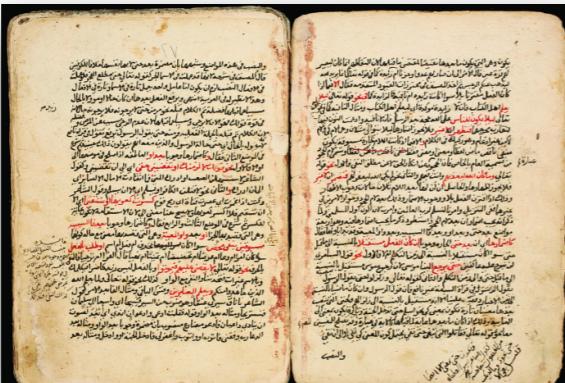


Automated Event Extraction for News-Based Counterdata

Katie Keith
Williams Statistics Colloquium

October 19, 2022

Age of abundant digitized texts



RETWEETS 10,496 FAVORITES

9:57 AM - 4 Sep 2010

燕子SXNSXS 8 min ago from 燕子铤碎了Android

这是出版成本下降的必然结果哈。

@中出宪政柏拉图:现在的出版物不是太少了,而是太多了,而且大部分都是垃圾,其中一小部分是垃圾中的战斗机,反而浪费读者生命。以后应该少出版一点儿,书贵精不贵多。

Repost Comments 1



GE 2019 third quarter performance

Financial results & Company highlights

October 30, 2019

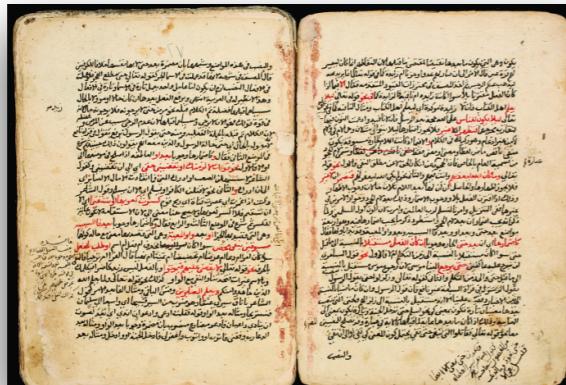
CAUTION CONCERNING FORWARD-LOOKING STATEMENTS:
This document contains forward-looking statements related to future events by the issuer or others that are, in different degrees, uncertain. For details on the uncertainties that may cause our actual future results to differ from those projected, see the cautionary statement in the section "Forward-Looking Statements" in the notes to the financial statements.

NON-GAAP FINANCIAL MEASURES:
The non-GAAP financial measures described below are presented in our financial statements prepared in accordance with U.S. generally accepted accounting principles (GAAP). Certain of these data are unadjusted from GAAP financial measures filed with the U.S. Securities and Exchange Commission (SEC). These non-GAAP financial measures are not included in the GAAP financial statement or otherwise in the SEC's filing. The non-GAAP financial measures are not intended to be a substitute for, or superior to, the GAAP financial measures.

GE's financial services business is conducted by GE Capital Consumer Finance LLC (GCCF). In its disclosure, we refer to GE Capital as "GE Capital". We refer to the industrial businesses of the Company including GE Capital as an entity under the heading "GE Industrial".

GE is a trademark of General Electric Company. All other trademarks and service marks used in this document are the property of their respective owners. GE Capital Consumer Finance LLC ("GCCF") and GE Capital are trademarks of General Electric Company.

Text data for social sciences questions



What drives newspapers' political slant?

Gentzkow and Shapiro,
Econometrica, 2010

The First Lady @FLOTUS

Happy birthday to the one and only, Queen 🎂! Thank you for being a role model for young girls around the world, @Beyonce. -mo

RETTWEETS 10,496 FAVORITES

9:57 AM - 4 Sep 2013

燕子SXSXSD 8 min ago from 燕子铤碎了Android

这是出版成本下降的必然结果哈。

@中出宪政柏拉图:现在的出版物不是太少了,而是太多了,而且大部分都是垃圾,其中一小部分是垃圾中的战斗机,反而浪费读者生命。以后应该少出版一点儿,贵精不贵多。

Repost Comments 1



What is the nature of online censorship in China?

King et al., *American Political Science Review*, 2013

GE 2019 third quarter performance

Financial results & Company highlights
October 30, 2019

CAUTION CONCERNING FORWARD-LOOKING STATEMENTS:
This document contains forward-looking statements made pursuant to the Private Securities Litigation Reform Act of 1995. Such statements are based on current expectations and involve risks and uncertainties. Actual results may differ significantly from those expressed or implied. This document also contains forward-looking financial information that is based on current estimates and forecasts.

NON-GAAP FINANCIAL MEASURES:
The non-GAAP financial measures described in this presentation are not presented in accordance with U.S. generally accepted accounting principles (GAAP). Certain of these non-GAAP financial measures require us to adjust financial information for items that are either not included in, or are excluded from, GAAP financial measures. We believe these non-GAAP financial measures provide useful information to investors about our operating results. These non-GAAP financial measures are not intended to be a substitute for, or superior to, GAAP financial measures, and should not be considered in isolation or as a substitute for GAAP financial measures.

GE's financial services business is conducted by GE Capital Global Finance LLC ("GE Capital"). In this document, we refer to "GE Capital" as "GE Capital". We refer to the industrial businesses of the Company including GE Capital as an "entity". We refer to the Company as "GE".

GE's Investor Relations website at www.ir.gecapital.com and its corporate blog at gecapitalblog.com, are not GE's Facebook page and Twitter account, contain a significant amount of information about GE, including financial and other information for investors. GE disclaims responsibility for these websites from time to time, as well as GE's Facebook and Twitter accounts.

Manual analysis is costly at scale

What drives newspapers' political slant?

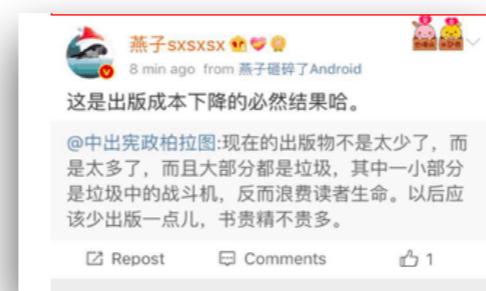
Gentzkow and Shapiro,
Econometrica, 2010



All articles for
400 news outlets

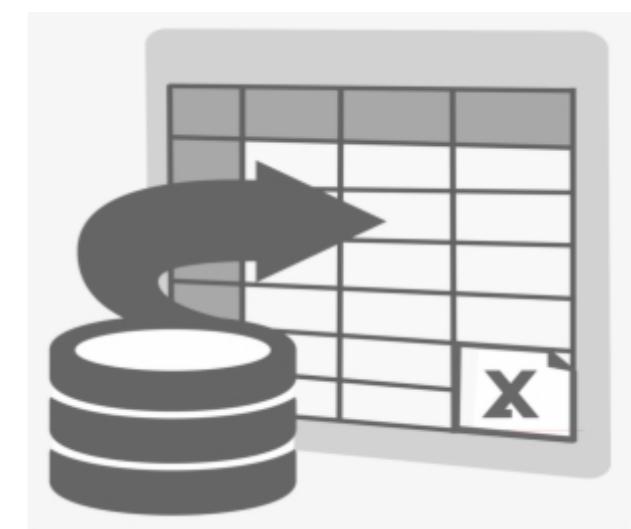
What is the nature of online censorship in China?

