

AI for Medicine

Lecture 12: Medical NLP

Dr. Gabriel Stanovsky

January 16, 2023



Text-processing models which exhibit facets of **human intelligence** with benefits for users in **real-life applications**

Grand Challenges in NLP



Machine translation

"The universal translator, invented in 2151, is used for deciphering unknown languages"



Information retrieval

"What's the second largest star in this galaxy?"



Automated assistants

"I got one of those terrible headaches from lack of sleep. Can you give me something for it?"

Grand Challenges in NLP



Machine translation

“the *universal translator*, invented in 2151, is used for *deciphering unknown languages*”

NLP models need to capture the
meaning behind our words and interact accordingly



galaxy?”



Automated assistance

“I got one of those *terrible headaches* from *lack of sleep*. Can you give me something for *it*?”

Today

- 1 Why is NLP hard?
- 2 Case Study: Recognizing Side Effects on Social Media
- 3 Machine Learning in NLP
 - Obtaining Labelled Data
 - Language Modelling
 - Model Evaluation
- 4 Putting it All Together
- 5 Conclusion

Why is NLP hard?

There are many languages (1000s)



Why is NLP hard?

Language is often imprecise



Saw it... A good movie to fall asleep to.



Panic Patrol at the disco @SONATORE_ · Dec 21

...

I've been eating a **ton of fruit the past month** since I've cut out added **sugar**. My skin and hair feel so good rn 🥰

Why is NLP hard?

Language is often ambiguous

מינוי בכיר

פייסבוק מודיעה: אדם מוסרי יעמוד בראש אינסטגרם

מוסרי, מנהל בכיר בפייסבוק, ימלא את מקומם של המייסדים הפורשים קווין סיסטרום ומייק קריגר • "אנו שמחים למסור את המושכות למנהיג מוצר עם רקע עיצובי חזק ומיקוד במלאכה ובפשטות", אמרו הפורשים



אדם מוסרי / צילום: רויטרס

Example from Mohr Wenger

San Jose cops kill man with knife

Text Paper

Translate Listen

CHS

San Jose cops kill man with knife

Ex-college football player, 23, shot 9 times allegedly charged police at fiancée's home

By Eliana Alvarez
and Victoria Ho

A man fatally shot by two San Jose police officers while allegedly charging at them with a knife was a 23-year-old former football player at San Jose State University who was reportedly well-known, his family said.

Thursday
Police officials said two officers opened fire Wednesday outside the home of the man's fiancée.

But, it turned out, had been made by William Hoover. The mother of William Hoover, who also lives in the home on the corner of the street, said she witnessed the shooting and described it as chaotic. Page 10 of 10

shooting after the victim's wife intervened. Hoover is in the hospital in hopes of getting him out of the hospital.

Hoover's wife said she was at the home when the man charged at the officers with a knife. She said she was in the hallway when the man charged at the officers with a knife. She said she was in the hallway when the man charged at the officers with a knife.

Hoover's wife said she was at the home when the man charged at the officers with a knife. She said she was in the hallway when the man charged at the officers with a knife.

Hoover's wife said she was at the home when the man charged at the officers with a knife. She said she was in the hallway when the man charged at the officers with a knife.

Hoover's wife said she was at the home when the man charged at the officers with a knife. She said she was in the hallway when the man charged at the officers with a knife.

Hoover's wife said she was at the home when the man charged at the officers with a knife. She said she was in the hallway when the man charged at the officers with a knife.

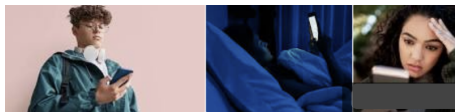
Students Cook & Serve Grandparents

On Thursday, September 9, Gorman School hosted the first annual Grandparent's Day.

All Grandparents were invited to a school wide pancake breakfast. Upper grade students served as excellent chefs, as well as taking responsibility for serving the food and the clean up after-

Why is NLP hard?

