other.
   (English) be very frank.

- Can we extend these results computationally?

Multipersonality Multilinguals

Rob Voigt (robvoigt@stanford.edu)

- Specifically, the Hamburg Adult Bilingual LAnguage (HABLA) corpus
  - Bilingual German and Italian/French speakers, conversations with transcripts and audio, 80 hrs
- Look for variation across languages, controlling for speaker:
  - Pitch range, speech rate, lexical choices, etc.
- Could well be publishable!

Smart Lexicon Expansion

Rob Voigt (robvoigt@stanford.edu)

- Lexicons and gazeteers can be really useful!
  - LIWC, HGI, MRC Psycholinguistic Databse, etc.
- But they're fragile:
  - One-hot representation, words are either in or out
- Could we soften or expand these lexicons to make them more general and robust?

Smart Lexicon Expansion

Rob Voigt (robvoigt@stanford.edu)

- Some existing work:
  - ex., graph-based semi-supervised learning (Das and Smith 2012)
- Some possibilities forward:
  - Distributed representations!
  - Replication-based: optimize learning of new words explicitly to maintain existing results
- Could well be publishable!

Lets talk about Food!

"What's good here?"

- Menu-item focused sentiment analysis
- Challenge: mapping review mentions to menu items
  - "Their minced chicken basil saute was yumm!!!"


"กระเพราไข่ดาว"
Gra Prow Kai Dow*** 11
Minced chicken (or pork) basil with bird's eye chili, over rice, with Thai style fried egg.

- Do New Yorkers, Californians and Texans talk about the dishes they love (and hate) differently?
- Data: Yelp reviews and menus (Jurafsky et al. (2014))

Contact: vinod@cs.stanford.edu

Modeling the Social Context in Interactions

- Data: Wikipedia talk pages; Trip Advisor forums?
- Do people exhibit different “dialog behavior” in different "social contexts"?

- Dialog behavior
  - Politeness (Danescu-Niculescu-Mizil et al. (2013))
  - Overt display of power (Prabhakaran et al. (2012))
  - …
- Social contexts
  - Wikipedia discussions: User talk pages vs. Article talk pages vs. Request for Adminship (RfA) discussions
  - Wiki topics: expert area vs. non-expert area
  - Trip Advisor: destination expert for SF interacting in LA forum vs. SF forum
  - …

Contact: vinod@cs.stanford.edu

Let Computers Do Reading Comprehension Test

Danqi Chen

(danqi@cs.stanford.edu)

MCTest (http://research.microsoft.com/en-us/um/redmond/projects/mctest/)

James the Turtle was always getting in trouble. Sometimes he'd reach into the freezer and empty out all the food. Other times he'd sled on the deck and get a splinter. His aunt Jane tried as hard as she could to keep him out of trouble, but he was sneaky and got into lots of trouble behind her back.

One day, James thought he would go into town and see what kind of trouble he could get into. He went to the grocery store and pulled all the pudding off the shelves and ate two jars. Then he walked to the fast food restaurant and ordered 15 bags of fries. He didn't pay, and instead headed home.

His aunt was waiting for him in his room. She told James that she loved him, but he would have to start acting like a well-behaved turtle.After about a month, and after getting into lots of trouble, James finally made up his mind to be a better turtle.

1) What is the name of the trouble making turtle?
   A) Fries
   B) Pudding
   C) James
   D) Jane

2) What did James pull off of the shelves in the grocery store?
   A) pudding
   B) fries
   C) food
   D) splinters

3) Where did James go after he went to the grocery store?
   A) his deck
   B) his freezer
   C) a fast food restaurant
   D) his room

4) What did James do after he ordered the fries?
   A) went to the grocery store
   B) went home without paying
   C) ate them
   D) made up his mind to be a better turtle

multiple-choice reading comprehension test

fictional stories

150-300 words

a 7-year old child is expected to understand

very challenging!

- Lexical matching / coreference / syntax / frame semantics / discourse …
- **Objective #1**: Design your favorite features / feed them into one classifier, figure out what is needed for deeper understanding.
- **Objective #2 (advanced)**: Learn how to train an end-to-end neural network for modeling reading comprehension!

- Lexical matching / coreference / syntax / frame semantics / discourse …
- **Objective #1**: Design your favorite features / feed them into one classifier, figure out what is needed for deeper understanding.
- **Objective #2 (advanced)**: Learn how to train an end-to-end neural network for modeling reading comprehension!

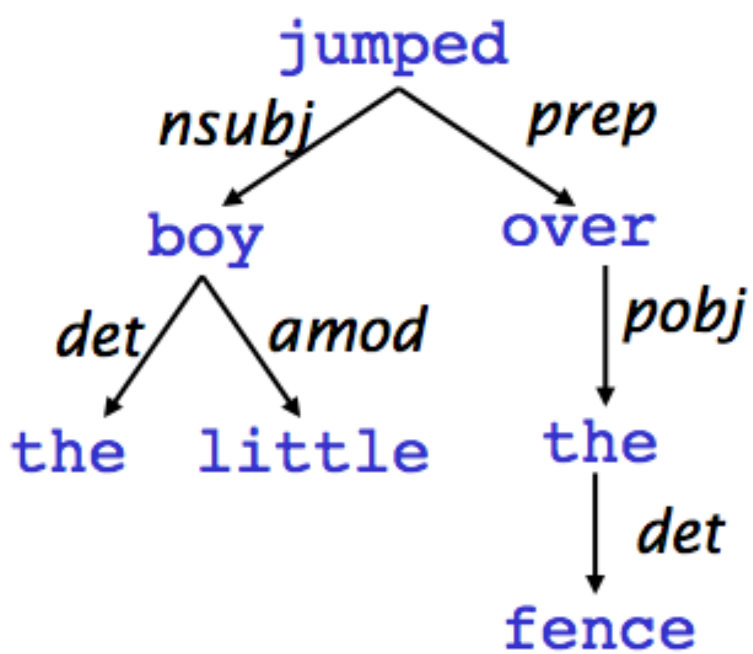
Random: 25%
Simple BOW features: 51.5%
SOTA: ~70%

Grammar Transfer between Two Languages

Danqi Chen

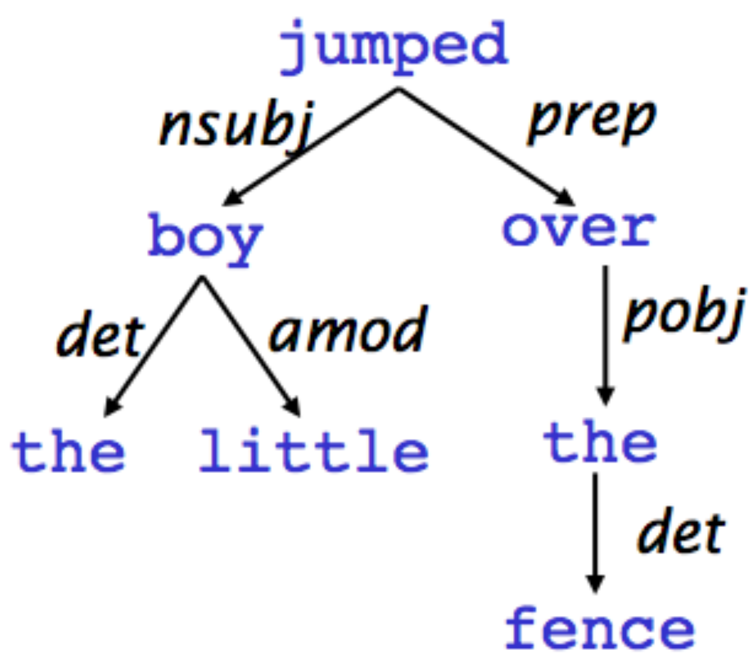
(danqi@cs.stanford.edu)

- You will learn syntactic parsing and dependency grammar soon (from Thursday)..