King et al., *American Political Science Review*, 2013



11 million posts

Natural language processing (NLP)



Focus of today's talk

**Corpus-Level Evaluation for Event QA:
The IndiaPoliceEvents Corpus Covering the 2002 Gujarat Violence**

Andrew Halterman*
Massachusetts Institute of Technology
ahalt@mit.edu

Katherine A. Keith*
University of Massachusetts Amherst
kkeith@cs.umass.edu

Sheikh Muhammad Sarwar*
University of Massachusetts Amherst
smsarwar@cs.umass.edu

Brendan O'Connor
University of Massachusetts Amherst
brenocon@cs.umass.edu

Abstract

Automated event extraction in social science applications often requires corpus-level evaluations: for example, aggregating text predictions across metadata and unbiased estimates of recall. We combine corpus-level evaluation requirements with a real-world, social science setting and introduce the INDIAPOLEVENTS corpus—all 21,391 sentences from 1,257 English-language *Times of India* articles about events in the state of Gujarat during March 2002. Our trained annotators read and label every document for mentions of police activity events, allowing for unbiased recall evaluations. In contrast to other datasets with structured event representations, we gather annotations by posing natural questions, and evaluate off-the-shelf models for three different tasks: sentence classification, document ranking, and temporal aggregation of target events. We present baseline results from zero-shot BERT-based models fine-tuned on natural language inference and passage retrieval tasks. Our novel corpus-level evaluations and annotation approach can guide creation of similar social-science-oriented resources in the future.

1 Introduction

Understanding the actions taken by political actors is at the heart of political science research: How do actors respond to contested elections (Daxencker et al., 2019)? How many people attend protests (Chenoweth and Lewis, 2013)? Which religious groups are engaged in violence (Brathwaite and Park, 2018)? Why do some governments try to prevent anti-minority riots while others do not (Wilkinson, 2006)? In the absence of official records, social scientists often turn to news data to extract the actions of actors and surrounding events. These

* Indicates joint first-authorship.

4240

Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, pages 4240–4253
August 1–6, 2021. ©2021 Association for Computational Linguistics

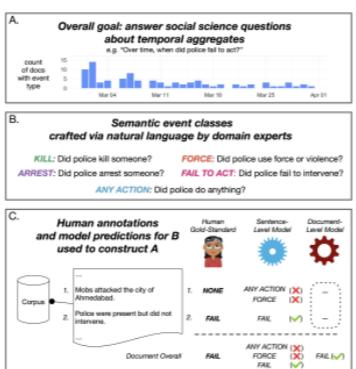


Figure 1: Motivation (A-B) and procedures (B-C) for this paper: A. Social scientists often use text data to answer substantive questions about temporal aggregates. B. To answer these questions, domain experts use natural language to define semantic event classes of interest. C. Our INDIAPOLEVENTS dataset: Humans annotate *every* sentence in the corpus in order to evaluate whether a system achieves full recall of relevant events. In production, computational models run B's queries to classify or rank sentences or documents, which are aggregated to answer A.



Andy Halterman
Political Science



Katie Keith
Computer Science



Sheikh Sarwar
Computer Science



Brendan O'Connor
Computer Science

Political science-motivated research questions



Andy Halterman
Political Science

1.

Q: Does variation in party control of state government affect whether police failed to intervene in communal violence?

Violence in Gujarat, India 2002



Train fire kills Hindu Pilgrims, Feb. 27, 2002
Photo Credit: New York Times

2.

Challenges

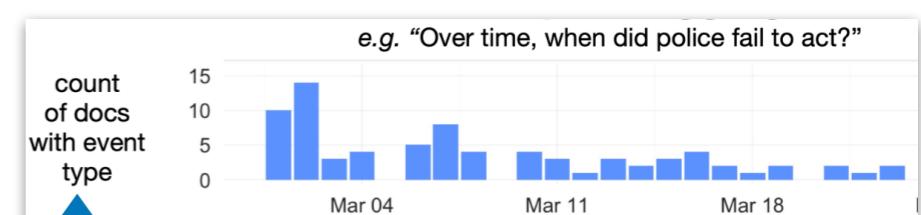
- No official records.
- Only news articles.
- Reading documents manually is costly.

3.

Use NLP to automate extracting events

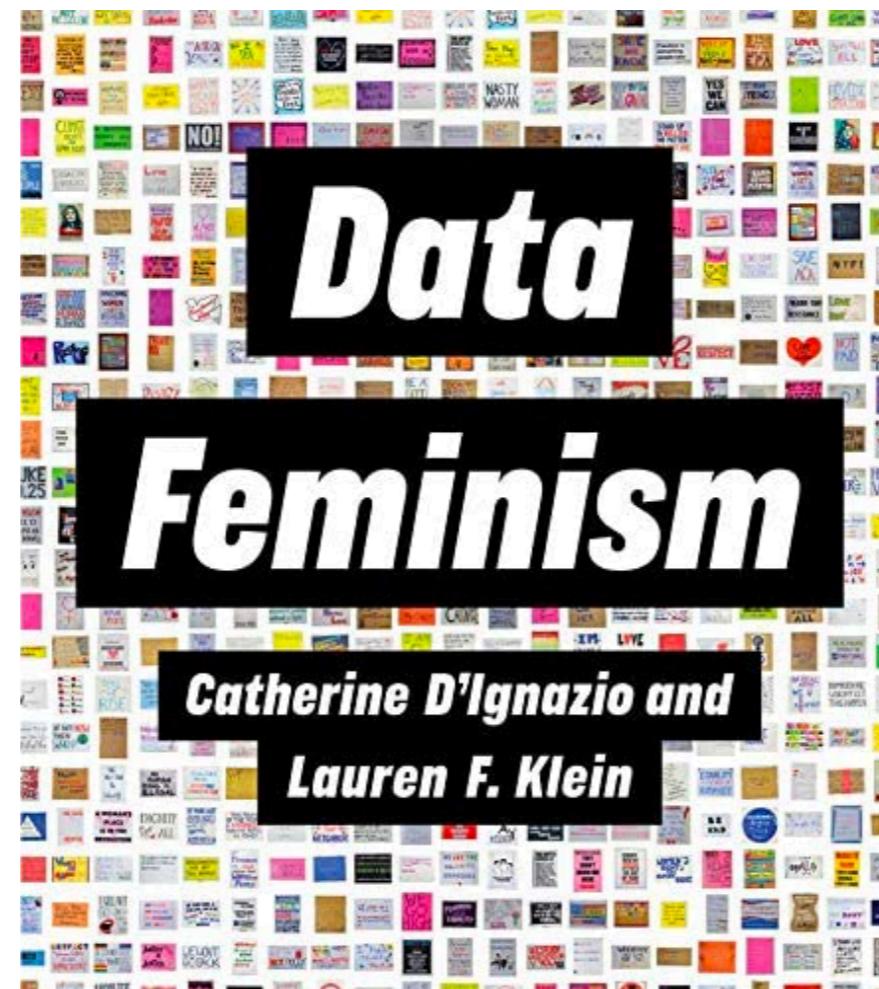


count of docs with event type



Media bias outside the scope of this talk

**Counterdata is the grassroots collection
of missing datasets**



Events

Who did what to whom?

Police killed [PERSON].

Even simple event types present challenges

Police killed PERSON.

Police officers spotted the butt of a handgun in Alton Sterling's front pocket and saw him reach for the weapon before opening fire, according to a Baton Rouge Police Department search warrant filed Monday that offers the first police account of the events leading up to **his fatal shooting**.

Keith et al. Identifying civilians killed by police with distantly supervised entity-event extraction. EMNLP, 2017.

Even simple event types present challenges

Police killed PERSON.

long-range dependencies

Police officers spotted the butt of a handgun in **Alton Sterling**'s front pocket and saw him reach for the weapon before **opening fire**, according to a Baton Rouge Police Department search warrant filed Monday that offers the first police account of the events leading up to **his fatal shooting**.