Language changes all the time



Doomscrolling

Doomscrolling is the act of consuming an endless procession of negative online news, to the detriment of the scroller's mental wellness. [Wikipedia](#)

- **Medical NLP** receives a lot of attention
- **Lots of applications**
 - Medical chatbots
 - Information extraction (more in a sec)
 - Summarizing doctor's notes
 - and more...

Lots of interest in Academia and Industry

מערכות בינה מלאכותית נכנסות לקודש הקודשים של הרפואה. האם ניתן לסמוך עליהן?

אחרי שרובוטים כבר השתלבו בחדי הניתוח, גם תהליך האיבחון הרפואי עובר מהפכה, ובקרוב נפגוש יותר ויותר מחשבים בחדר הבדיקה

8 תגובות קריאת נ 1



עמותת טריריות חולים בשנחאי. חפזת מההשפעות חיוביות, לטוב ולרע צילום: China News Service / Getty Images

Microsoft Israel uses AI to translate medical data for doctors



By Alice Chambers on 22 November 2022

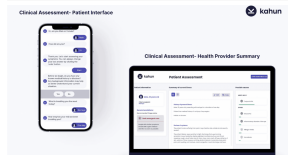


Kahun raises \$8M for AI enabled clinical reasoning chatbot

The startup's first product asks patients questions about their concerns and provides a summary with recommendations for follow-up to the physician.

By Emily Olsen | September 29, 2022 | 01:56 pm

SHARE



TheMarker | TechNation

K Health: סמארטפון, מה יש לי?

אפליקצית קיי-הלת' משווה בין דיווחים של משתמשים על מצבם הבריאותי רפואיים שנעשו למטופלים דומים להם

Identifying Adverse Drug Reaction in Social Media

Adverse Drug Reaction

Unwanted reaction clearly associated with the intake of a drug



I stopped taking Ambien after a week, it gave me **a terrible headache!**

Motivation

- Discover unknown side effects
- Monitor drug reactions over time
- Respond to patient's complaints

Challenges

- Context dependent
 - Ambien **gave me** terrible headaches
 - Ambien made my terrible headaches **go away**
- Colloquial
 - been having a hard time **getting some Z's** after ambien

ML Pipeline for NLP



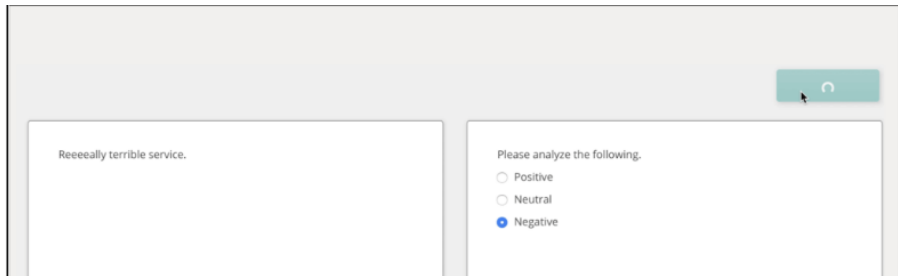
annotated data

model training

evaluation

Obtaining Labeled Data

- Usually done via **crowdsourced** annotations



The image shows a screenshot of a crowdsourcing annotation interface. It features a light gray background with a teal button in the top right corner containing a circular arrow icon. Below this, there are two white rectangular boxes. The left box contains the text "Reeeeeeally terrible service." The right box contains the text "Please analyze the following." followed by three radio button options: "Positive", "Neutral", and "Negative". The "Negative" option is selected, indicated by a blue dot next to the radio button.

Reeeeeeally terrible service.

Please analyze the following.

☐ Positive

☐ Neutral

☒ Negative

Language is Subjective

- Are these positive or negative reviews?



Hummingbirds **Top Contributor: Running**

These sound good, but not the best earhook



Victoria Chen

Good software, bad physical design specs

Obtaining Labelled Data

- Much harder in medicine where **crowdsourcing won't work**
 - As annotations often require expertise
 - E.g., identifying medication names, and their known side effects
- Instead often relies on **in-house annotators**
 - and hence the cost per sample is much higher



annotated data



model training



evaluation

How to Represent Text?

