EECS 598-017: Natural Language Processing with Deep Learning

Instructor: Prof. Lu Wang

Computer Science and Engineering

University of Michigan

https://web.eecs.umich.edu/~wangluxy/

Time and Location

- **Time**: Mondays and Wednesdays, 12 pm - 1:30 pm
- **Location**: online via Zoom (link is provided on piazza, anyone with umich email can join piazza and the discussions too)

Overall Goals

- Recent advances of using deep learning methods for solving natural language processing problems

- More importantly, we will learn: *how to define a research problem, propose a solution, and evaluate the solution*
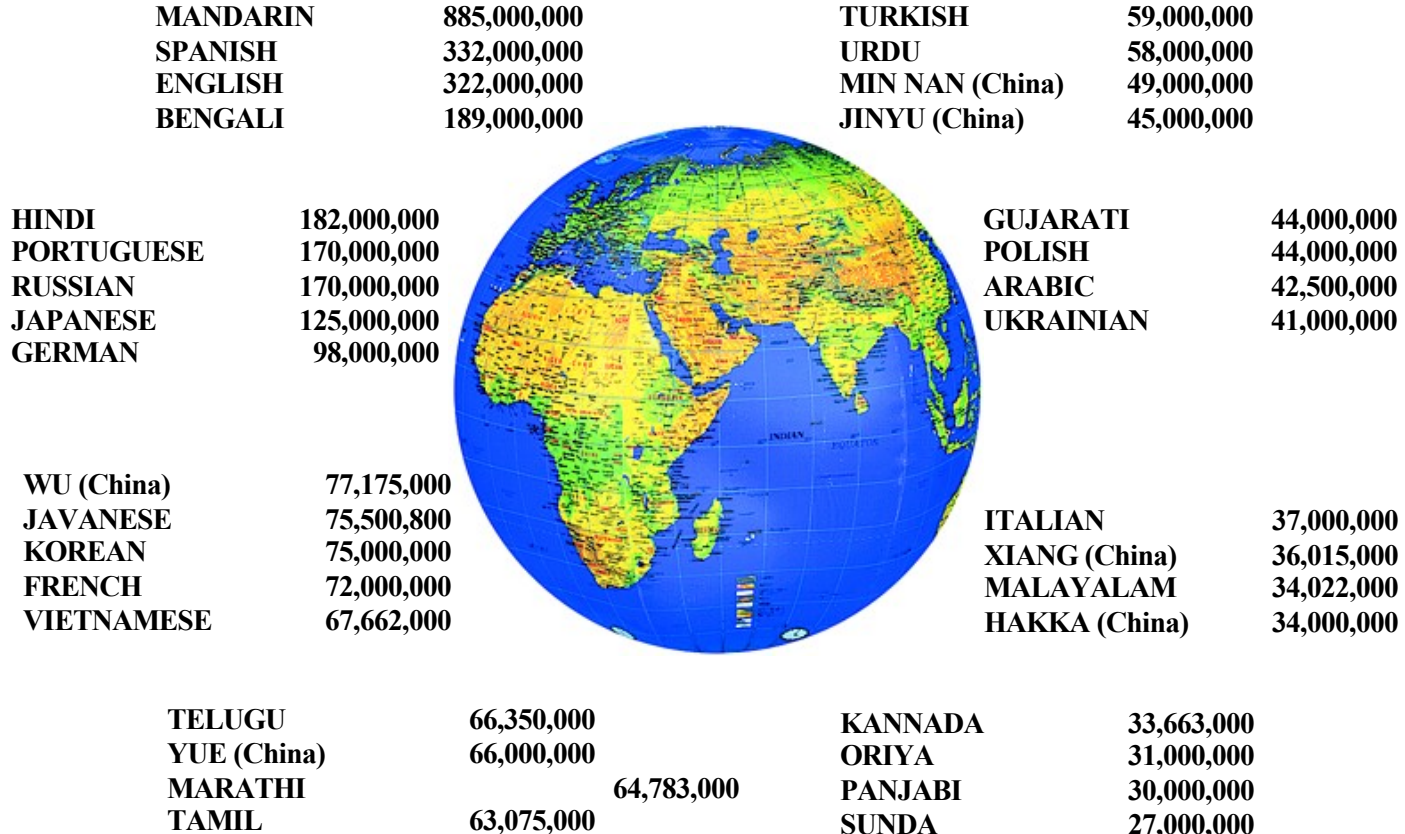
Some real-world applications

