NPTEL ONLINE CERTIFICATION COURSES

DEEP LEARNING FOR NATURAL LANGUAGE PROCESSING

Lecture 05 : NLP Tasks and Paradigms

PROF. PAWAN GOYAL

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

CONCEPTS COVERED

- Paradigms in NLP
- Text Classification, Sequence Labeling, Text Generation, Structured Prediction
- Some NLP Tasks

NLP Paradigms

We generally try to map NLP problems to various (ML) paradigms

- Sentiment Analysis, news article groupings, etc. → **Text Classification**
- Named entity recognition, code-mixing, etc. → **Sequence Labeling**
- Machine Translation, summarization, chatbots, etc. → **Text Generation**

Other popular paradigm: **Structured Prediction**

Example of NLP tasks

Word / Span Level: Word sense disambiguation, Entity Linking
Sentence Level: Sentence Similarity, Natural Language Inference
Paragraph / Document Level: Question Answering

NLP Paradigms: Classification

Classification : Positive or negative review?

\+ *...zany characters and richly applied satire, and some great plot twists*

− *It was pathetic. The worst part about it was the boxing scenes...*

\+ *...awesome caramel sauce and sweet toasty almonds. I love this place!*

− *...awful pizza and ridiculously overpriced...*

https://web.stanford.edu/~jurafsky/slp3/

Why sentiment analysis?

Movie: is this review positive or negative?

Products: what do people think about the new iPhone?

Politics: what do people think about this candidate or issue?

Prediction: predict election outcomes or market trends from sentiment

https://web.stanford.edu/~jurafsky/slp3/

Text Classification: formal definition

Input:

- a document d
- a fixed set of classes $C = \{c_1, c_2, \ldots, c_J\}$

Output: a predicted class $c \in C$

Classification Methods: Machine Learning

Input:

- a document d
- a fixed set of classes $C = \{c_1, c_2, \ldots, c_J\}$
- A training set of m hand-labeled documents $(d_1,c_1),\ldots,(d_m,c_m)$

Output:

- a learned classifier $\gamma:d \rightarrow c$

Any kind of classifier

- Naïve Bayes
- Support Vector Machines
- Neural networks
- k-Nearest Neighbors
- …

https://web.stanford.edu/~jurafsky/slp3/

Evaluation for Text Classification

Let's consider just binary text classification tasks

Imagine you're the CEO of Delicious Pie Company

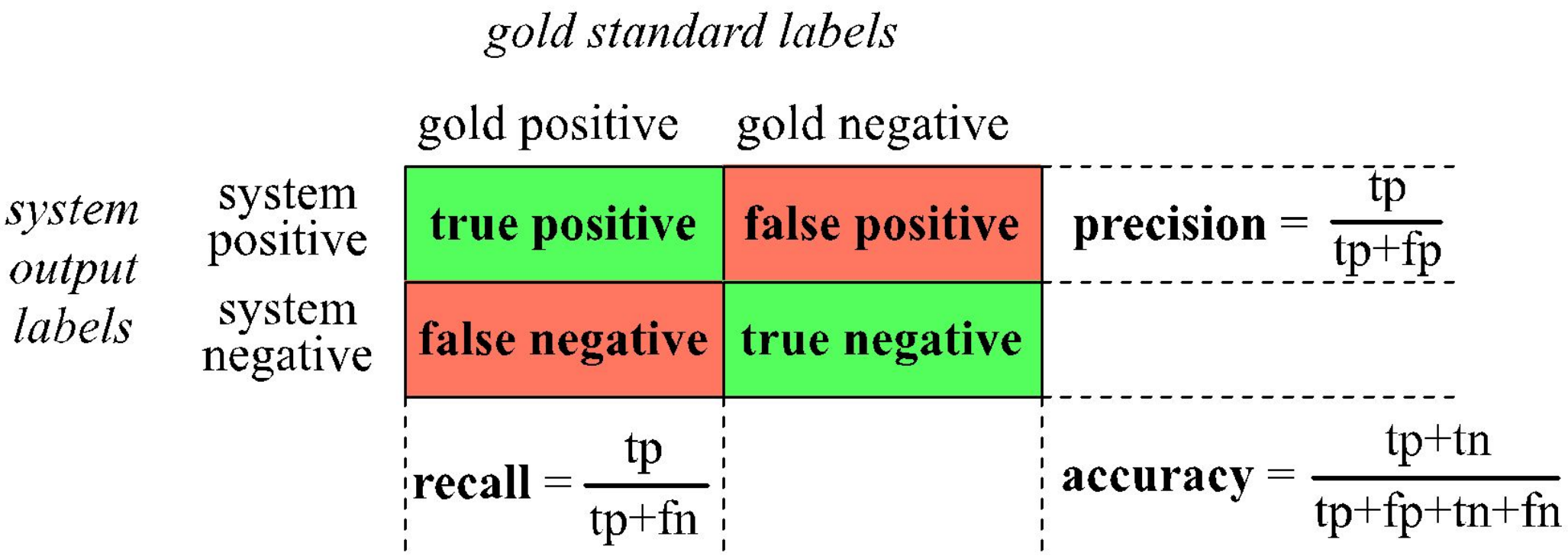
You want to know what people are saying about your pies