Even simple event types present challenges

Police killed PERSON.

long-range dependencies

Police officers spotted the butt of a handgun in Alton Sterling's front pocket and saw him reach for the weapon before opening fire, according to a Baton Rouge Police Department search warrant filed Monday that offers the first police account of the events leading up to his fatal shooting.

coreference

Even simple event types present challenges

Police killed PERSON.

long-range dependencies

Police officers spotted the butt of a handgun in Alton Sterling's front pocket and saw him reach for the weapon before opening fire, according to a Baton Rouge Police Department search warrant filed Monday that offers the first police account of the events leading up to his fatal shooting.

coreference

*event
coreference*

Events

Who did what to whom?

Automated event extraction has a large academic literature...

in the social sciences

Schrodt et al., 1994; King and Lowe, 2003; Hanna, 2014; Hammond and Weidmann, 2014; Boschee et al., 2015; Beieler et al., 2016; Osorio and Reyes, 2017

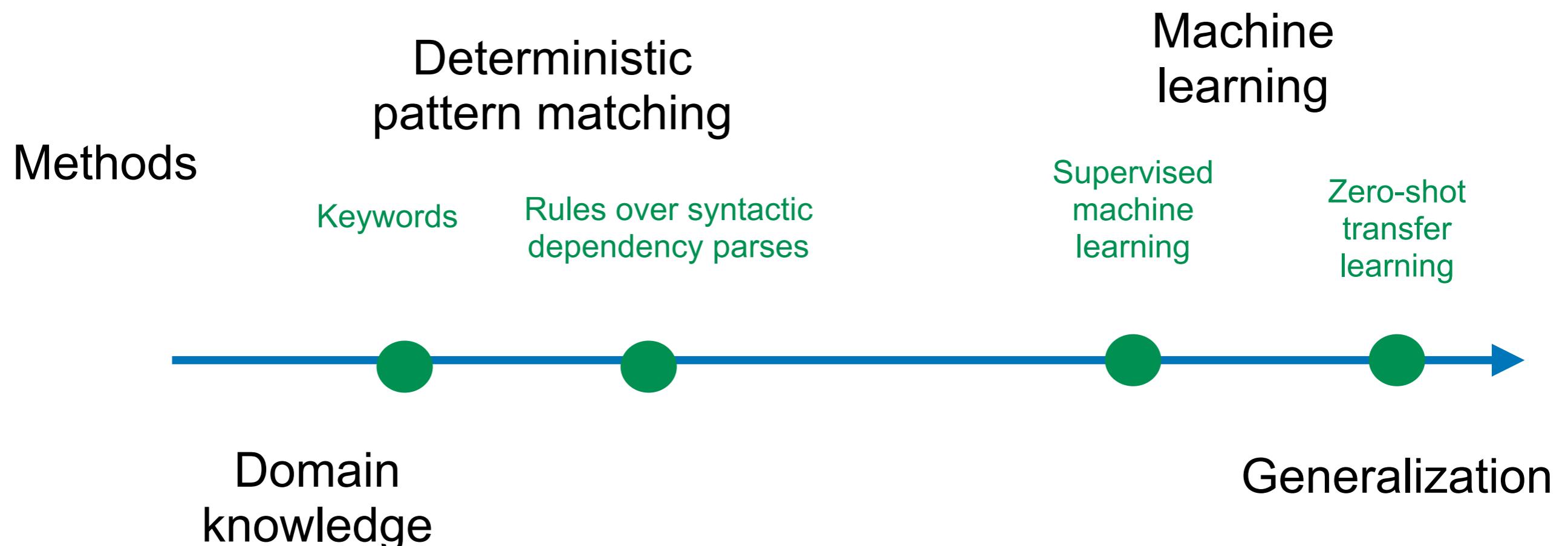
Hovy et al. Events are Not Simple: Identity, Non-Identity, and Quasi-Identity. Workshop on EVENTS, 2013.

Abend and Rapport. The State of the Art in Semantic Representation. ACL, 2017.

in computer science

Grishman, 1997; McCallum, 2005; Aguilar et al., 2014; Hovy et al., 2013; Levy et al., 2017; Abend and Rappoport, 2017; Grishman, 2019; Liu et al., 2020; Du and Cardie, 2020

Approaches to Automated Event Extraction



Mitchell. The Need for Biases in Learning Generalizations. 1980.