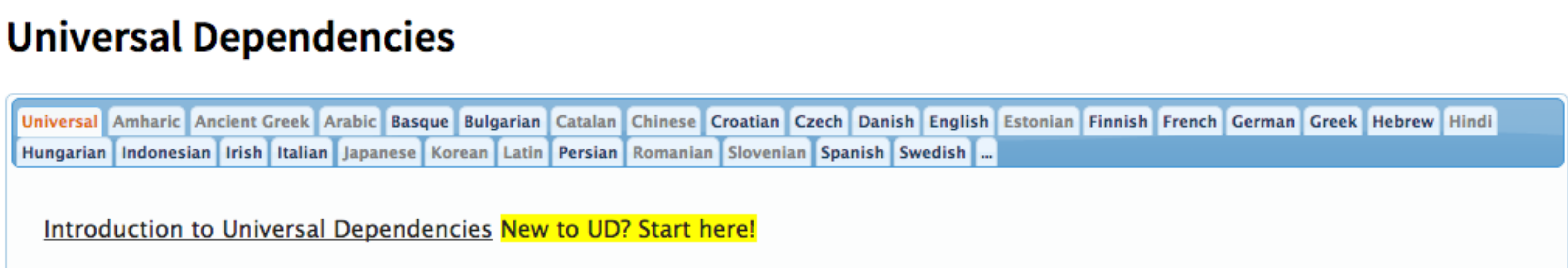
binary head-modifier relations between words

- You will learn syntactic parsing and dependency grammar soon (from Thursday)..

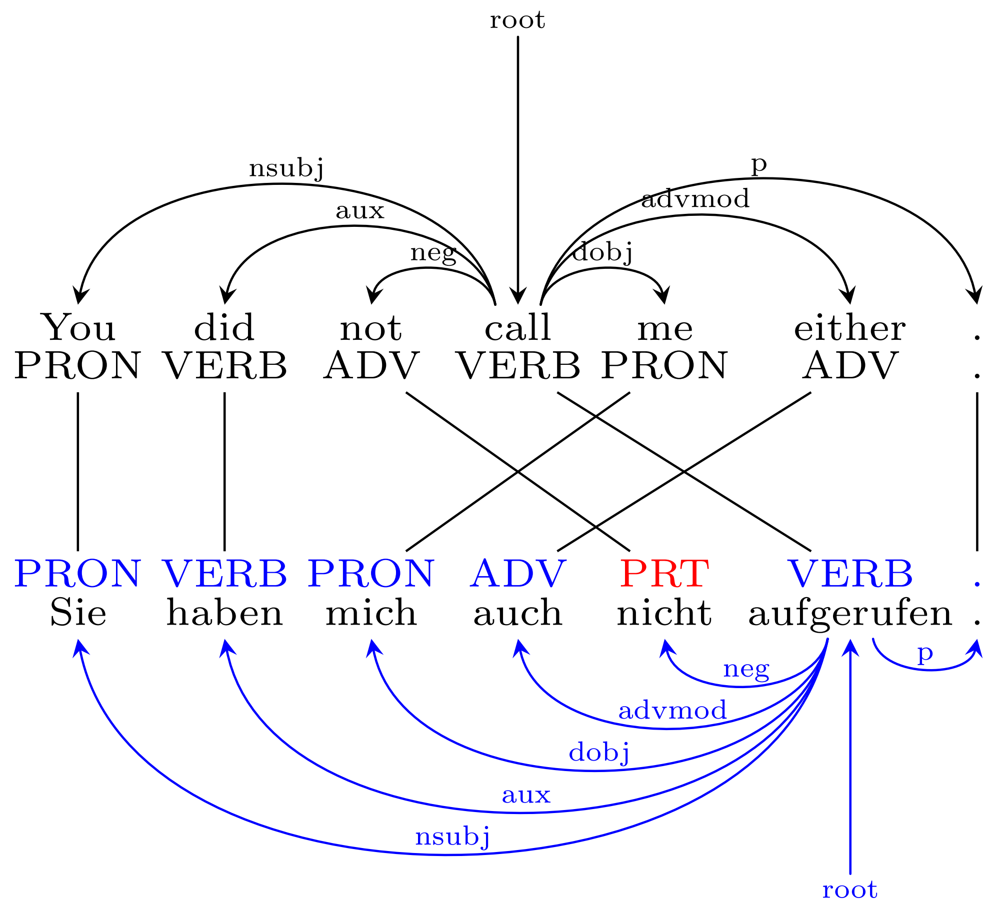


binary head-modifier relations between words

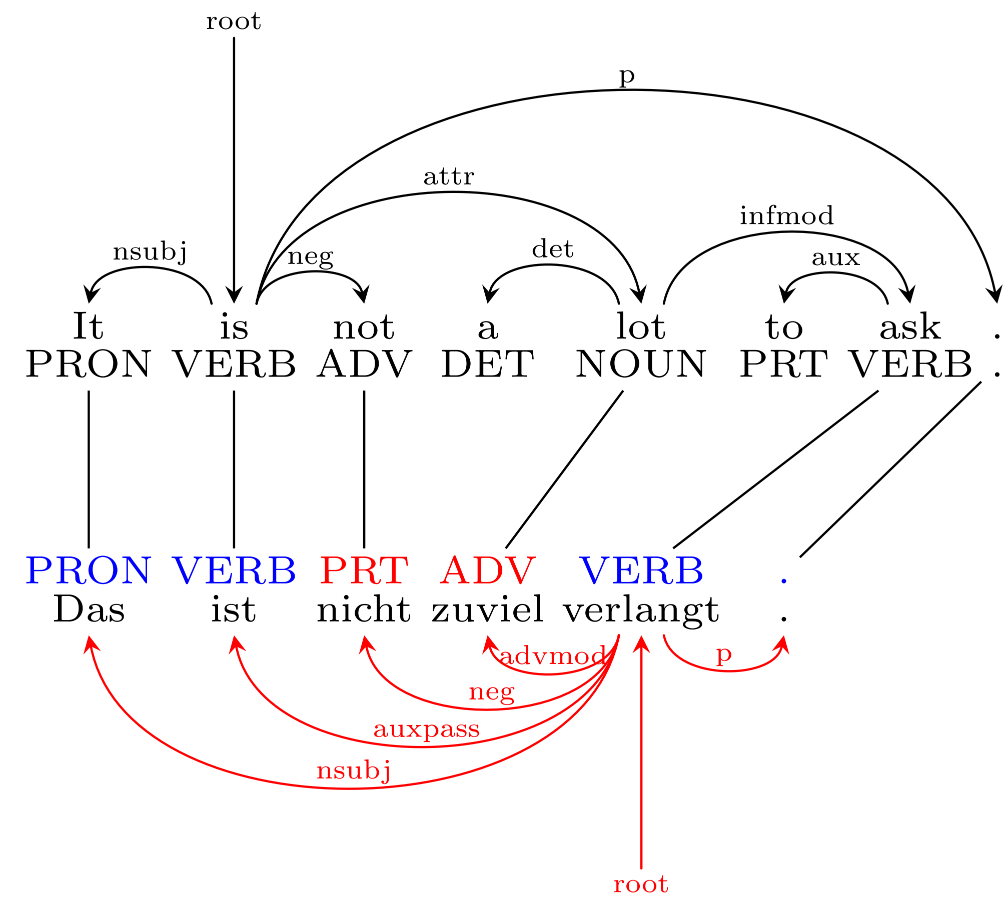
https://universaldependencies.github.io/docs/