So you build a "Delicious Pie" tweet detector

- Positive class: tweets about Delicious Pie Co
- Negative class: all other tweets

https://web.stanford.edu/~jurafsky/slp3/

The 2-by-2 confusion matrix



https://web.stanford.edu/~jurafsky/slp3/

Evaluation: Accuracy

*Why don't we use **accuracy** as our metric?*

Imagine we saw 1 million tweets

- 100 of them talked about Delicious Pie Co.
- 999,900 talked about something else

We could build a dumb classifier that just labels every tweet "not about pie"

- It would get 99.99% accuracy!!! Wow!!!!
- But useless! Doesn't return the comments we are looking for!
- That's why we use **precision** and **recall** instead

https://web.stanford.edu/~jurafsky/slp3/

Why Precision and recall

Our dumb pie-classifier

- Just label nothing as "about pie"

Accuracy=99.99%

but

Recall = 0

- (it doesn't get any of the 100 Pie tweets)

Precision and recall, unlike accuracy, emphasize true positives:

- finding the things that we are supposed to be looking for.

https://web.stanford.edu/~jurafsky/slp3/

A combined measure: F

F measure: a single number that combines P and R:

$$F_\beta = \frac{(\beta^2 + 1)PR}{\beta^2 P + R}$$

We almost always use balanced F_1 (i.e., $\beta = 1$)

$$F_1 = \frac{2PR}{P + R}$$

https://web.stanford.edu/~jurafsky/slp3/

Confusion Matrix for 3-class classification

system output \ *gold labels*	urgent	normal	spam	
urgent	8	10	1	$\textbf{precision}_u = \frac{8}{8+10+1}$
normal	5	60	50	$\textbf{precision}_n = \frac{60}{5+60+50}$
spam	3	30	200	$\textbf{precision}_s = \frac{200}{3+30+200}$
	$\textbf{recall}_u = \frac{8}{8+5+3}$	$\textbf{recall}_n = \frac{60}{10+60+30}$	$\textbf{recall}_s = \frac{200}{1+50+200}$	

https://web.stanford.edu/~jurafsky/slp3/

How to combine P/R from 3 classes to get one metric

Macro-averaging:

- compute the performance for each class, and then average over classes

Micro-averaging:

- collect decisions for all classes into one confusion matrix
- compute precision and recall from that table.