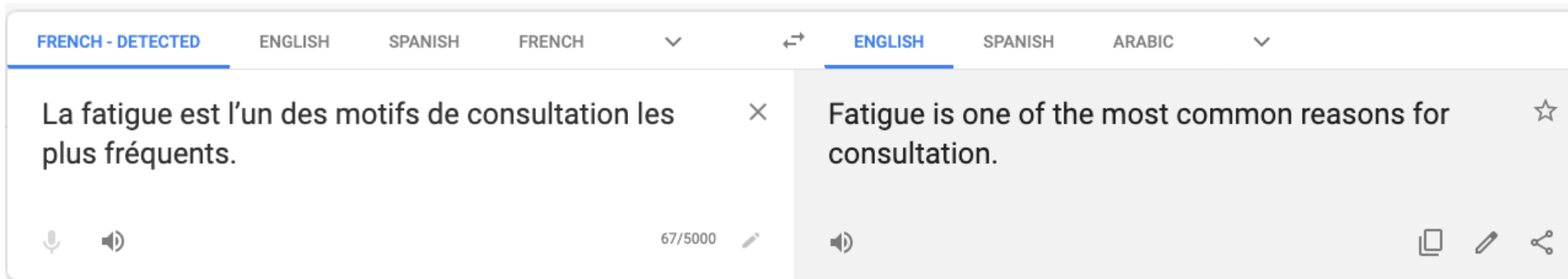- Machine translation

Mary didn't slap the green witch.

Maria no dió una bofetada a la bruja verde.
(Mary do not gave a slap to the witch green.)

Thousands of Languages Are Spoken

MANDARIN	885,000,000	TURKISH	59,000,000
SPANISH	332,000,000	URDU	58,000,000
ENGLISH	322,000,000	MIN NAN (China)	49,000,000
BENGALI	189,000,000	JINYU (China)	45,000,000
HINDI	182,000,000	GUJARATI	44,000,000
PORTUGUESE	170,000,000	POLISH	44,000,000
RUSSIAN	170,000,000	ARABIC	42,500,000
JAPANESE	125,000,000	UKRAINIAN	41,000,000
GERMAN	98,000,000		
WU (China)	77,175,000		
JAVANESE	75,500,800	ITALIAN	37,000,000
KOREAN	75,000,000	XIANG (China)	36,015,000
FRENCH	72,000,000	MALAYALAM	34,022,000
VIETNAMESE	67,662,000	HAKKA (China)	34,000,000
TELUGU	66,350,000	KANNADA	33,663,000
YUE (China)	66,000,000	ORIYA	31,000,000
MARATHI	64,783,000	PANJABI	30,000,000
TAMIL	63,075,000	SUNDA	27,000,000

Source: Ethnologue

Some real-world applications

- Machine translation

Google Translate

| FRENCH - DETECTED | ENGLISH | SPANISH | FRENCH | ⌄ | ⇄ | ENGLISH | SPANISH | ARABIC | ⌄ |

La fatigue est l'un des motifs de consultation les plus fréquents.   ✕

Fatigue is one of the most common reasons for consultation.   ☆

🎤 🔊   67/5000 ✏️   🔊   📋 ✏️ ⌣

Some real-world applications

- Spoken Dialogue Systems
  - Q: Is it going to rain today?
  - A: It will be mostly sunny. No rain is expected.

Examples of modern Virtual Assistant dialogue systems

- Android Google Now
  - Predictive search assistant
- Windows Cortana
  - Works across different Windows devices
  - Aims to be able to "talk about anything"

Examples of modern Virtual Assistant dialogue systems

- Amazon Echo – home assistant device
  - Plays music
    - With voice commands
  - Question answering
    - Get weather, news
    - More complex questions, like
      - "how many spoons are in a cup?"
  - Setting timer
  - Manages TODO lists

News Article Writing

Artificial Intelligence Can Now Write Amazing Content -- What Does That Mean For Humans?

https://www.forbes.com/sites/bernardmarr/2019/03/29/artificial-intelligence-can-now-write-amazing-content-what-does-that-mean-for-humans

YAHOO!

