

Natural Language Processing Applications

Lecture 2: NLP Applications



fit@hcmus

KHOA CÔNG NGHỆ THÔNG TIN
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

- ❑ Development Status
- ❑ Sample Applications



NLPA - NLP Applications

DEVELOPMENT STATUS



Development Status

mostly solved

Spam detection

Let's go to Agra! ✓

Buy V1AGRA ... ✗

Part-of-speech (POS) tagging

ADJ ADJ NOUN VERB ADV

Colorless green ideas sleep furiously.

Named entity recognition (NER)

PERSON ORG LOC

Einstein met with UN officials in Princeton

making good progress

Sentiment analysis

Best roast chicken in San Francisco! 👍

The waiter ignored us for 20 minutes. 👎

Coreference resolution

Carter told Mubarak he shouldn't run again.

Word sense disambiguation (WSD)

I need new batteries for my *mouse*.

Parsing

I can see Alcatraz from the window!

Machine translation (MT)

第13届上海国际电影节开幕...

The 13th Shanghai International Film Festival...

Information extraction (IE)

You're invited to our dinner party, Friday May 27 at 8:30

Party
May 27
add

still really hard

Question answering (QA)

Q. How effective is ibuprofen in reducing fever in patients with acute febrile illness?

Paraphrase

XYZ acquired ABC yesterday

ABC has been taken over by XYZ

Summarization

The Dow Jones is up

The S&P500 jumped

Housing prices rose

Economy is good

Dialog

Where is Citizen Kane playing in SF?

Castro Theatre at 7:30. Do you want a ticket?

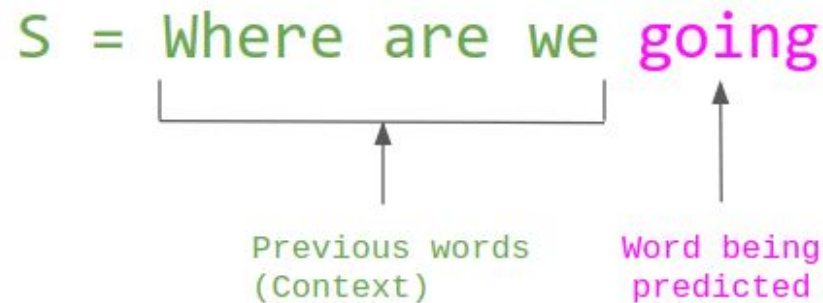
NLPA - NLP Applications

SAMPLE APPLICATIONS



Language Models

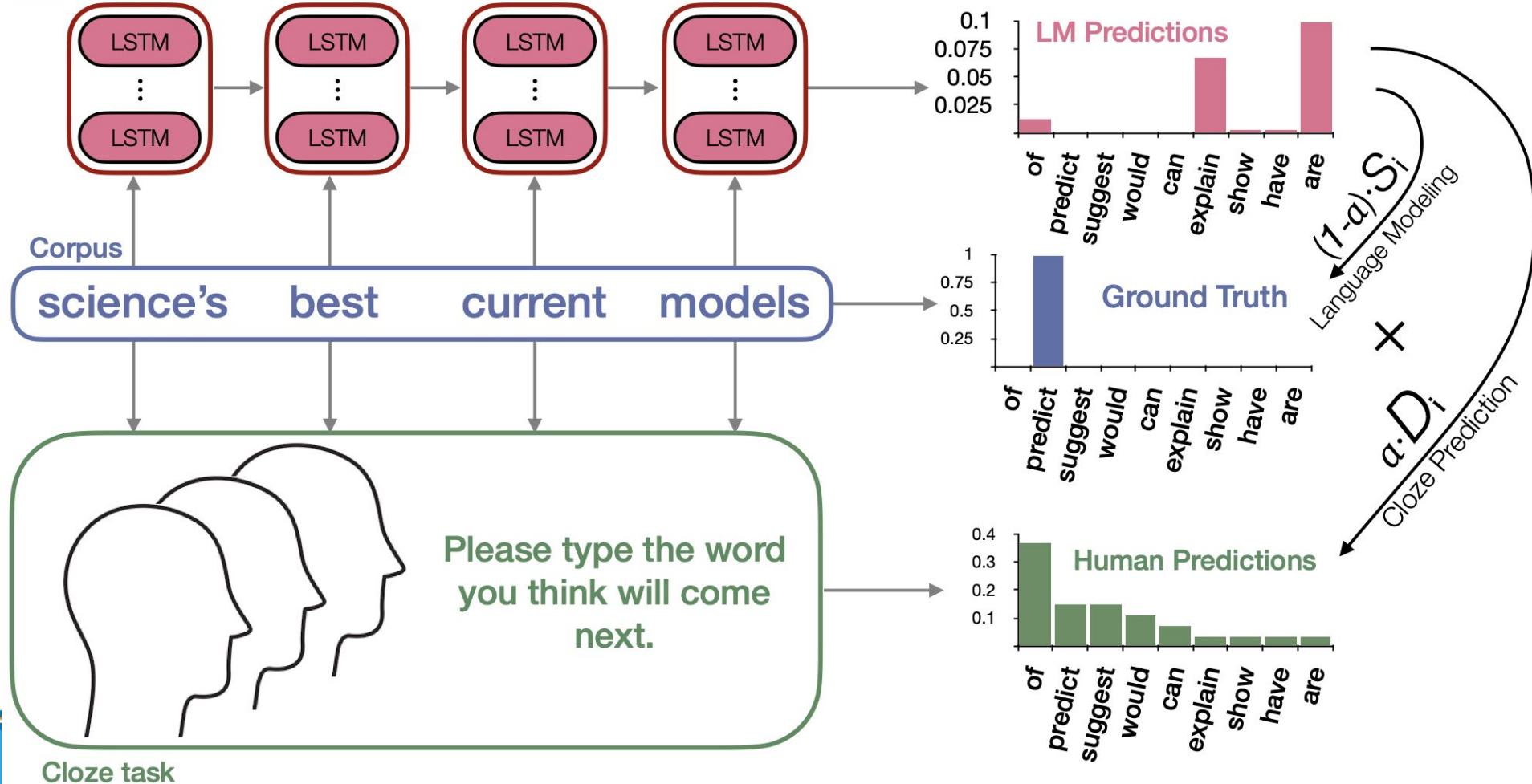
- ❑ Prediction of next words:
 - ❑ Given a context C : Calculate the probability of the next word w ?