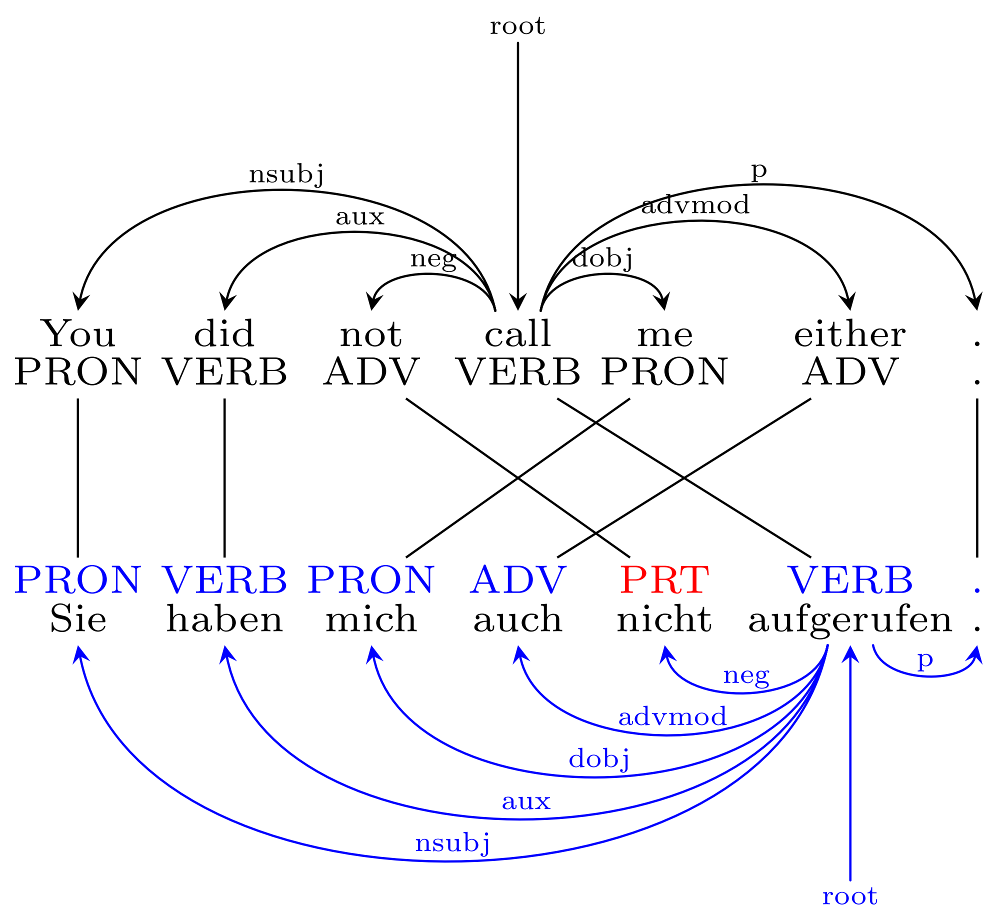
Good


root
nsubj
aux
neg
dobj
advmod
p
You did not call me either .
PRON VERB ADV VERB PRON ADV .
PRON VERB PRON ADV PRT VERB .
Sie haben mich auch nicht aufgerufen .
neg
advmod
dobj
aux
nsubj
p
root

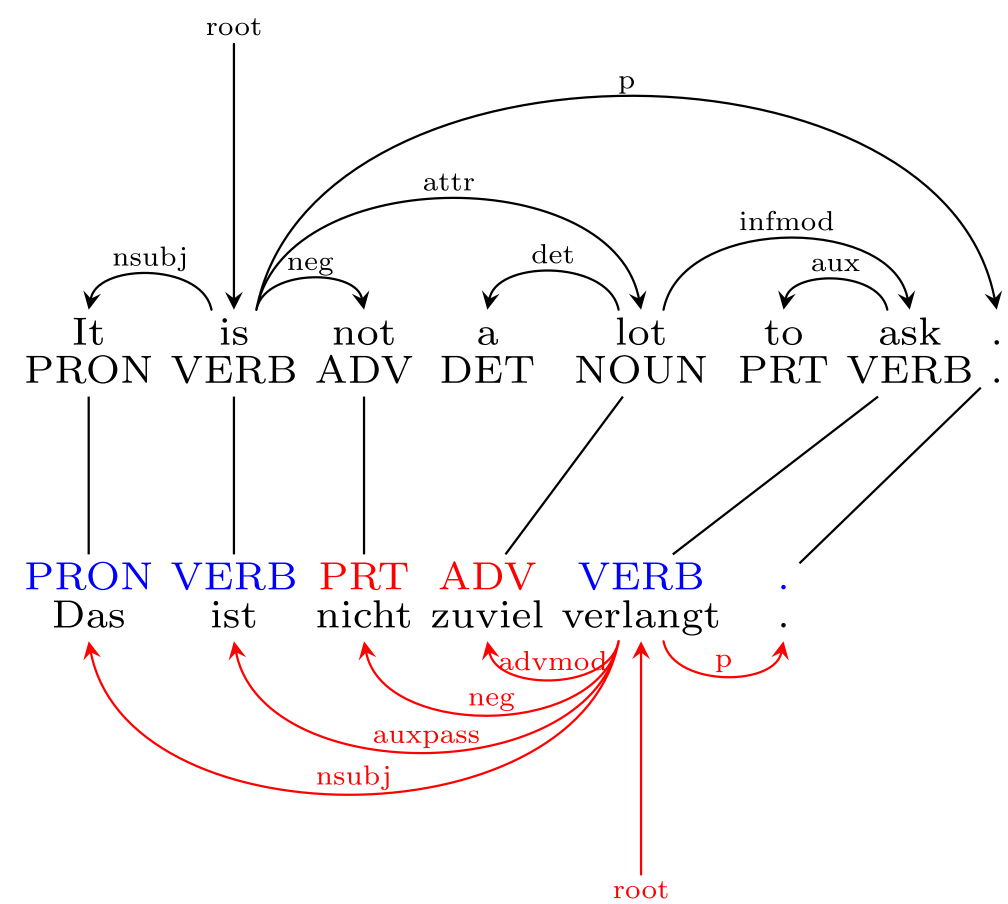
Bad


root
p
attr
infmod
nsubj
neg
det
aux
It is not a lot to ask .
PRON VERB ADV DET NOUN PRT VERB .
PRON VERB PRT ADV VERB .
Das ist nicht zuviel verlangt .
advmod
neg
auxpass
nsubj
p
root