Yahoo! Sports uses NLG to produce over 70 million reports and match recaps—each one unique—that help engage, monetize, and delight its massive user base.

- Over 70 million fantasy football player recaps
- Increased time spent on page for website visitors

What is Natural Language Processing?

What is Natural Language Processing?

- Allowing machines to communicate with human

- Natural language understanding + natural language generation

What is Natural Language Processing?

- Allowing machines to communicate with human

- Natural language understanding + **natural language generation**

We will read papers mainly on NLG

Overall Goals

- Recent advances of using deep learning methods for solving natural language processing problems, especially in natural language generation

- More importantly, we will learn: *how to define a research problem, propose a solution, and evaluate the solution*

Prerequisites

- Programming
  - Being proficient in some programming languages (Python is encouraged).

- Courses
  - Familiarity with natural language processing concepts and machine learning fundamentals, e.g., have done projects with machine learning tools to train and evaluate computational and statistical models.

  - If you have not taken any NLP course before and would like to learn fundamental concepts and techniques in NLP, you should consider **EECS 595** (graduate-level NLP, offered every fall or winter semester) or **EECS 495** (undergraduate-level NLP, scheduled in winter 2021).

References (no required textbook)

- Jacob Eisenstein, "Introduction to Natural Language Processing", The MIT Press, 2019

- Dan Jurafsky and James H. Martin, "Speech and Language Processing, 2nd Edition", Prentice Hall, 2009.
  - 3rd edition http://web.stanford.edu/~jurafsky/slp3/

Course Format

- This is a seminar-style course:

    1. paper reading, presentation, and discussion

    -> We will learn *how to define a problem*, *how to design model and experiments, how to properly evaluate the proposed model and results*, etc

    -> Each class discusses two papers on a selected topic. For each paper, we will start with (tentatively) 15 minutes presentation (delivered by 1 or 2 students), followed by (1) 10 minutes group discussion and 15 minutes class discussion (e.g. based on questions collected from groups) or (2) two groups discussions and class discussions

    2. working on a course project in small groups and reporting results

    -> You will select a relevant topic of your interests and gain hands-on research experience.

Grading

- **Paper critiques** (38%): write a short (half a page to one page) critique for each paper , each submission worth 2%; the critiques are due the day before the class
  - e.g. for papers discussed on Sep 2, critiques are due on Sep 1
- **Paper presentation** and discussion leading (10%): each student will present papers and lead corresponding discussions twice (with another student or individually), each of 5%; share slides with instructor two days before presentation to get feedback (e.g. if you present on Sep 9, please email the instructor by Sep 7)

- **Project** (40%): team of 2 to 3 students, proposal (5%), reports (10%+15%, mid-tern and final), presentations (5%+5%, mid-term and final)
- **Project feedback** (6%): each student will write short feedback to other groups' projects after mid-term presentation

- **Participation** (6%): attendance, participating in-class discussions, etc

Course Webpage

- https://web.eecs.umich.edu/~wangluxy/courses/eecs598_fa2020/eecs598_fa2020.html
  - Up-to-date schedule with reading list and due dates
  - Zoom links for lectures and office hours

- You can also go to the instructor's web page and find it from there:
  - https://web.eecs.umich.edu/~wangluxy

Outlines

- Selected topics for this semester
- How to present a paper
- How to write a paper critique
- How to choose a project topic

Themes to be covered

- Text Summarization
- Text Generation
- Dialogue Generation
- Question Generation
- NLG Evaluation
- Bias and Ethics for Neural Generation

Text Summarization

- **Goal**: produce an abridged version of a text that contains information that is important or relevant to a user.