https://web.stanford.edu/~jurafsky/slp3/

Macro-averaging and Micro-averaging

Class 1: Urgent

	true urgent	true not
system urgent	8	11
system not	8	340

Class 2: Normal

	true normal	true not
system normal	60	55
system not	40	212

Class 3: Spam

	true spam	true not
system spam	200	33
system not	51	83

Pooled

	true yes	true no
system yes	268	99
system no	99	635

$$\text{precision} = \frac{8}{8+11} = .42 \qquad \text{precision} = \frac{60}{60+55} = .52 \qquad \text{precision} = \frac{200}{200+33} = .86$$

$$\text{microaverage precision} = \frac{268}{268+99} = \mathbf{.73}$$

$$\text{macroaverage precision} = \frac{.42+.52+.86}{3} = \mathbf{.60}$$

https://web.stanford.edu/~jurafsky/slp3/

NLP Paradigms: Sequence Labeling

Parts-of-Speech Tagging

Sequence Labeling: Parts of Speech

From the earliest linguistic traditions (Yaska and Panini 5th C. BCE, Aristotle 4th C. BCE), the idea that words can be classified into grammatical categories

- part of speech, word classes, POS, POS tags

8 parts of speech attributed to Dionysius Thrax of Alexandria (c. 1st C. BCE):

- noun, verb, pronoun, preposition, adverb, conjunction, participle, article

https://web.stanford.edu/~jurafsky/slp3/

Open vs. Closed Class

Open class ("content") words

- Nouns
  - Proper: *Janet, Italy*
  - Common: *cat, cats, mango*
- Verbs
  - Main: *eat, went*
- Adjectives: *old, green, tasty*
- Adverbs: *slowly, yesterday*
- Numbers: *122,312, one*
- Interjections: *Ow, hello*
- … more

Closed class ("function")

- Determiners: *the, some*
- Conjunctions: *and, or*
- Pronouns: *they, its*
- Auxiliary: *can, had*
- Prepositions: *to, with*
- Particles: *off, up*
- … more

https://web.stanford.edu/~jurafsky/slp3/

Part-of-Speech Tagging

Assigning a part-of-speech to each word in a text.

Words often have more than one POS.

book:

- VERB: (***Book*** *that flight*)
- NOUN: (*Hand me that* ***book***).

https://web.stanford.edu/~jurafsky/slp3/

Popular tag-set: Penn Treebank

Tag	Description	Example	Tag	Description	Example	Tag	Description	Example
CC	coordinating conjunction	*and, but, or*	PDT	predeterminer	*all, both*	VBP	verb non-3sg present	*eat*
CD	cardinal number	*one, two*	POS	possessive ending	*'s*	VBZ	verb 3sg pres	*eats*
DT	determiner	*a, the*	PRP	personal pronoun	*I, you, he*	WDT	wh-determ.	*which, that*
EX	existential 'there'	*there*	PRP$	possess. pronoun	*your, one's*	WP	wh-pronoun	*what, who*
FW	foreign word	*mea culpa*	RB	adverb	*quickly*	WP$	wh-possess.	*whose*
IN	preposition/ subordin-conj	*of, in, by*	RBR	comparative adverb	*faster*	WRB	wh-adverb	*how, where*
JJ	adjective	*yellow*	RBS	superlatv. adverb	*fastest*	$	dollar sign	*$*
JJR	comparative adj	*bigger*	RP	particle	*up, off*	#	pound sign	*#*
JJS	superlative adj	*wildest*	SYM	symbol	+,%, &	"	left quote	' or "
LS	list item marker	*1, 2, One*	TO	"to"	*to*	"	right quote	' or "
MD	modal	*can, should*	UH	interjection	*ah, oops*	(	left paren	[, (, {, <
NN	sing or mass noun	*llama*	VB	verb base form	*eat*	)	right paren	],), }, >
NNS	noun, plural	*llamas*	VBD	verb past tense	*ate*	,	comma	,
NNP	proper noun, sing.	*IBM*	VBG	verb gerund	*eating*	.	sent-end punc	. ! ?
NNPS	proper noun, plu.	*Carolinas*	VBN	verb past part.	*eaten*	:	sent-mid punc	: ; ... – -

https://web.stanford.edu/~jurafsky/slp3/

Methods and Evaluation

Methods:

Hidden Markov Models

Maximum Entropy Markov Models

Conditional Random Fields

RNNs, Transformers

Evaluation:

Accuracy

Macro-F1 (giving equal importance to each tag)

https://web.stanford.edu/~jurafsky/slp3/

NLP Paradigms: Text Generation

Dialogs

Example: Dialogs

- Generating responses
  - That are **consistent** and **coherent** with the dialog history
  - That are **interesting** and **engaging**
  - That meaningfully progress the dialog **towards a goal**

https://web.stanford.edu/~jurafsky/slp3/

Two kind of conversational agents

1. Chatbots
   - mimic informal human chatting
   - for fun, or even for therapy
2. (Task-based) Dialogue Agents
   - interfaces to personal assistants
   - cars, robots, appliances
   - booking flights or restaurants

https://web.stanford.edu/~jurafsky/slp3/

Chatbot Architectures

Rule-based

Pattern-action rules (ELIZA)

+ A mental model (PARRY):

The first system to pass the Turing Test!

Corpus-based

Information Retrieval (XiaoIce)

Neural encoder-decoder (BlenderBot)

https://web.stanford.edu/~jurafsky/slp3/

Response by generation

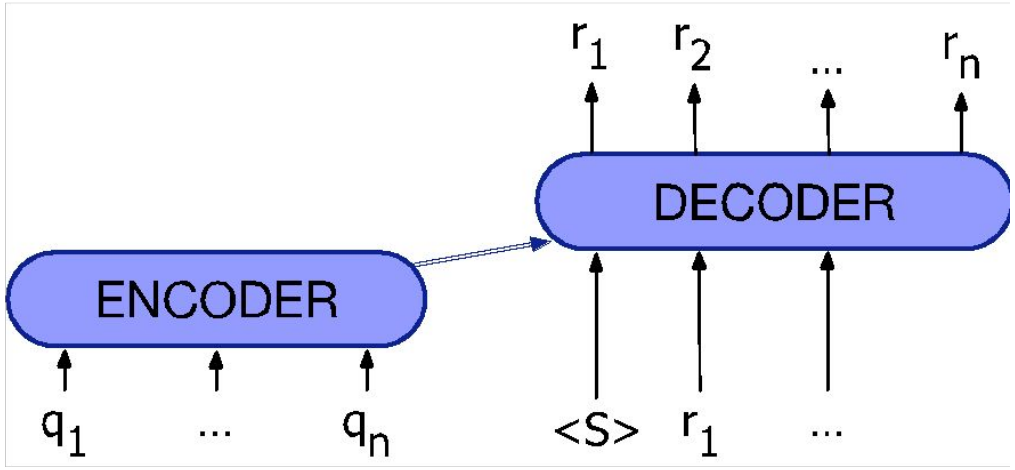
Think of response production as an encoder-decoder task

Generate each token r_t of the response by conditioning on the encoding of the entire query q and the response so far $r_1 \ldots r_{t-1}$