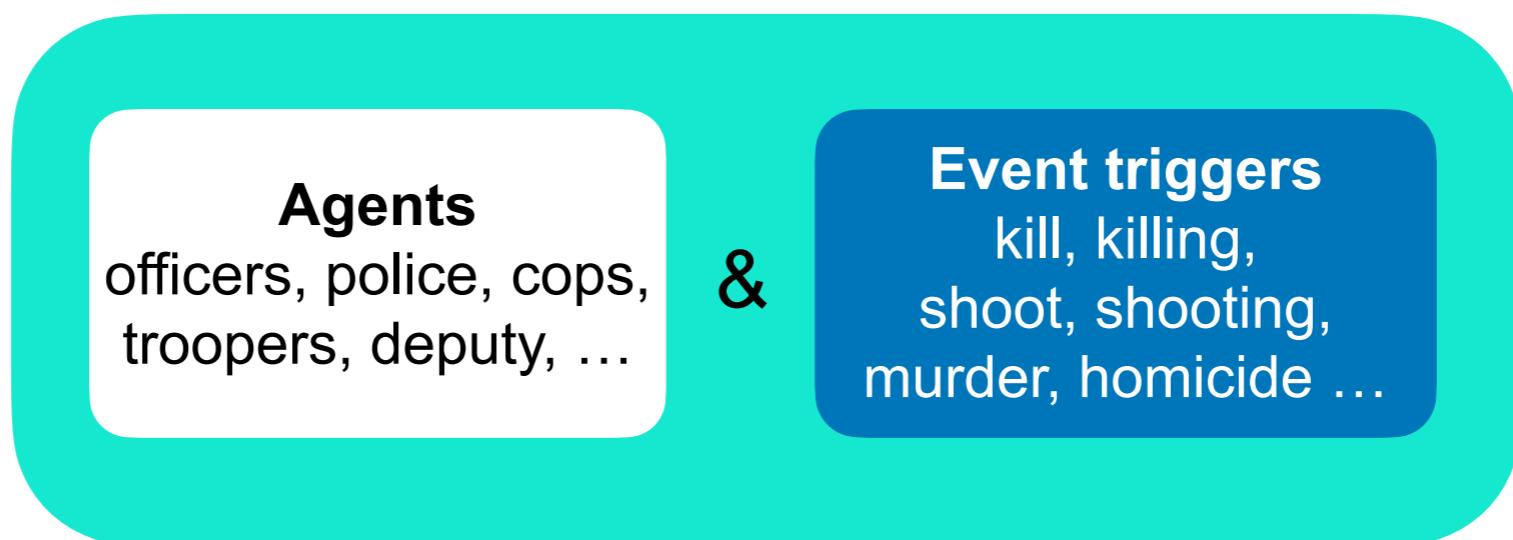
Deterministic Keyword Matching

Input: sentences

*PERSON was **fatally shot** by **police**.*

***Officers** reported **PERSON** was **killed** in a car accident.*

Method:
Keyword matching



Output: Classification

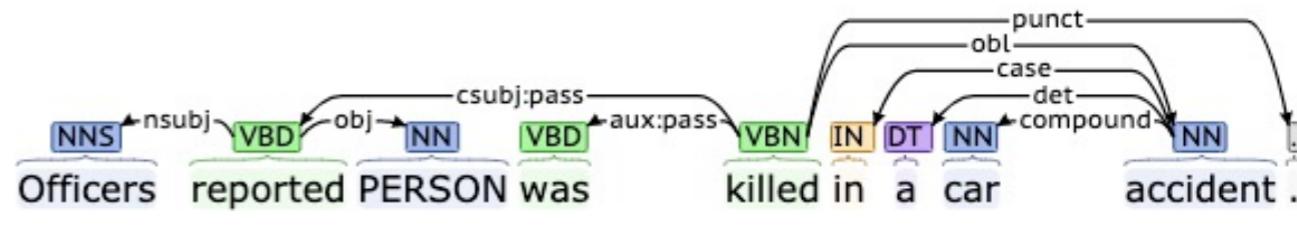
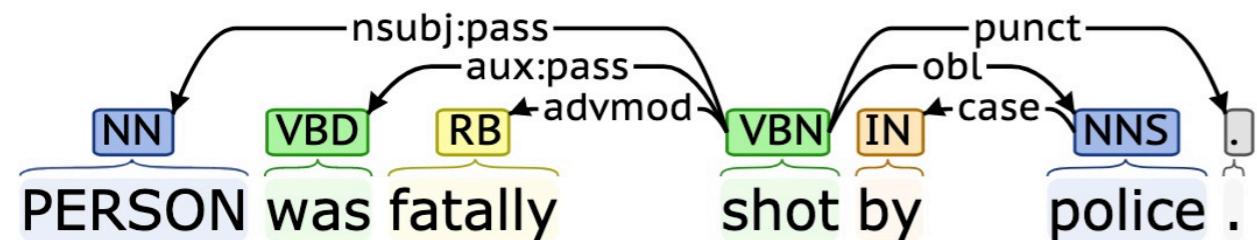
Yes

~~Yes~~

Issue: many
false positives
(low precision)

Deterministic Syntax Matching

Input: automatically infer dependency parse trees over sentences



Method: Rules over dependency paths

PERSON <-nsubj:pass <-

kill, killing,
shoot, shooting,
murder, homicide ...

->obl ->

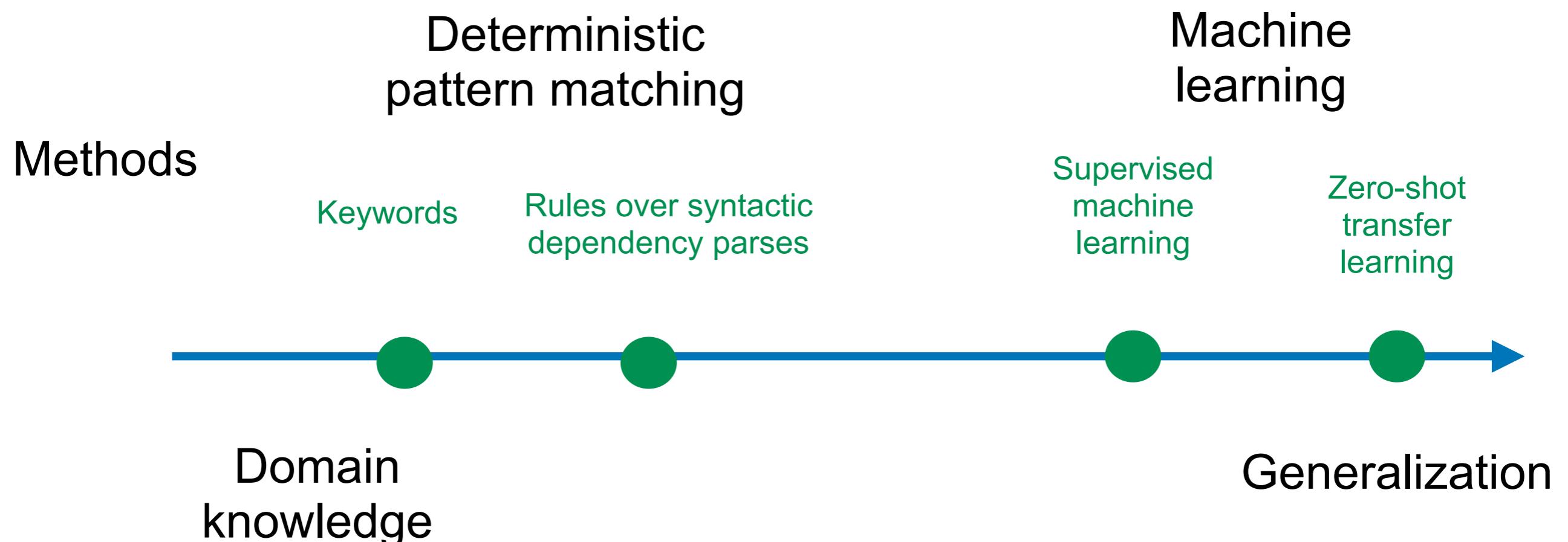
officers, police,
cops, troopers,
deputy, ...

Output: Classification

Yes

Issue: Difficult for
a domain expert to
list all possible
rules (*low recall*)

Approaches to Automated Event Extraction



Mitchell. The Need for Biases in Learning Generalizations. 1980.

Supervised Machine Learning

1. Gather training data

Police killed PERSON.



2. Humans label training data

Issue: Costly

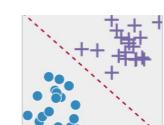
3. Train model: statistical pattern matching between inputs and labels

Our 2017 work:

- *logistic regression* with bag of words features
- *convolutional neural networks* initialized with pre-trained word embeddings

4. Inference: (generalization) apply trained model on unseen inputs

PERSON died in a police homicide.



Yes

AI Hype

The New York Times

A Learning Advance in Artificial Intelligence Rivals Human Abilities

Give this article

Human or Machine?

Humans and machines were given an image of a novel character (represented atop each grid) and then asked to copy it. Brenden Lake

By John Markoff
Dec. 10, 2015

Latest Issues

SCIENTIFIC AMERICAN

Cart 0 Sign In | Newsletters

navirus Health Mind & Brain Environment Technology Space & Physics Video Podcasts Opinion Store Q

Unlimited Knowledge Awaits. [Subscribe](#)

ARTIFICIAL INTELLIGENCE

Google Engineer Claims AI Chatbot Is Sentient: Why That Matters

Is it possible for an artificial intelligence to be sentient?

By Leonardo De Cosmo on July 12, 2022

Credit: Boris SV/Getty Images

READ THIS NEXT

SPONSORED
The Custom Coatings that Transformed High-Tech Industry

CONSCIOUSNESS
We Shouldn't Try to Make Conscious Software--Until We Should
Jim Davies | Opinion

ENGINEERING
The Search for a New Test of Artificial Intelligence
Gary Marcus

Pre-training with large-scale language models

Huge performance gains in recent years with **large-scale language models** trained on scrapes of the web