Good



Bad



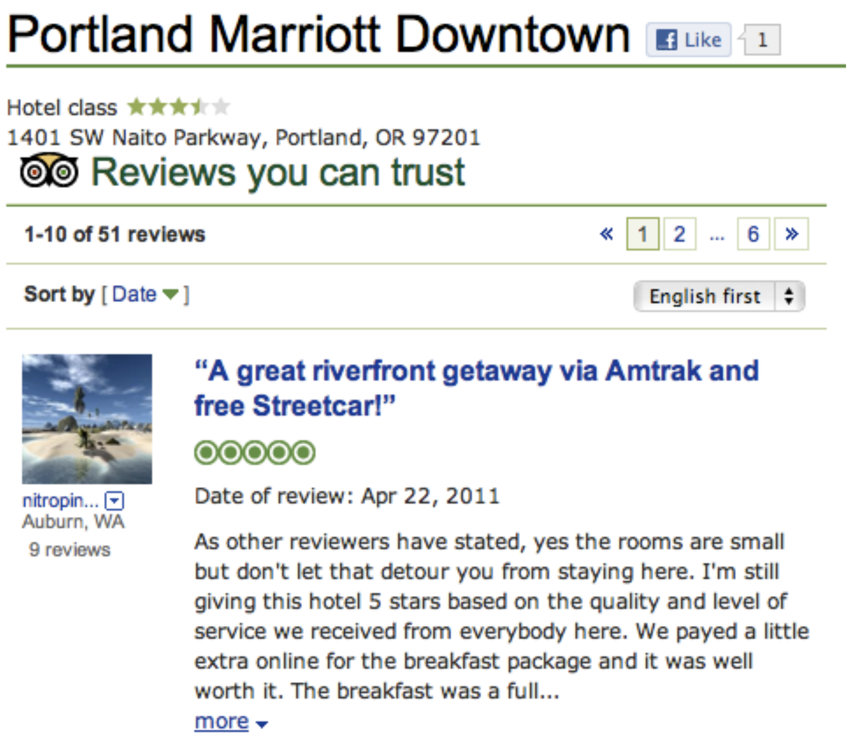
- **Objective #1**: analysis on how / when dependencies can be transferred through word alignments.
- **Objective #2 (advanced)**: use one language to (help) parse another language.
- **Recommended**: knowledge with one of those languages is a significant benefit.

Identifying Deceptive Opinion Spam

Jiwei Li
jiweil@stanford.edu

Motivation

- Consumers increasingly rate, review and research products online
- Potential for opinion spam
  - Deceptive opinion spam



Motivation

Which of these two hotel reviews is deceptive opinion spam?

Date of review: Jun 9, 2006

4 people found this review helpful

I have stayed at many hotels traveling for both business and pleasure and I can honestly stay that The James is tops. The service at the hotel is first class. The rooms are modern and very comfortable. The location is perfect within walking distance to all of the great sights and restaurants. Highly recommend to both business travellers and couples.

Date of review: Jun 9, 2006

4 people found this review helpful

My husband and I stayed at the James Chicago Hotel for our anniversary. This place is fantastic! We knew as soon as we arrived we made the right choice! The rooms are BEAUTIFUL and the staff very attentive and wonderful!! The area of the hotel is great, since I love to shop I couldn't ask for more!! We will definatly be back to Chicago and we will for sure be back to the James Chicago.

Motivation

Which of these two hotel reviews is deceptive opinion spam?

Answer:

Date of review: Jun 9, 2006

4 people found this review helpful

My husband and I stayed at the James Chicago Hotel for our anniversary. This place is fantastic! We knew as soon as we arrived we made the right choice! The rooms are BEAUTIFUL and the staff very attentive and wonderful!! The area of the hotel is great, since I love to shop I couldn't ask for more!! We will definatly be back to Chicago and we will for sure be back to the James Chicago.

Goal

Automates the process of fake review identification :

Using machine learning models !

Challenging but Meaningful!

CS224N Project Ideas

All of them are cool

Real Research Problems

David Jurgens

jurgens@stanford.edu

Or come by Gates 405

Publishable Results

Large-scale Language Identification

- **Objective**: Reliably classify what language a text is in when you have 200+ languages to choose from

Euskara Kiswahlil Cymraeg Gaelige

Frysk Hausa Yorùbá Igbo

Māori Occitan Wolof Удмурт

- **Social Good**: Help raise the visibility of indigenous languages — spoken by millions!

Large-scale Language Identification (2)

- **Requirements**: ability to train a classifier (or willingness to learn)
  - Python or Java would be great
  - You don't have to speak these languages!
- **Side Benefit**: Learn all about other cultures!
- **Most data is ready — but more is available**

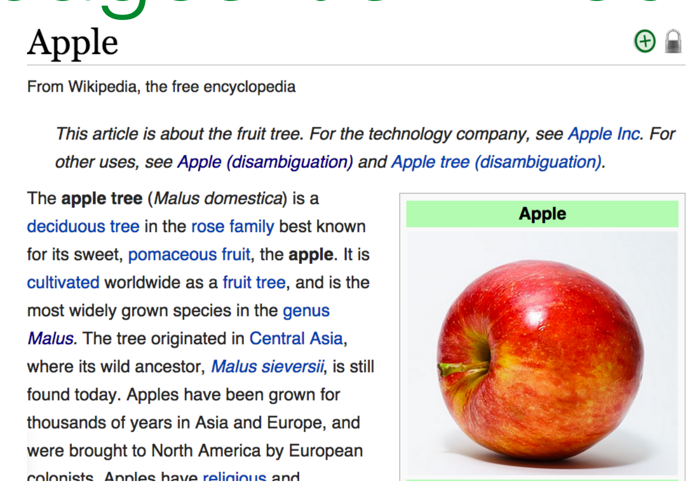
Collaboratively-curated Knowledge for Word Sense Disambiguation

- **Objective**: Improve Word Sense Disambiguation using Wikipedia text

What's the meaning of this usage of "apple"?