- All models require **numerical input feature vectors**: $x \in \mathbb{R}^d$
 - E.g., representing a patient as (*height, age, blood pressure*)
 - Major design choice
- Perhaps most important question in NLP: **how to represent text?**

Input representation is clearer in e.g. computer vision

- Think about an object detection problem



Input representation is clearer in e.g. computer vision

- Think about an object detection problem



Input representation is clearer in e.g. computer vision

- Think about an object detection problem



Input representation is clearer in e.g. computer vision

- Think about an object detection problem



Input representation is clearer in e.g. computer vision

- Think about an object detection problem



Input representation is clearer in e.g. computer vision

- Think about an object detection problem



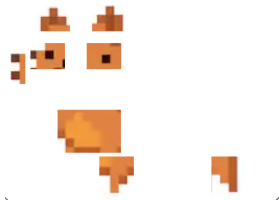
Input representation is clearer in e.g. computer vision

- Think about an object detection problem



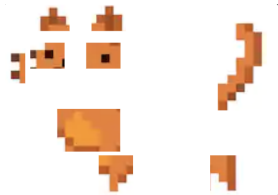
Input representation is clearer in e.g. computer vision

- Think about an object detection problem



Input representation is clearer in e.g. computer vision

- Think about an object detection problem



Input representation is clearer in e.g. computer vision

- Think about an object detection problem



What's desired from an input representation?

- **Compositional**

- An image of a cat is composed of a tail, ears, whiskers, ...
- A tail is composed of a slim and narrow column of pixels ...

- **Continuous**

- Small change in representation $\Rightarrow$ small change in semantics
- Similar images are near in representation space
- Changing one pixel of a cat is still a cat

- **Dense**

- Not a lot of zero entries (sparse)
- Condensed representation



Natural language defies desired properties

✗ Often isn't compositional

- E.g., idioms: *Kick the bucket*, *Fall off the wagon* ...

✗ Isn't dense

- E.g., one-hot vectors

✗ Isn't continuous

- Small change in representation $\nrightarrow$ small change in semantics

Natural language defies desired properties

✗ Often isn't compositional

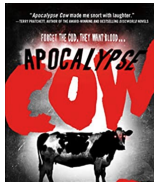
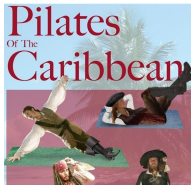
- E.g., idioms: *Kick the bucket*, *Fall off the wagon* ...

✗ Isn't dense

- E.g., one-hot vectors

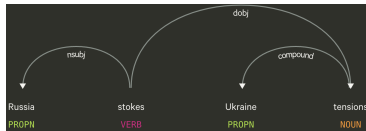
✗ Isn't continuous

- Small change in representation $\nrightarrow$ small change in semantics

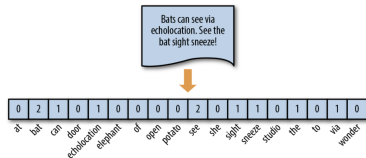


Text Representation Before LMs: Feature Engineering

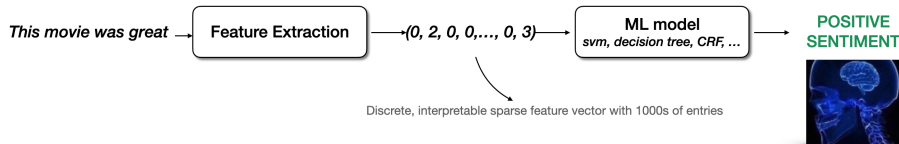
- **Part of speech and/or syntax**
 - E.g., number of nouns as a feature



- **Task-specific features**
 - E.g., indicators for negation words
- **Bag of words**