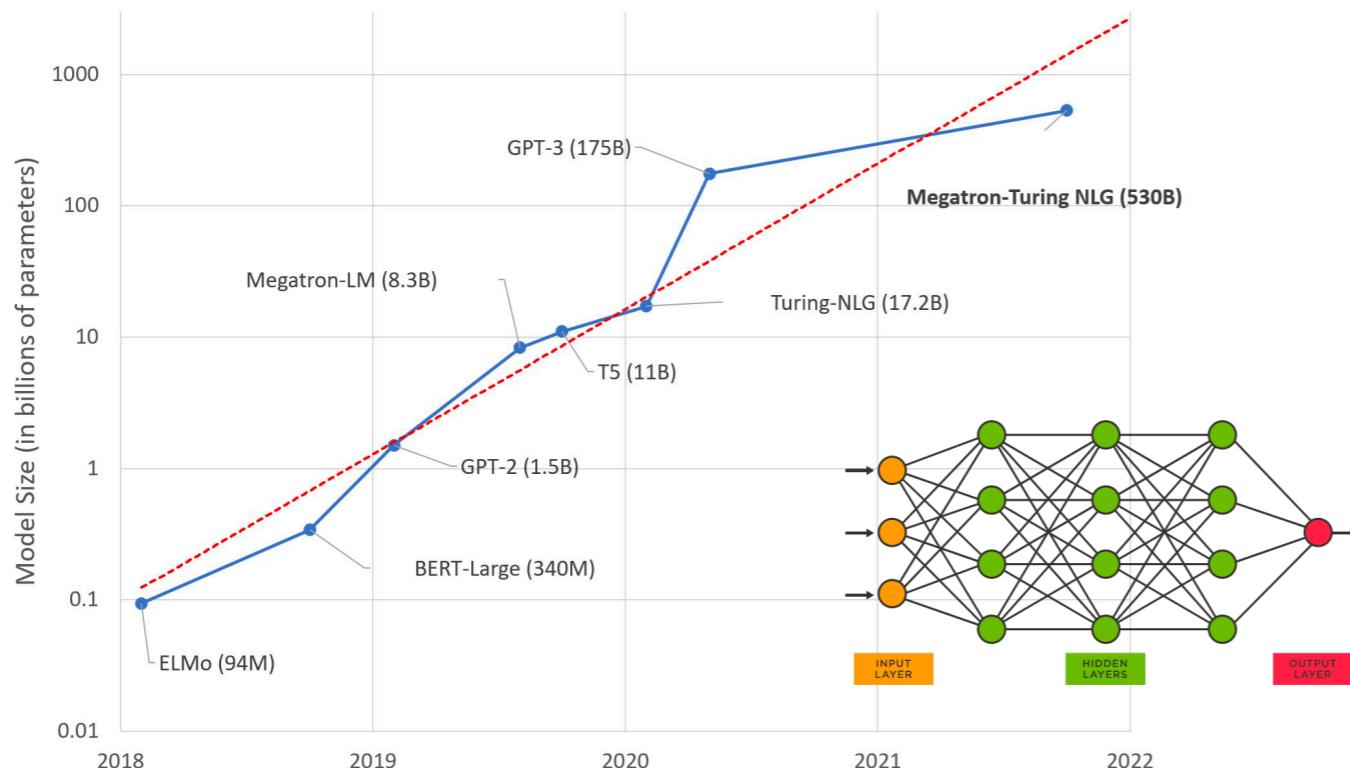


Figure credit: Hugging Face

Self-supervision: Masked language modeling (MLM) objective function

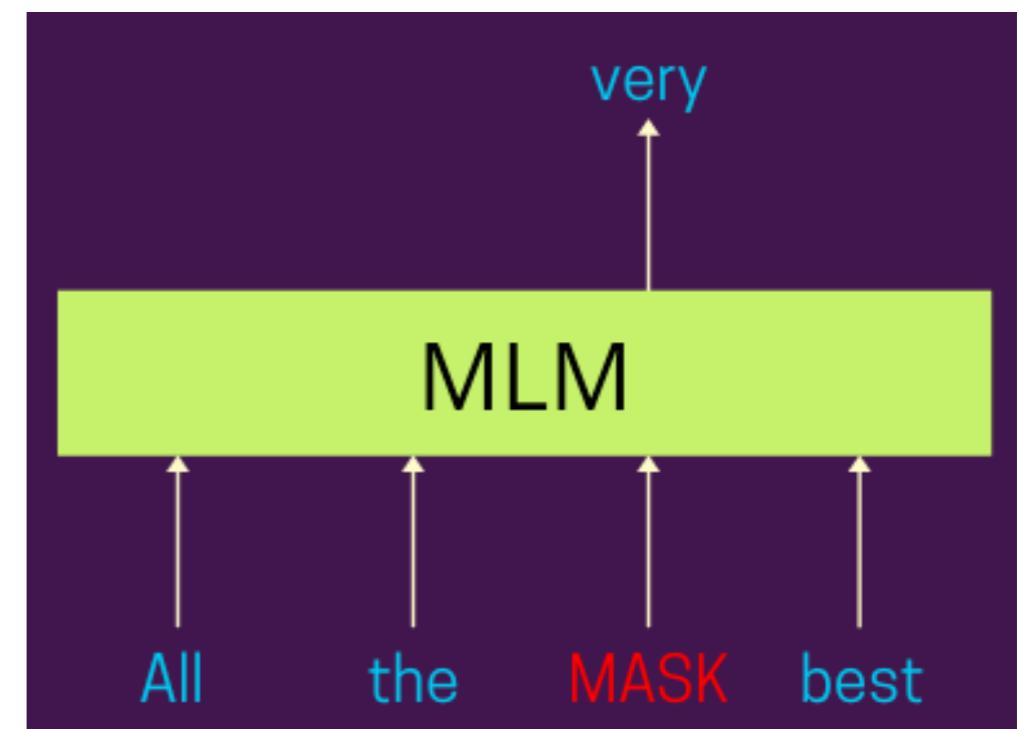


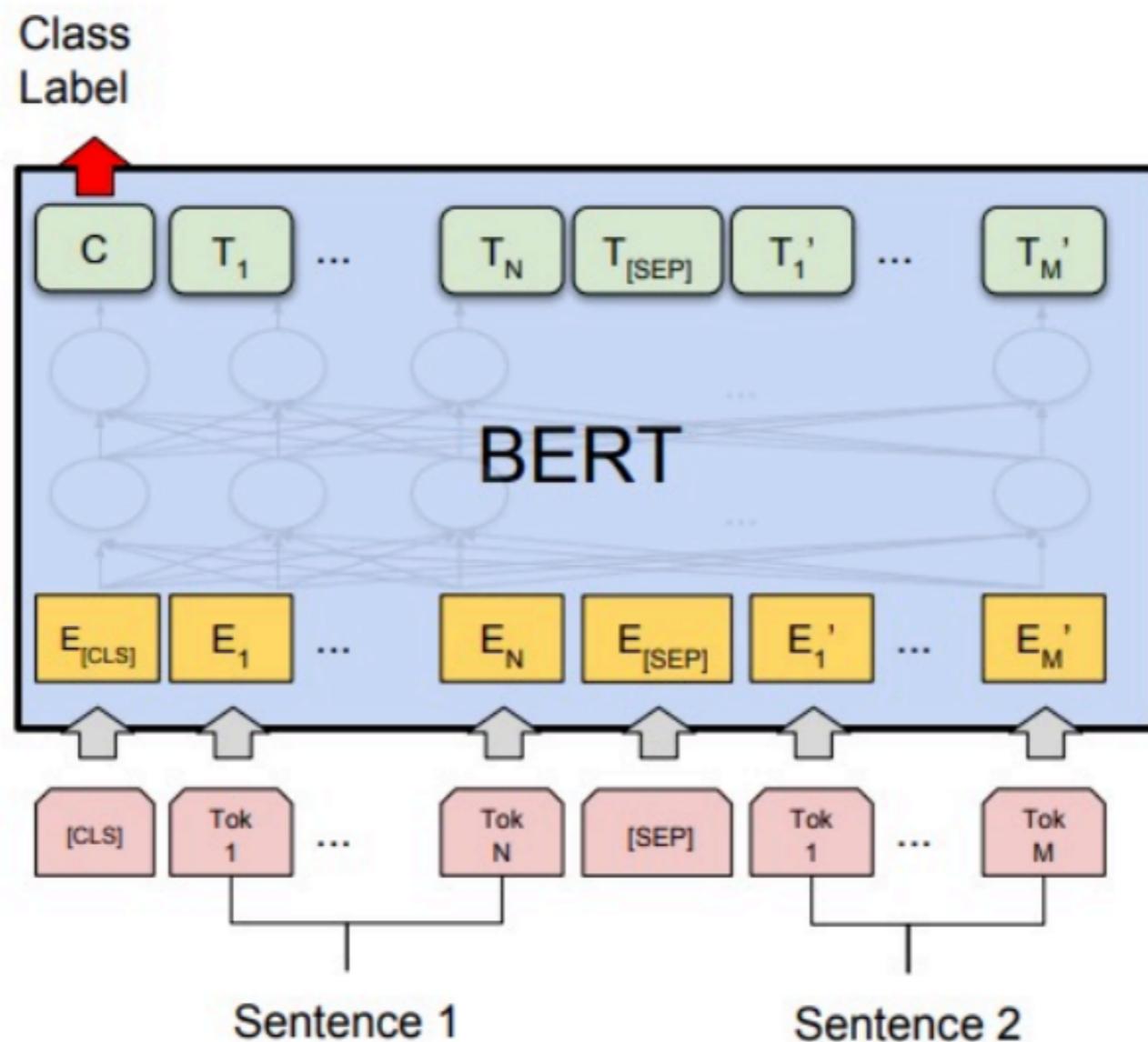
Figure credit: Prakhar Mishra, [blog](#)