Malus (/ˈmeɪləs/[3] or /ˈmæləs/) is a genus of about 30–55 species[4] of small deciduous apple trees or shrubs in the family Rosaceae, including the domesticated orchard apple (*M domestica*). The other species are generally known as **crabapples**, crab apples, crabs, or wild apples.

Learn by looking at the pages it's linked to!


Apple

From Wikipedia, the free encyclopedia

This article is about the fruit tree. For the technology company, see Apple Inc. For other uses, see Apple (disambiguation) and Apple tree (disambiguation).

The **apple tree** (*Malus domestica*) is a deciduous tree in the rose family best known for its sweet, pomaceous fruit, the **apple**. It is cultivated worldwide as a fruit tree, and is the most widely grown species in the genus *Malus*. The tree originated in Central Asia, where its wild ancestor, *Malus sieversii*, is still found today. Apples have been grown for thousands of years in Asia and Europe, and were brought to North America by European colonists. Apples have religious and

Collaboratively-curated Knowledge for Word Sense Disambiguation (2)

- **Requirements**: ability to train a classifier (or willingness to learn)
    - Python or Java would be great
    - Need to work with Wikipedia data (~10GB)
- **Side Benefit**: Learn esoteric meanings of words and appreciate just how complex semantics can be! Also, learn all about Wiki markup.

Overcoming Sense Annotation Sparsity through Paraphrase Generation

- **Objective**: Improve word sense disambiguation performance by expanding the training data with paraphrases
  - **Key question**: We need more training data, so can we just automatically make more?
- **Requirements**: Basic familiarity training a classifier
- **Side benefit**: Learn all about semi-supervised learning
- **Paraphrase Data is already prepared!**

Pertainymy Identification

- **Objective**: Figure out the topical domain of specific concepts, if it it exists
  - Doctor -> medicine, Tractor -> farming, Blackboard -> teaching, etc.
- Open-ended methodology: (a) learn from examples in WordNet? (b) combine new information from Wiktionary? (c) use corpus statistics? — Your choice!
- **Side Benefit**: Work on a semantic relationship *other than* hypernymy or meronymy
- **Requirements**: Familiarity with WordNet helps

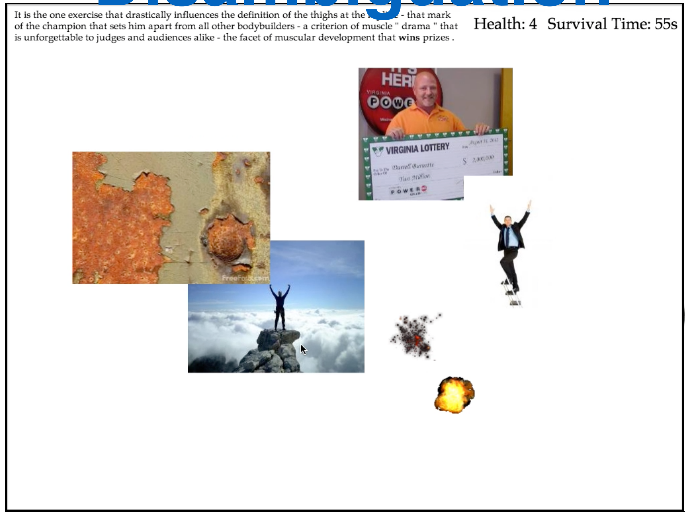
Video Games for Semantic Annotation

- **Objective**: Annotate linguistic data with a video game!

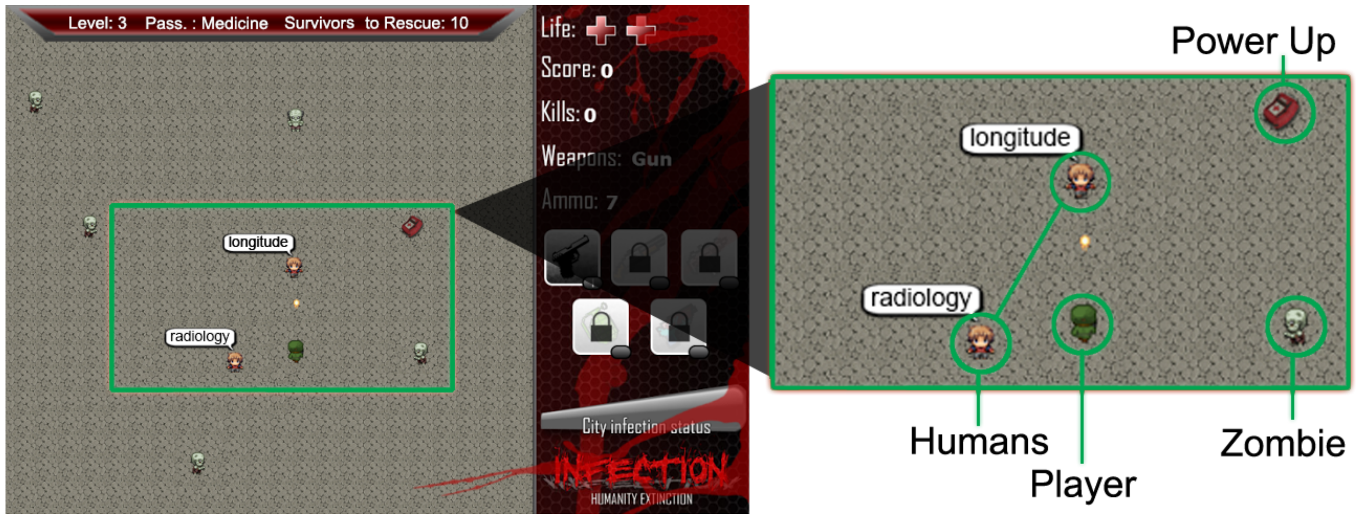
Image-Concept mapping

Image-Concept mapping

Word Sense Disambiguation



Relation Extraction



Video Games for Semantic Annotation (2)

- **Warning**: High-risk, high reward
- **Requirements**: Video game design will help (a lot)
  - Remember, it has to work *and* produce annotations — but there are game-dev toolkits that make it easier!
- **Side benefit**: probably the coolest demo you'll ever produce from a class project
- **Lots of data is available** (and I built other games and can give you tips)

Literary Social Network Extraction

- **Objective**: Find the characters in the book and the social network for how they interact — Not as easy as it seems!
- **Open-ended Methodology with a Concrete Application**: Lots of new ideas to try: Project can match your interest and use off-the-shelf tools for the rest!
- **Side-benefit**: You'll become familiar with a huge set of NLP tools: POS tagging, WordNet, Named Entity Recognition, Coreference Resolution, etc.
- **Requirements:** Using Java and CoreNLP is easiest, but Python is also possible. Experience with classifiers is helpful but not necessary.
- **Data is easily available** (Project Gutenberg!)