$$\hat{r}_t = \text{argmax}_{w \in V} P(w | q, r_1 \ldots r_{t-1})$$

Conditional LM

Evaluation is tricky



https://web.stanford.edu/~jurafsky/slp3/

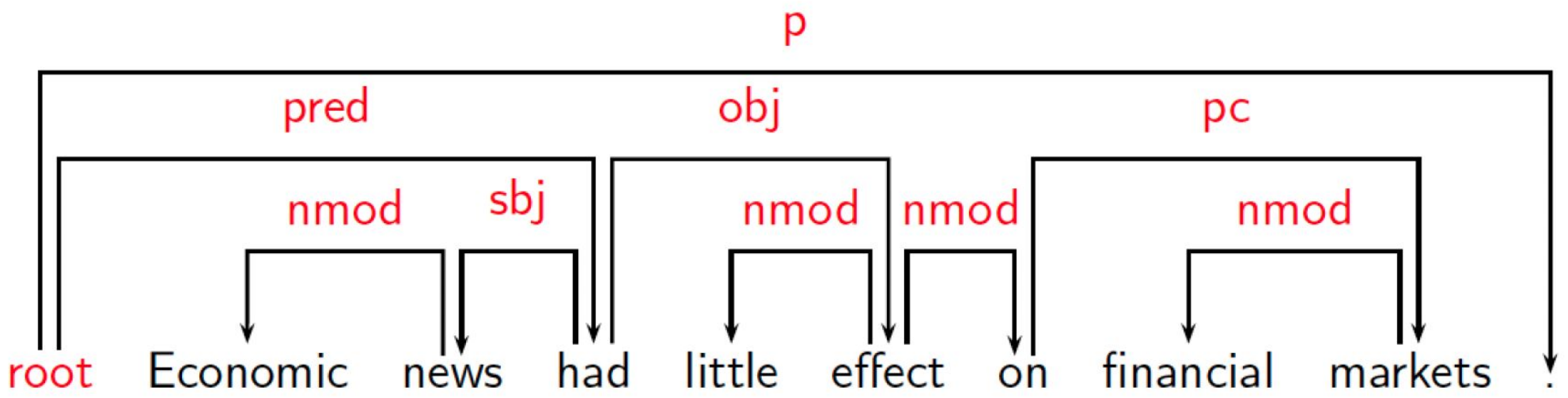
NLP Paradigms: Structured Prediction

Dependency Parsing

Dependency Parsing

Dependency Parsing

- **Input:** Sentence $x = w_1, \ldots, w_n$
- **Output:** Dependency graph G



Other NLP Tasks: Some Examples

Word Sense Disambiguation (WSD)

Sense ambiguity

- Many words have several meanings or senses
- The meaning of **bass** depends on the context
- Are we talking about music, or fish?
  - An electric guitar and **bass** player stand off to one side, not really part of the scene, just as a sort of nod to gringo expectations perhaps.
  - And it all started when fishermen decided the striped **bass** in Lake Mead were too skinny.

Disambiguation

- The task of disambiguation is to determine which of the senses of an ambiguous word is invoked in a particular use of the word.
- This is done by looking at the context of the word's use.

Entity Linking

Iranian POW negotiator holds talks with Iraqi ministers

The head of **Iran**'s **prisoner of war** commission met with two **Iraqi** Cabinet ministers Saturday in a bid to glean information about thousands of Iranian POWs allegedly in Iraq, the official Iraqi News Agency reported.

Iraqi Foreign Minister **Mohammed Saeed al-Sahhaf** told Abdullah al-Najafi that the two states needed to ``speed up the closure of what remains from the POW and Missing-In-Action file,'' INA said.

The issue of POWs and missing persons remains a stumbling block to normalizing relations between the two neighbors.

Iraq has long maintained that it has released all Iranian prisoners captured in the **1980**-88 **Iran-Iraq War**. The countries accuse each other of hiding POWs and preventing visits by the **International Committee of the Red Cross** to prisoner camps.

The ICRC representative in **Baghdad**, Manuel Bessler, told The **Associated Press** that his organization has had difficulty visiting POWs on both sides on a regular basis.

In April, Iran released 5,584 fied as civil law detainees in the largest exchange since **1990**.

More than 1 million people w

Baghdad

Baghdad is the capital of Iraq and of Baghdad Governorate. With a metropolitan area estimated at a population of 7,000,000, it is the largest city in Iraq. It is the second-largest city in the Arab world (after Cairo) and the second-largest city in southwest Asia (after Tehran).

open in wikipedia

Sentence Similarity

id	qid1	qid2	question1	question2	is_duplicate
447	895	896	What are natural numbers?	What is a least natural number?	0
1518	3037	3038	Which pizzas are the most popularly ordered pizzas on Domino's menu?	How many calories does a Dominos pizza have?	0
3272	6542	6543	How do you start a bakery?	How can one start a bakery business?	1
3362	6722	6723	Should I learn python or Java first?	If I had to choose between learning Java and Python, what should I choose to learn first?	1

Quora Question Pairs

Quora Question Pairs (QQP) dataset consists of over 400,000 question pairs, and each question pair is annotated with a binary value indicating whether the two questions are paraphrase of each other.