Zero-Shot Transfer Learning

1. **Pre-train** large-scale language model
2. **Fine-tune** on a task with labeled data
3. Apply trained model **zero-shot** to our dataset

2. Fine-tuning on a task with labeled data

Entailment
Neutral
Contradiction



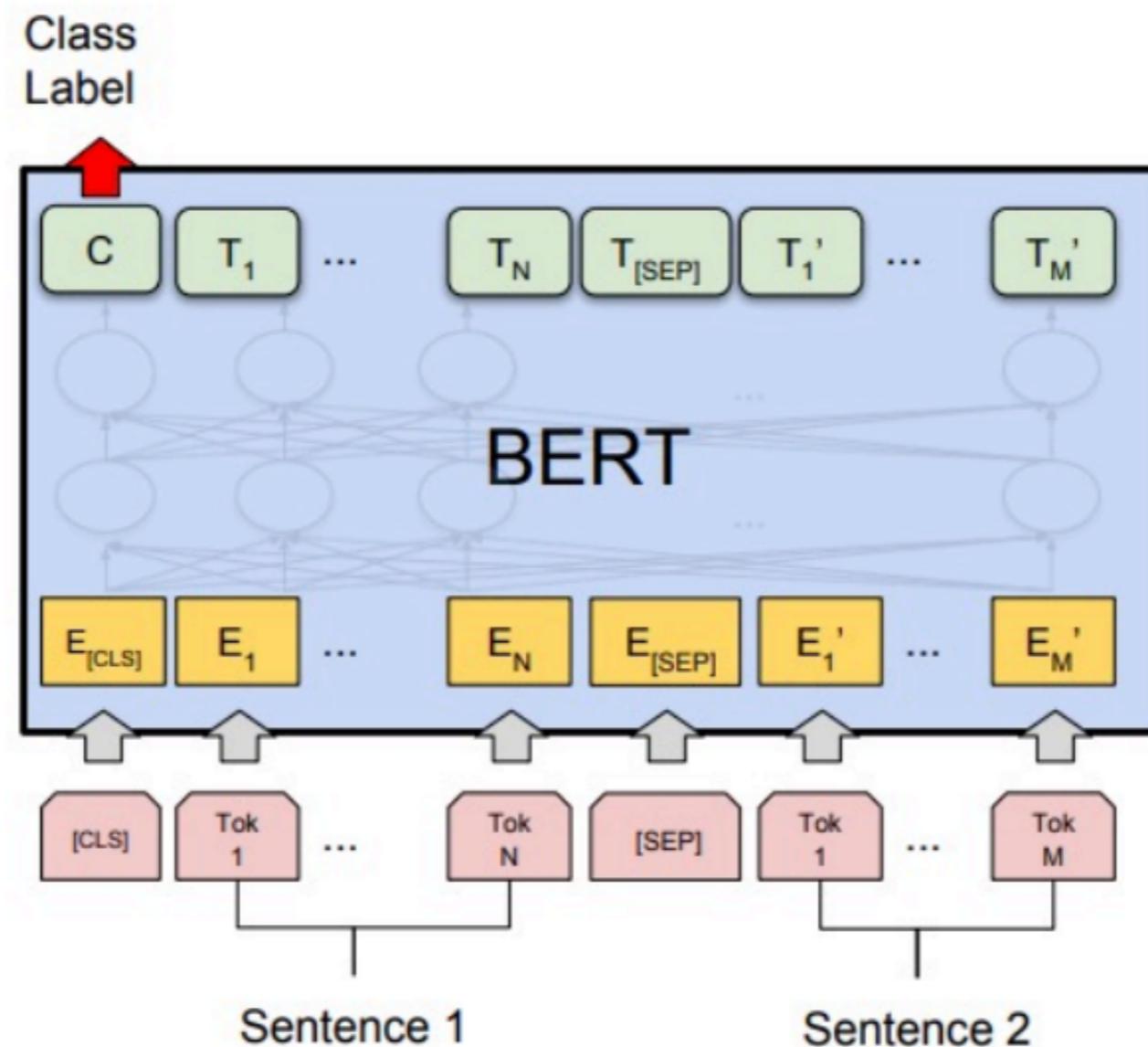
A soccer game with
multiple males playing.

Some men are playing
a sport.

Bowman et al. ACL, 2015

3. Apply trained model **zero-shot** to our dataset

Entailment
Neutral
Contradiction



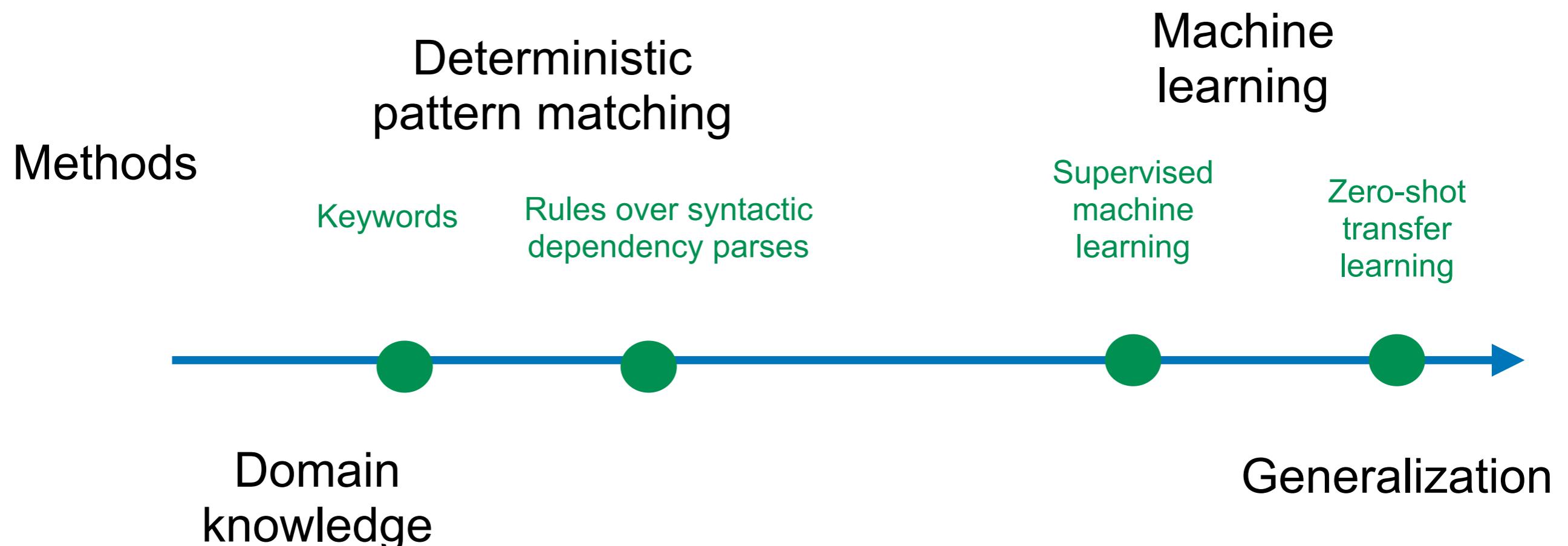
Prompt

Police killed
someone.