Figuring out what kind of thing something is

- **Objective**: Given a word and its definition, Identify either: (a) the WordNet synset that is its hypernym or (b) the WordNet synset that is synonymous
  - changing_room#n - A room, especially in a gym, designed for people to change their clothes.
  - mudslide#n - a mixed drink consisting of vodka, Kahlua and Bailey's.
- **Requirements**: familiarity with classifiers and WordNet will help, but other approaches are possible!
- **Data is already prepared!**

Challenging but Meaningful!

Contact David Jurgens for more details on any of these projects

Real Research Problems

jurgens@stanford.edu

Or come by Gates 405

Publishab
le
Results

Deep Learning for Medical Record Understanding

Kelvin Guu, Stanford NLP

Medical records

Date: October 5, 2015

The patient is a healthy-appearing 56-year-old male. His height is 5 feet 9 inches. He weighs 150 pounds. Examination of the upper extremities reveals normal muscle tone, bulk, and strength with normal sensation throughout. Review of his MRI shows multilevel degenerative disease with some degenerative scoliosis …

Every time you see the doctor, you might not know that they write a 1-5 page note about you.
It's used to support disease diagnoses, and justify hefty insurance payments.

So, it's a critical piece of data that supports a very large and complex medical decision-making system.

But it has the potential to do even more than that.

If we could do automated large-scale analyses of these records, there's the potential to discover entirely new patterns in disease progression and treatment outcomes.

This is part of a growing movement to transform medical decision making with NLP and machine learning. Basically, to get us to a point where a single doctor can learn from the countless experiences and decisions of all other doctors.

Medical records

Date: October 5, 2015

The patient is a healthy-appearing 56-year-old male. His height is 5 feet 9 inches. He weighs 150 pounds. Examination of the upper extremities reveals normal muscle tone, bulk, and strength with normal sensation throughout. Review of his MRI shows multilevel degenerative disease with some degenerative scoliosis …

- Support **disease diagnoses**

Every time you see the doctor, you might not know that they write a 1-5 page note about you.
It's used to support disease diagnoses, and justify hefty insurance payments.

So, it's a critical piece of data that supports a very large and complex medical decision-making system.

But it has the potential to do even more than that.

If we could do automated large-scale analyses of these records, there's the potential to discover entirely new patterns in disease progression and treatment outcomes.

This is part of a growing movement to transform medical decision making with NLP and machine learning. Basically, to get us to a point where a single doctor can learn from the countless experiences and decisions of all other doctors.

Medical records

Date: October 5, 2015

The patient is a healthy-appearing 56-year-old male. His height is 5 feet 9 inches. He weighs 150 pounds. Examination of the upper extremities reveals normal muscle tone, bulk, and strength with normal sensation throughout. Review of his MRI shows multilevel degenerative disease with some degenerative scoliosis …

- Support **disease diagnoses**
- Justify **insurance payments**

Every time you see the doctor, you might not know that they write a 1-5 page note about you.
It's used to support disease diagnoses, and justify hefty insurance payments.

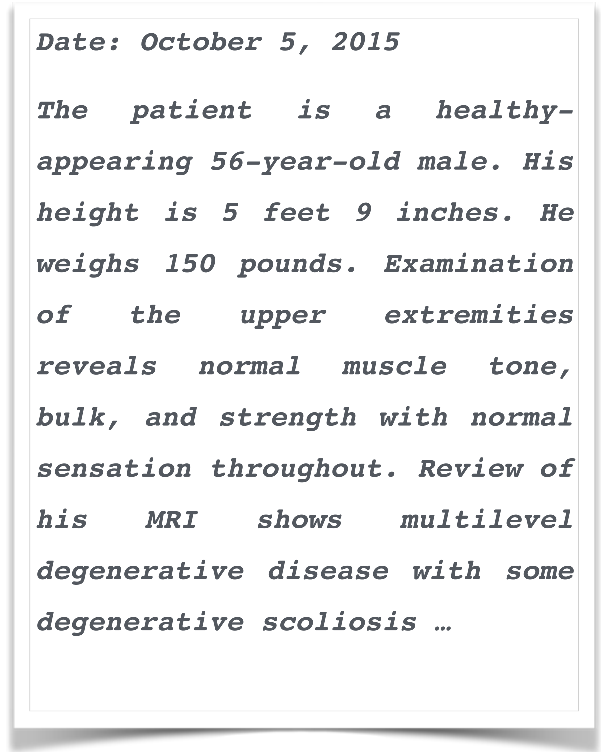
So, it's a critical piece of data that supports a very large and complex medical decision-making system.

But it has the potential to do even more than that.

If we could do automated large-scale analyses of these records, there's the potential to discover entirely new patterns in disease progression and treatment outcomes.

This is part of a growing movement to transform medical decision making with NLP and machine learning. Basically, to get us to a point where a single doctor can learn from the countless experiences and decisions of all other doctors.

Medical records



Date: October 5, 2015

The patient is a healthy-appearing 56-year-old male. His height is 5 feet 9 inches. He weighs 150 pounds. Examination of the upper extremities reveals normal muscle tone, bulk, and strength with normal sensation throughout. Review of his MRI shows multilevel degenerative disease with some degenerative scoliosis …

- Support **disease diagnoses**
- Justify **insurance payments**
- Large scale analysis: **Discover new patterns in disease progression and treatment outcomes**