$$P(S) = P(\text{Where}) \times P(\text{are} \mid \text{Where}) \times P(\text{we} \mid \text{Where are}) \times P(\text{going} \mid \text{Where are we})$$

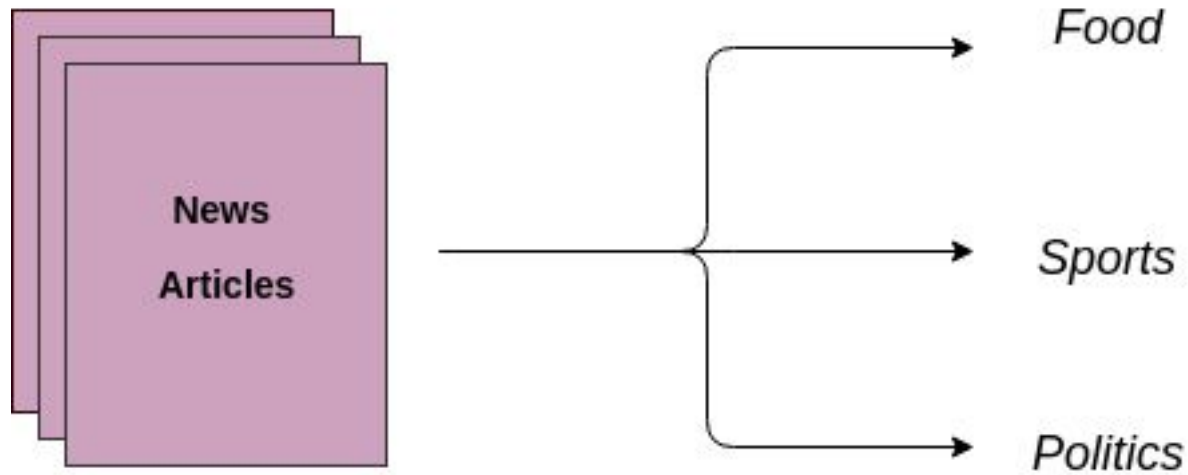
Language Models (Cont)