Sentence from
our dataset

Yesterday, 97 died in
police firing.

Approaches to Automated Event Extraction



Mitchell. The Need for Biases in Learning Generalizations. 1980.

What was our original problem again?

International-relations motivated research questions



Andy Halterman
Political Science

1.

Q: Does variation in party control of state government affect whether police failed to intervene in communal violence?

2.

Challenges

- No official records.
- Only news articles.
- Reading documents manually is costly.

3.

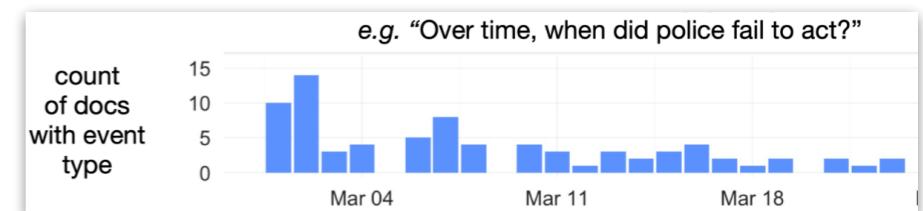
Use NLP to
automate
extracting events



Violence in Gujarat, India 2002



Train fire kills Hindu Pilgrims, Feb. 27, 2002
Photo Credit: New York Times



Novel dataset created for empirical evaluation



- *Times of India*
- Filter to March 2002 and “Ayodha” OR “Gujarat”
- **Results in 1,257 articles, 21,391 sentences**
- Every sentence annotated with 2 annotators + adjudication round

Annotation interface

On Sunday, a mob gathered carrying swords, hockey sticks and other weapons. In response, the police rushed to the spot to quell the violence and arrested ten people. **Two people died due to police firing and another three were injured from the shooting.** An officer was detained due to unethical conduct.

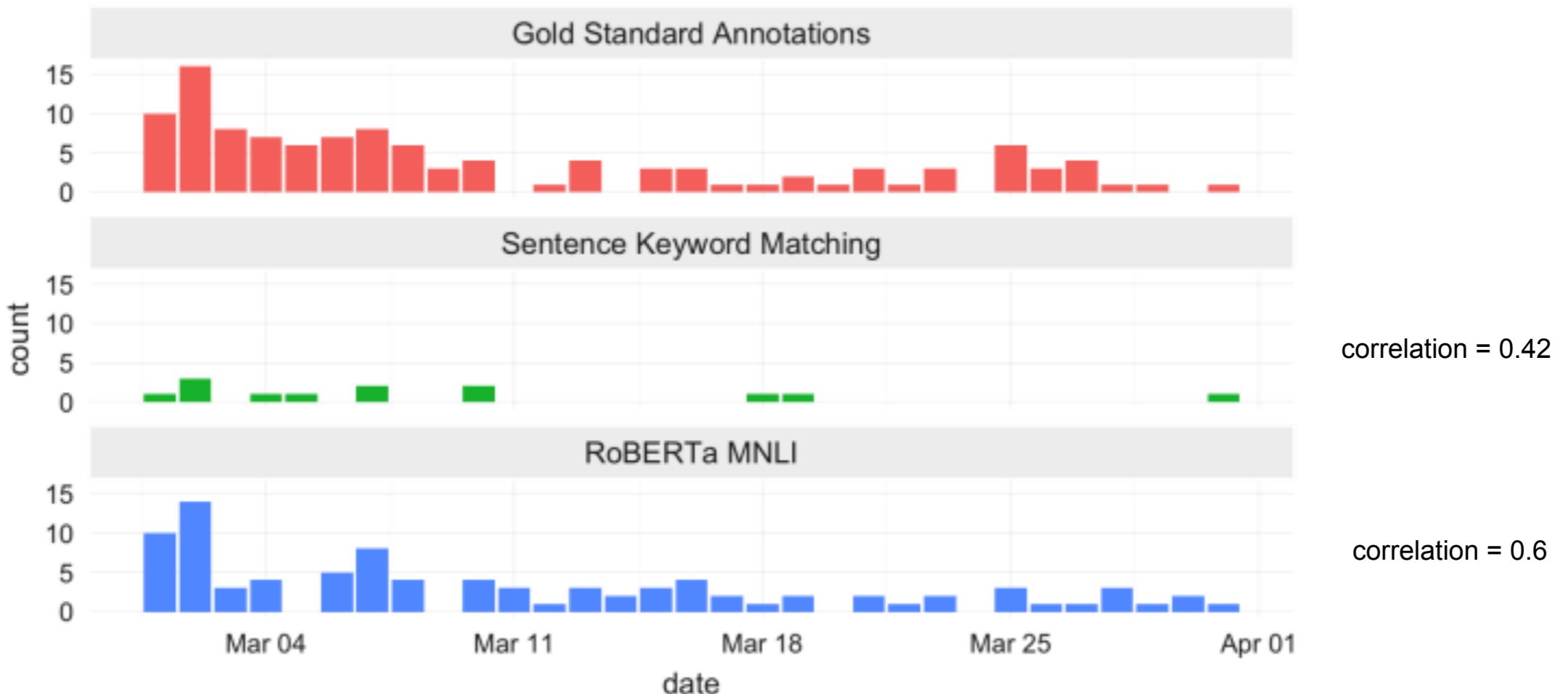
<input checked="" type="checkbox"/> Did police kill someone?	1
<input type="checkbox"/> Did police arrest someone?	2
<input type="checkbox"/> Did police fail to act or not intervene?	3
<input checked="" type="checkbox"/> Did police use other force or violence?	4
<input type="checkbox"/> Did police say or do something else (not included above)?	5