https://paperswithcode.com/dataset/quora-question-pairs

Question Answering

The Norman dynasty had a major political, cultural and military impact on medieval Europe and even the Near East. The Normans were famed for their martial spirit and eventually for their Christian piety, becoming exponents of the Catholic orthodoxy into which they assimilated. They adopted the Gallo-Romance language of the Frankish land they settled, their dialect becoming known as Norman, Normaund or Norman French, an important literary language. The Duchy of Normandy, which they formed by treaty with the French crown, was a great fief of medieval France, and under Richard I of Normandy was forged into a cohesive and formidable principality in feudal tenure. The Normans are noted both for their culture, such as their unique Romanesque architecture and musical traditions, and for their significant military accomplishments and innovations. Norman adventurers founded the Kingdom of Sicily under Roger II after conquering southern Italy on the Saracens and Byzantines, and an expedition on behalf of their duke, William the Conqueror, led to the Norman conquest of England at the Battle of Hastings in 1066. Norman cultural and military influence spread from these new European centres to the Crusader states of the Near East, where their prince Bohemond I founded the Principality of Antioch in the Levant, to Scotland and Wales in Great Britain, to Ireland, and to the coasts of north Africa and the Canary Islands.

Ground Truth Answers: William the Conqueror | William the Conqueror | William the Conqueror

Who ruled the duchy of Normandy
Ground Truth Answers: Richard I | Richard I | Richard I

What religion were the Normans
Ground Truth Answers: Catholic | Catholic orthodoxy | Catholic

What type of major impact did the Norman dynasty have on modern Europe?
Ground Truth Answers: <No Answer>

Who was famed for their Christian spirit?
Ground Truth Answers: <No Answer>

Who assimilted the Roman language?
Ground Truth Answers: <No Answer>

SQuAD2.0

The Stanford Question Answering Dataset

More Example Tasks, Benchmarks

2892 dataset results for Texts ×

Search for datasets

Best match

Filter by Modality (clear)

- Texts ×
- Images 3021
- Videos 949
- Audio 453

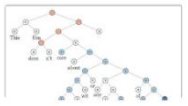
Filter by Task

- Question Answering 295
- Language Modelling 121
- Text Generation 115
- Text Classification 111

GLUE (General Language Understanding Evaluation benchmark)

General Language Understanding Evaluation (GLUE) benchmark is a collection of nine natural language understanding tasks, including single-sentence tasks CoLA and SST-2, similarity...

3,019 PAPERS • 24 BENCHMARKS



SST (Stanford Sentiment Treebank)

The Stanford Sentiment Treebank is a corpus with fully labeled parse trees that allows for a complete analysis of the compositional effects of sentiment in language. The corpus is bas...

2,223 PAPERS • 10 BENCHMARKS

SQuAD (Stanford Question Answering Dataset)

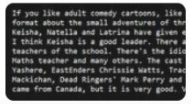
The Stanford Question Answering Dataset (SQuAD) is a collection of question-answer pairs derived from Wikipedia articles. In SQuAD, the correct answers of questions can be any se-...

2,077 PAPERS • 13 BENCHMARKS

MultiNLI (Multi-Genre Natural Language Inference)

The Multi-Genre Natural Language Inference (MultiNLI) dataset has 433K sentence pairs. Its size and mode of collection are modeled closely like SNLI. MultiNLI offers ten distinct genre...

1,762 PAPERS • 3 BENCHMARKS



IMDb Movie Reviews

The IMDb Movie Reviews dataset is a binary sentiment analysis dataset consisting of 50,000 reviews from the Internet Movie Database (IMDb) labeled as positive or negative. The datas...

1,716 PAPERS • 11 BENCHMARKS

https://paperswithcode.com/datasets?mod=texts&page=1

REFERENCES

Daniel Jurafsky and James H. Martin. 2024. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models, 3rd edition. Online manuscript released August 20, 2024.


INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR
1951
योगः कर्मसु कौशलम्
NPTEL
THANK YOU