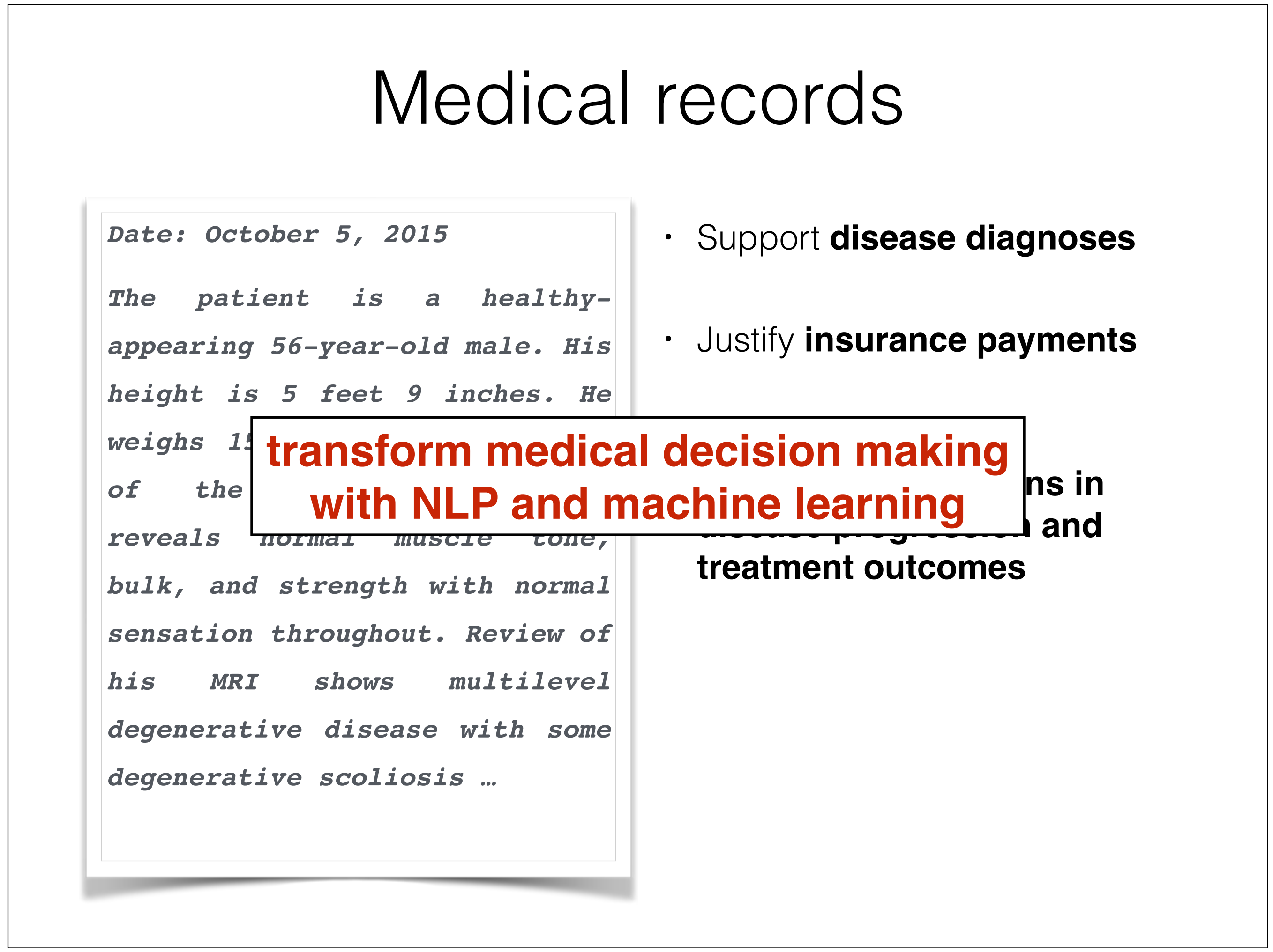
Every time you see the doctor, you might not know that they write a 1-5 page note about you.
It's used to support disease diagnoses, and justify hefty insurance payments.

So, it's a critical piece of data that supports a very large and complex medical decision-making system.

But it has the potential to do even more than that.

If we could do automated large-scale analyses of these records, there's the potential to discover entirely new patterns in disease progression and treatment outcomes.

This is part of a growing movement to transform medical decision making with NLP and machine learning. Basically, to get us to a point where a single doctor can learn from the countless experiences and decisions of all other doctors.



Every time you see the doctor, you might not know that they write a 1-5 page note about you.
It's used to support disease diagnoses, and justify hefty insurance payments.

So, it's a critical piece of data that supports a very large and complex medical decision-making system.

But it has the potential to do even more than that.

If we could do automated large-scale analyses of these records, there's the potential to discover entirely new patterns in disease progression and treatment outcomes.

This is part of a growing movement to transform medical decision making with NLP and machine learning. Basically, to get us to a point where a single doctor can learn from the countless experiences and decisions of all other doctors.

Challenges

Today, this potential is just completely unrealized. Suppose we want to answer a simple question like this.

We face two problems with simple keyword-style search. On the one hand, language has a lot of variability. But if you try to cast your net too wide, you end up getting distracted.

Challenges

"How many patients feel dizzy after taking sinus medication?"

Today, this potential is just completely unrealized. Suppose we want to answer a simple question like this.

We face two problems with simple keyword-style search. On the one hand, language has a lot of variability. But if you try to cast your net too wide, you end up getting distracted.

Challenges

"How many patients feel dizzy after taking sinus medication?"

language variability (hurt recall):

nausea after taking pseudoephedrine

felt dizzy from sudafed

Today, this potential is just completely unrealized. Suppose we want to answer a simple question like this.

We face two problems with simple keyword-style search. On the one hand, language has a lot of variability. But if you try to cast your net too wide, you end up getting distracted.

Challenges

"How many patients feel dizzy after taking sinus medication?"

language variability (hurt recall):

nausea after taking pseudoephedrine

felt dizzy from sudafed

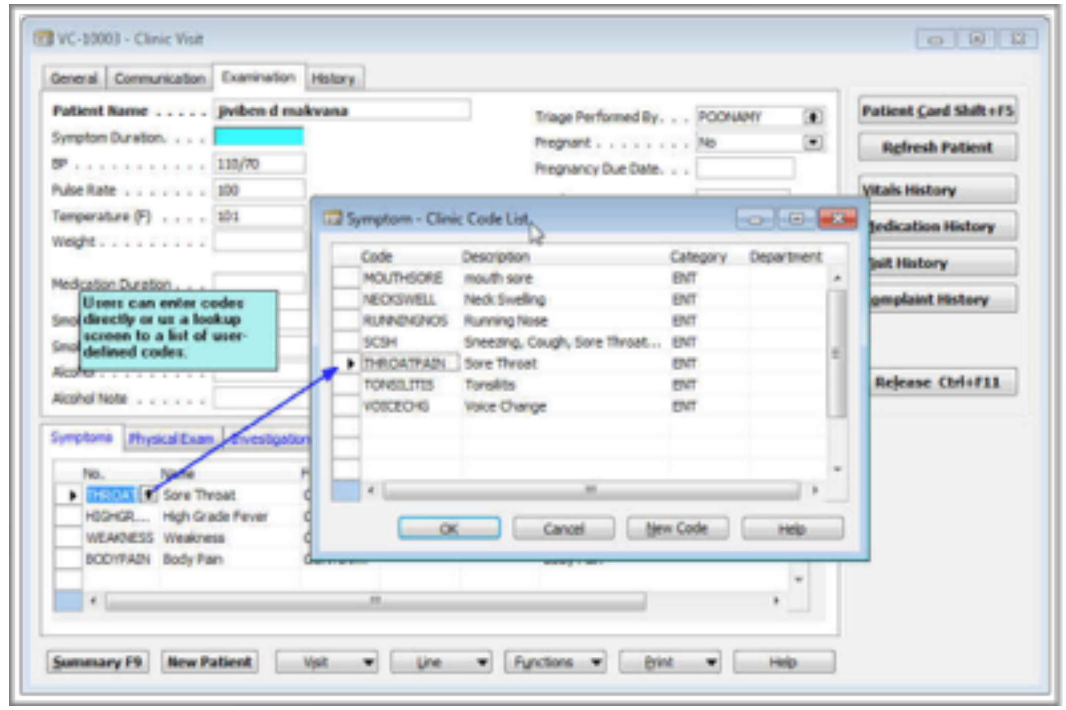
distractors (hurt precision):

patient was concerned that ***sudafed causes nausea***

Today, this potential is just completely unrealized. Suppose we want to answer a simple question like this.

We face two problems with simple keyword-style search. On the one hand, language has a lot of variability. But if you try to cast your net too wide, you end up getting distracted.