Dataset publicly available

<https://github.com/slanglab/IndiaPoliceEvents>

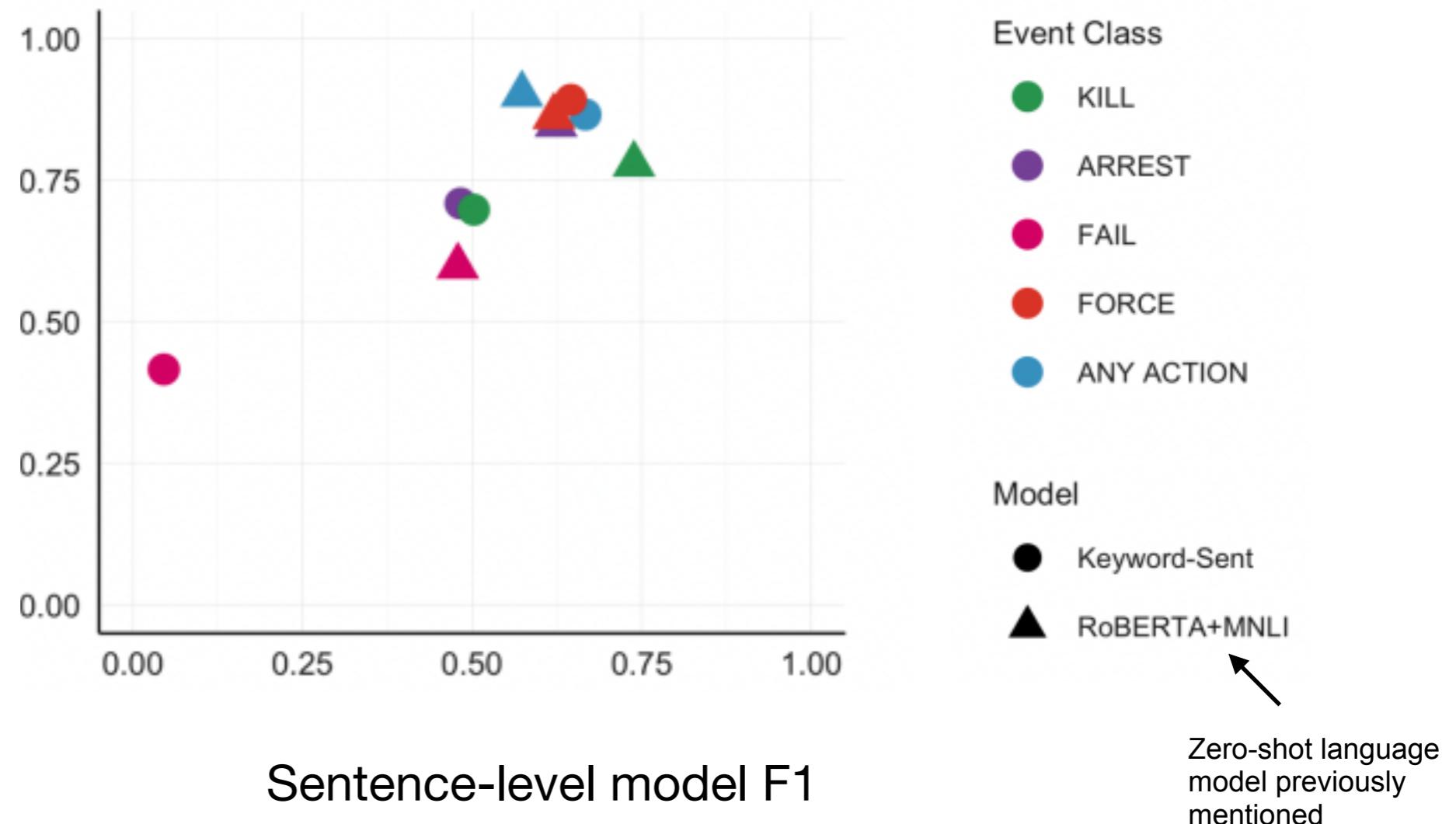
Evaluation highlights

FAIL
Zero-shot language model previously mentioned



Evaluation highlights

Temporal aggregates:
correlation between
human gold-standard and model



Manual error analysis



Model assigns high positive probability to sentences that should be classified as **negatives**

“ [...] scores of people have been killed in rural Gujarat due to police failure...

“Police said that two persons had been killed [...]”

Please read our paper for more details!

**Corpus-Level Evaluation for Event QA:
The IndiaPoliceEvents Corpus Covering the 2002 Gujarat Violence**

Andrew Halterman*
Massachusetts Institute of Technology
ahalt@mit.edu

Katherine A. Keith*
University of Massachusetts Amherst
kkeith@cs.umass.edu

Sheikh Muhammad Sarwar*
University of Massachusetts Amherst
smsarwar@cs.umass.edu

Brendan O'Connor
University of Massachusetts Amherst
brenocon@cs.umass.edu

Abstract

Automated event extraction in social science applications often requires corpus-level evaluations: for example, aggregating text predictions across metadata and unbiased estimates of recall. We combine corpus-level evaluation requirements with a real-world, social science setting and introduce the INDIAPOLEVENTS corpus—all 21,391 sentences from 1,257 English-language *Times of India* articles about events in the state of Gujarat during March 2002. Our trained annotators read and label every document for mentions of police activity events, allowing for unbiased recall evaluations. In contrast to other datasets with structured event representations, we gather annotations by posing natural questions, and evaluate off-the-shelf models for three different tasks: sentence classification, document ranking, and temporal aggregation of target events. We present baseline results from zero-shot BERT-based models fine-tuned on natural language inference and passage retrieval tasks. Our novel corpus-level evaluations and annotation approach can guide creation of similar social-science-oriented resources in the future.

1 Introduction

Understanding the actions taken by political actors is at the heart of political science research: How do actors respond to contested elections (Daxencker et al., 2019)? How many people attend protests (Chenoweth and Lewis, 2013)? Which religious groups are engaged in violence (Brathwaite and Park, 2018)? Why do some governments try to prevent anti-minority riots while others do not (Wilkinson, 2006)? In the absence of official records, social scientists often turn to news data to extract the actions of actors and surrounding events. These

* Indicates joint first-authorship.

4240

Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, pages 4240–4253
August 1–6, 2021. ©2021 Association for Computational Linguistics

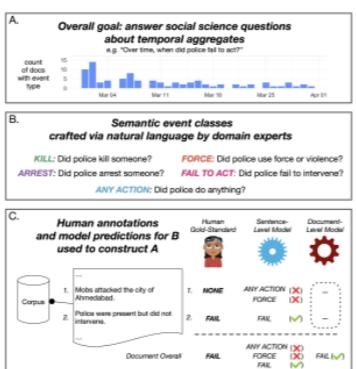


Figure 1: Motivation (A-B) and procedures (B-C) for this paper: A. Social scientists often use text data to answer substantive questions about temporal aggregates. B. To answer these questions, domain experts use natural language to define semantic event classes of interest. C. Our INDIAPOLEVENTS dataset: Humans annotate *every* sentence in the corpus in order to evaluate whether a system achieves full recall of relevant events. In production, computational models run B's queries to classify or rank sentences or documents, which are aggregated to answer A.



Andy Halterman
Political Science



Katie Keith
Computer Science



Sheikh Sarwar
Computer Science



Brendan O'Connor
Computer Science

Two shameless plugs

Williams

Search Menu

CATALOG Archive Registrar Student Records Syllabi Winter Study

Williams » Catalog » Courses and Programs 2022-23 » Computer Science » Computer Science Spring 2022-23 Class List » CSCI 375 Natural Language Processing 2022-23

CSCI 375
Natural Language Processing
Spring 2023

Format: lecture
Limit: 24
Expected: 24
Class#: 3296
Grading: no pass/fail option, no fifth course option

Requirements/Evaluation: Evaluation based on assignments, projects, and exams.

Prerequisites: CSCI 136, and either CSCI 256 or STAT 201/202.

Enrollment Preferences: current or expected Computer Science majors.

Distributions: Division III
Quantitative/Formal Reasoning

QFR Notes: The course will consist of programming assignments and problem sets in which quantitative/formal reasoning skills are practiced and evaluated.

[View Book Information](#)

Updated 9:31 AM

CLASSES	DREQ	INSTRUCTORS	TIMES	CLASS#	ENROLL	CONSENT
CSCI 375 - 01 (S) LEC Natural Language Processing	III Q	Katie A. Keith	MWF 10:00 am - 10:50 am	3296	Open	None
CSCI 375 - 02 (S) LEC Natural Language Processing	III Q	Katie A. Keith	MWF 11:00 am - 11:50 am	3297	Open	None

Diaries of Social Data Research
By Katherine A. Keith & Lucy Li

Large-scale data has become a major component of research about human behavior and society. But how are interdisciplinary collaborations that use large-scale social data formed and maintained? What obstacles are encountered on the journey from idea conception to publication? In this podcast, we investigate these questions by probing the "research diaries" of scholars in computational social science and adjacent fields. We unmask the research process with the hope of normalizing the challenges of and increasing

[Listen on Spotify](#) [Message](#) [Website](#)

WHERE TO LISTEN

Thanks!

Collaborators



Kaggle Data Science
Research Grant

kaggle

Bloomberg Data Science
PhD Fellowship