Building NLP Models Before Language Models

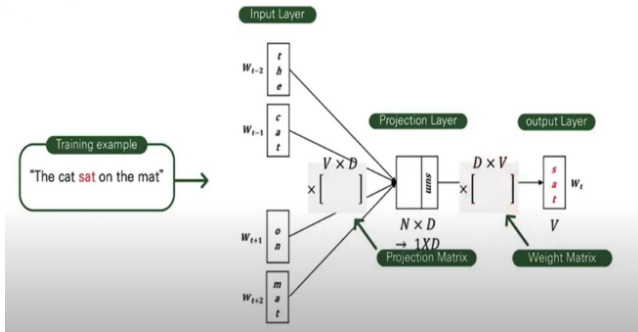


- **Mostly monolithic models**
 - Different representation for translations, summarization, etc.
 - Rare to reuse parts between tasks or architecture
- **Close coupling between representation and modelling**

Language Modelling

Predict the next word in a sequence

- **Several variants**
 - *Causal*: John asked Mary on a ___
 - *Bidirectional*: John ___ Mary on a date
- **Q: What's convenient about this task?**
 - Hint: Where can we get **A LOT** of training data?



- Start with **random initialization** of word features
- Then train them to **predict each word in context**
- End with hopefully a **meaningful representation for each word**
- ▶ Demo

- **Q: Why does this happen?**
 - What characterizes all words which can fill a gap?

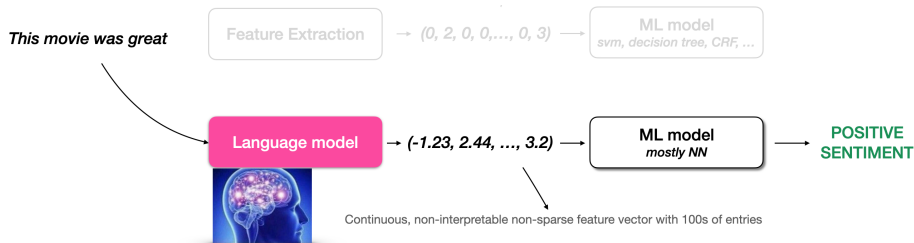
The Distributional Hypothesis

"You shall know a word by the company it keeps."

Firth (1957)

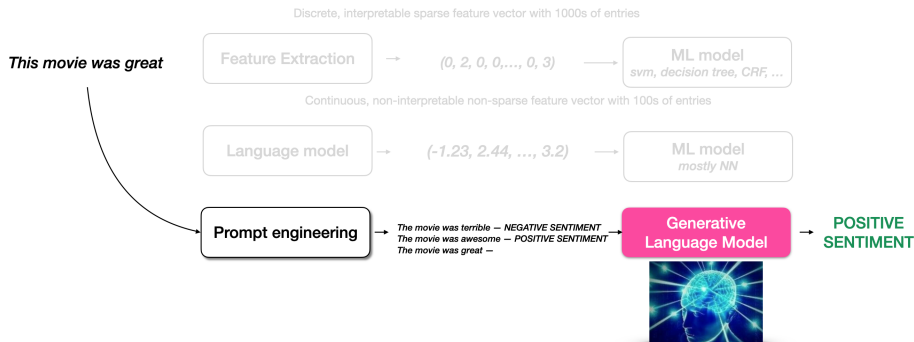
- **Q: Where can this break?**
 - What other words tend to co-occur?

Finetuning

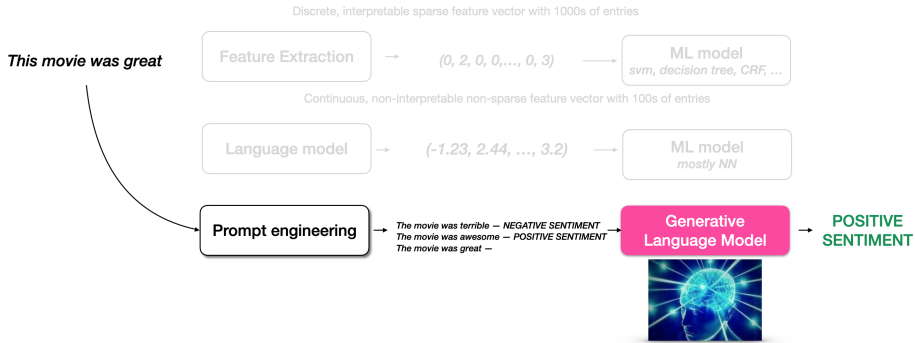


- First, language models are **pretrained**
 - These are made publicly available (e.g., BERT)
 - then **finetuned** for a specific task
- **Decoupling between representation and modelling**
 - LMs are reused between different tasks
- **Achieved state-of-the-art results on practically all NLP**