Pre-trained LM



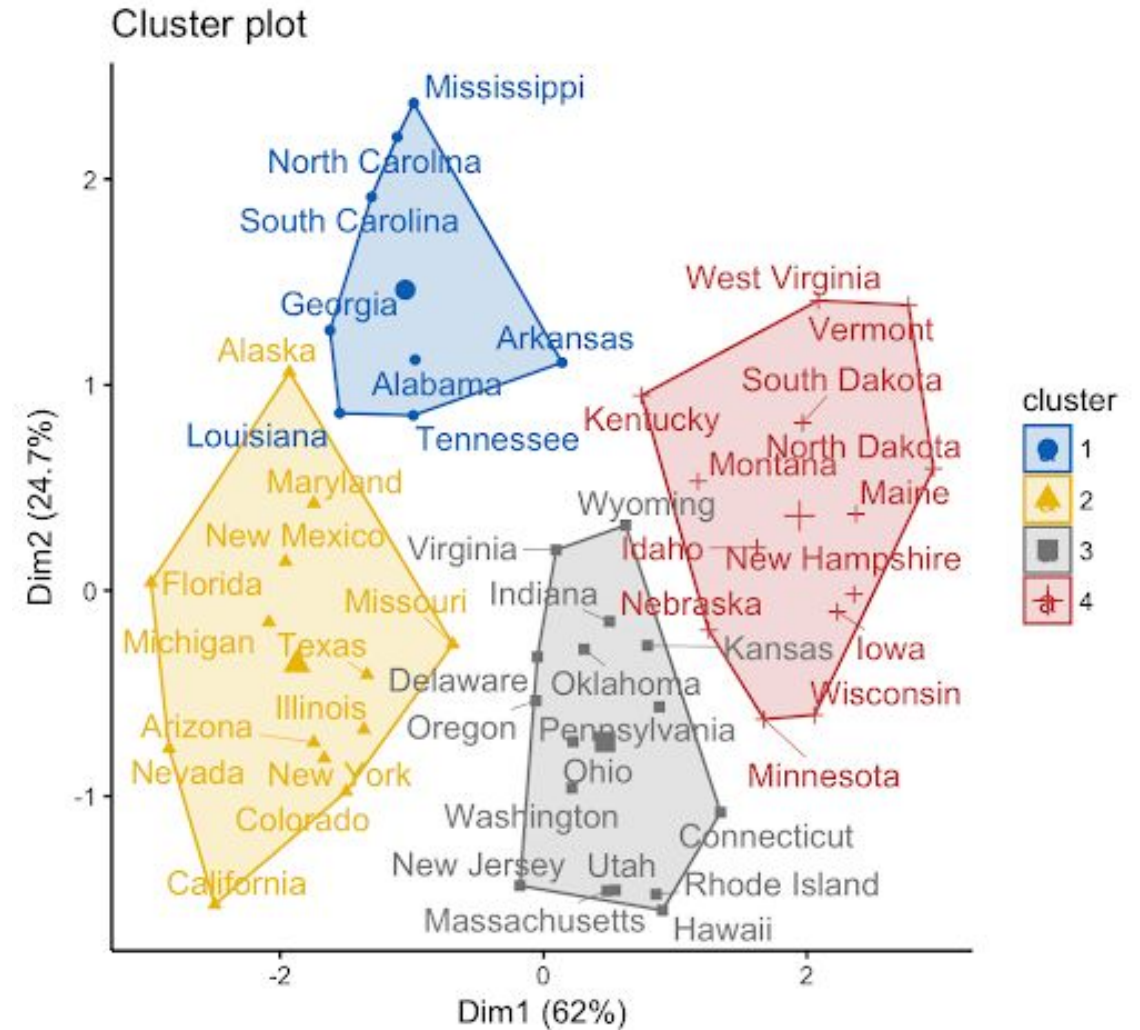
Text Classification

- Given a text T , calculate the probability of $t \in T$ for a pre-defined class $c \in \mathbf{C}$.



Text Clustering

- Given a text T and a number of clusters k , divide T to k clusters based on linguistic features.



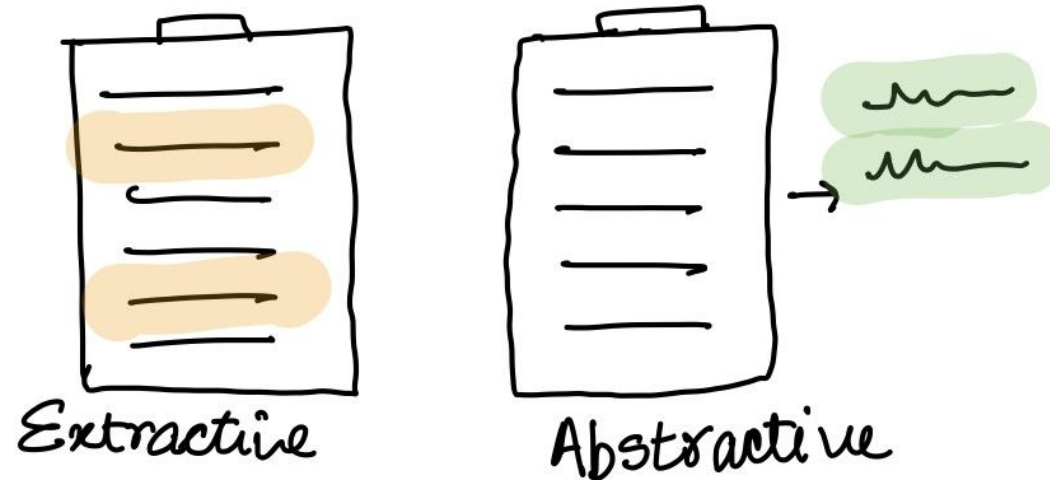
Text Similarity

- Given two texts T1 and T2, calculate the similarity score between T1 and T2.



Text Summarization

- Given a text T , return a new text T' containing the content of T but the length of T' is shorter the length of T .



Machine Translation

Text for Translation

Hola Mundo

你好世界

नमस्ते दुनिया

Bonjour le monde

ハローワールド

مرحباً بالعالم

Hallo Welt



Translated Text

Hello World

Hello World

Hello World

Hello World

Hello World

Hello World

Hello World



Q & A