Clickarrhea



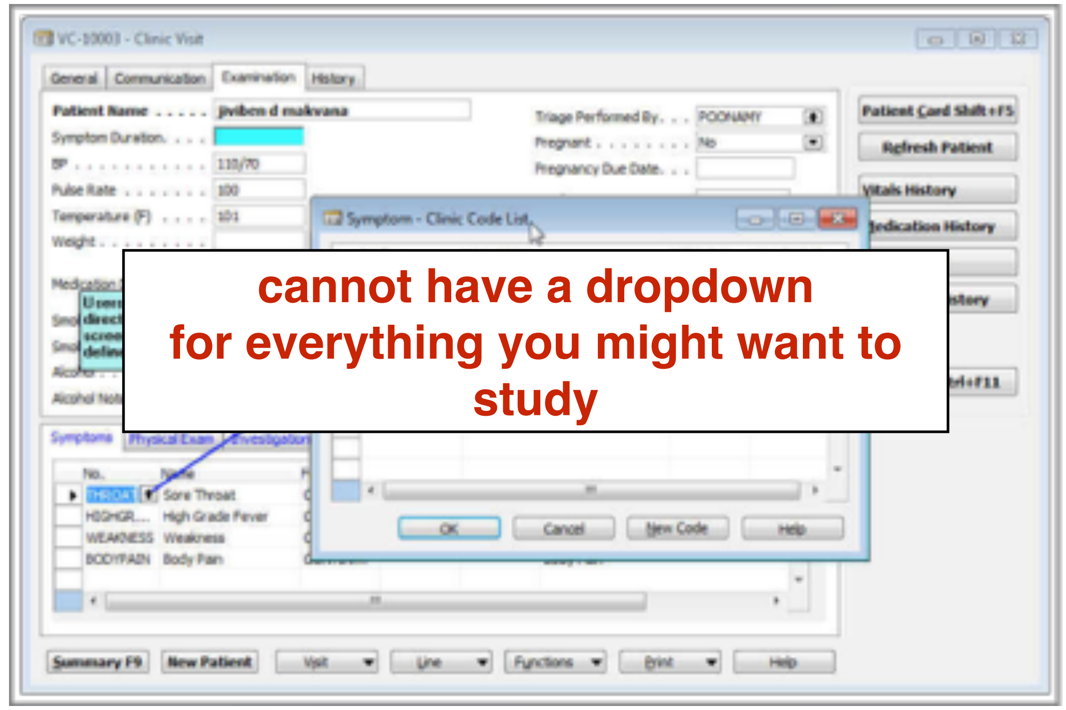
And in case you were thinking that maybe we could solve this with more dropdown menus, there are some serious obstacles to that solution.

Doctors and nurses already complain that they suffer from clickarrhea, which is the need to click through an ever-growing list of menus.

The problem is that you just can’t have a dropdown for everything you might want to retroactively study. And because insurance regulations and best treatment practices are constantly changing, predefined dropdown menus actually prevent institutions from rapidly adapting.

We think that you should just write what you want, and use NLP to extract what you want later.

Clickarrhea



And in case you were thinking that maybe we could solve this with more dropdown menus, there are some serious obstacles to that solution.

Doctors and nurses already complain that they suffer from clickarrhea, which is the need to click through an ever-growing list of menus.

The problem is that you just can't have a dropdown for everything you might want to retroactively study. And because insurance regulations and best treatment practices are constantly changing, predefined dropdown menus actually prevent institutions from rapidly adapting.

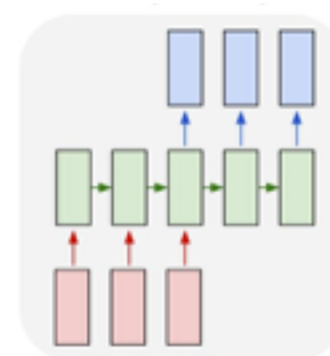
We think that you should just write what you want, and use NLP to extract what you want later.

Approach

Deep learning

Structured data sources

Large-scale unsupervised learning



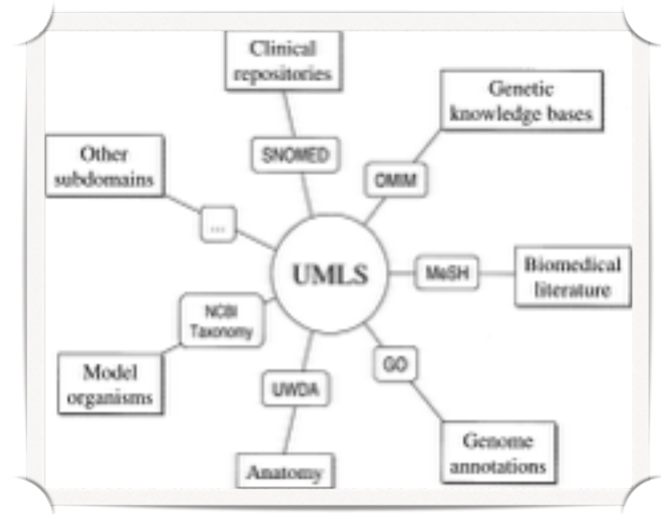
In this project, we'll be exploring new methods in deep learning for extracting structured knowledge. This could include the use of recently developed neural attention mechanisms, as well as flexible models for learning measures of sentence similarity.

Approach

Deep learning

Structured data sources

Large-scale unsupervised learning



Our source of labeled data will come from the structured databases that supplement medical notes. Fortunately for us, clinicians have been entering structured data for a long time, giving us something to learn from. We will also explore interesting ways to leverage medical ontologies and other knowledge sources.

Approach

Deep learning

Structured data sources

Large-scale unsupervised learning

And the previous two elements won't be enough to develop a successful approach. The medical datasets we have at the moment feature roughly 30,000 patient histories.

That may sound like a lot, but it's far too little to learn the rules of natural language from scratch.
To do that, we'll aim to learn from much larger natural language corpora that don't have anything to do with medicine, but can still transfer important linguistic information.

Why do this project?

If all this has you interested, here are just a few more reasons why this could be the project for you.

(The previous project that we mentored also led to a publication 3 months later.)

Why do this project?

- Serious mentorship

If all this has you interested, here are just a few more reasons why this could be the project for you.

(The previous project that we mentored also led to a publication 3 months later.)

Why do this project?

- Serious mentorship
- The latest tools/libraries for deep learning

If all this has you interested, here are just a few more reasons why this could be the project for you.

(The previous project that we mentored also led to a publication 3 months later.)

Why do this project?

- Serious mentorship
- The latest tools/libraries for deep learning
- Best practices for rapid experimentation

If all this has you interested, here are just a few more reasons why this could be the project for you.

(The previous project that we mentored also led to a publication 3 months later.)

Who we’re looking for

And in exchange, we’re looking for someone who’s really willing to work hard and get into the problem. You’ll need strong programming skills, but knowledge of NLP and deep learning are an optional plus.

Contact me

kguu@stanford.edu

Contact me if you’re interested!