Prompting



- **Generative LMs trained to complete a sequence**
- **In-context Learning from a few examples**
 - Without any weight updates
- **Getting better results every day**



- **Generative LMs claim to mimic human traits**
 - “Learn” by example
 - Single model solves many tasks without training
- **This is what GPT and its variants do**

Domain-Specific Pretraining

- **There's no single “generic” English** (or any other language)
 - “Cancer” means different things in astrology. . .
 - and would tend to appear in different contexts
- Lots of efforts into pretraining on **specialized domains**

PubMedGPT 2.7B

Authors: Elliot Bolton and [David Hall](#) and [Michihiro Yasunaga](#) and [Tony Lee](#) and [Chris Manning](#) and [Percy Liang](#)

We introduce a new 2.7B parameter language model trained on biomedical literature which delivers an improved state of the art for medical question answering.



annotated data



model training



evaluation

How do we evaluate?

- There are often many equally correct answers
 - Is there a single “correct” translation of a book?
 - A single “correct” summary?
- Compare to object detection where agreement is high
- This is even evident in our toy example



I stopped taking Ambien after a week, it gave me **a terrible headache!**

- **Q: Did taking ambien cause headache,
or did stopping taking it cause the headache?**

Inter-Annotator Agreement

- Ask several experts to annotate texts
 - and hence drives the annotation cost even higher
- Check their **agreement**, e.g., how often do they agree
- This is an **upper bound** for model performance



annotated data



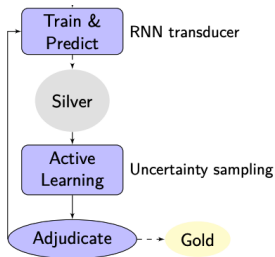
model training



evaluation

Collecting Data: Active Learning

- In-house doctor to annotate social media posts



- An **active learning** approach to reduce the number of annotations



annotated data



model training



evaluation

Model

- A pretrained language model
- Trained on medical entries from Wikipedia
- Finetuned on the medical annotations



annotated data

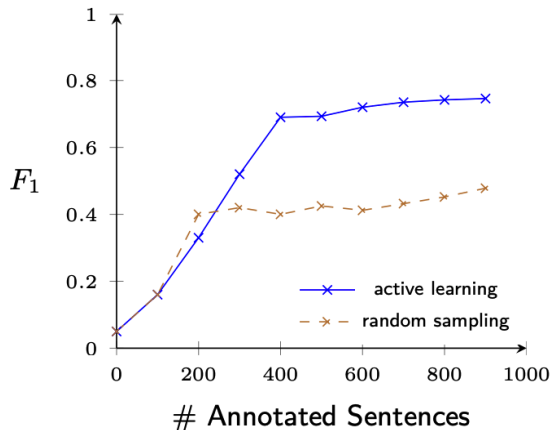


model training



evaluation

Evaluation



- Performance after 1hr of annotation: 74.2 F_1
- Uncertainty sampling boosts improvement rate

- Language modelling
- Pretraining
- Finetuning
- Prompting
- Active learning
- Crowdsourcing
- In-house annotation
- Inter-Annotator agreement

Conclusion

- NLP deals with understanding **real-world language usage**
- Lots of applications in the **medical domain**
 - Medical chatbots
 - Information extraction
(we've seen adverse drug reaction identification)
 - Summarizing doctor's notes
 - and more...
- **Language modelling** is a key technique
- The need for **domain experts** *in the loop* is crucial