- **Summarization Applications**
  - **outlines or abstracts** of any document, article, etc
  - **summaries** of email threads
  - **action items** from a meeting
  - **simplifying** text by compressing sentences

News articles

Emails

Social Media Streams

Scientific Articles

Books

Websites

Text Summarization

- Produce a short summary for a document
  - **Article:** An international team of scientists studied diet and mortality in 135,335 people between 35 and 70 years old in 18 countries, following them for an average of more than seven years. Diet information depended on self-reports, and the scientists controlled for factors including age, sex, smoking, physical activity and body mass index. The study is in The Lancet. Compared with people who ate the lowest 20 percent of carbohydrates, those who ate the highest 20 percent had a 28 percent increased risk of death. But high carbohydrate intake was not associated with cardiovascular death. …

  - **Summary:** Researchers found that people who ate higher amounts of carbohydrates had a higher risk of dying than those who ate more fats.

23

Abstractive Summarization

- Input: Congratulations to Australia for seeing sense and dropping the ridiculous policy of not selecting their best players if they are playing overseas.

- Summary: Australia have seen sense by revamping their overseas selection policy.

Abstractive Summarization

- Input: Congratulations to Australia for seeing sense and dropping the ridiculous policy of not selecting their best players if they are playing overseas.

- Summary: Australia have seen sense by revamping their overseas selection policy.
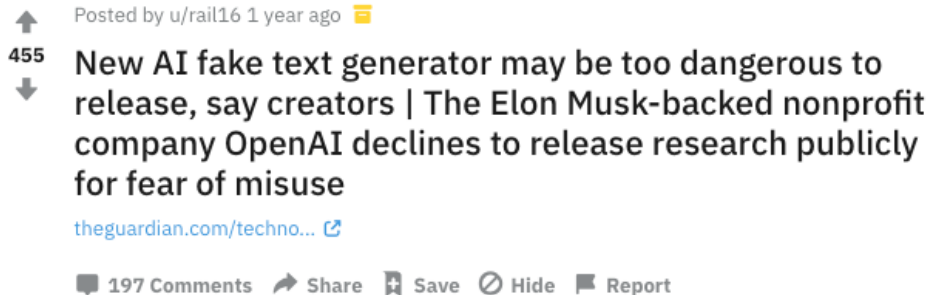
How to achieve human-level abstraction?

Natural Language Generation

- GPT-2!

Posted by u/rail16 1 year ago

455

New AI fake text generator may be too dangerous to release, say creators | The Elon Musk-backed nonprofit company OpenAI declines to release research publicly for fear of misuse

theguardian.com/techno... ↗

197 Comments   ↱ Share   Save   ⊘ Hide   ⚑ Report

Gun control is the only way to solve gun violence. We must enact gun control legislation. We must prevent gun purchases.

LGBT people are also victims of gun violence. They too must not be denied access to their right to bear arms.

And if you've made it this far, we will work together to prevent gun violence. Together, we can do it.

Natural Language Generation

- GPT-2!

Posted by u/rail16 1 year ago

455

New AI fake text generator may be too dangerous to release, say creators | The Elon Musk-backed nonprofit company OpenAI declines to release research publicly for fear of misuse

theguardian.com/techno...

197 Comments   Share   Save   Hide   Report

Gun control is the only way to solve gun violence. Second Amendment would not exist without it."

Westchester County Legislator, Charles French, a Democrat who supports the measure said, "In our society, you can't have a high crime rate, high number of killings and not be making some kind of adjustment and have a criminal penalty. The Second Amendment is about people being able to have a good time, protect their families and it should not prevent those things."

Natural Language Generation

- GPT-2!

Posted by u/rail16 1 year ago

455

New AI fake text generator may be too dangerous to release, say creators | The Elon Musk-backed nonprofit company OpenAI declines to release research publicly for fear of misuse

theguardian.com/techno...

💬 197 Comments ↗ Share 🔖 Save ⊘ Hide 🚩 Report

Gun control is the only way to solve gun violence. Second Amendment would not exist without it."

Westchester County Legislator, Charles French, a Democrat who supports the measure said, "In our society, you can't have a high crime rate, high number of killings and not be making some kind of adjustment and have a criminal penalty. The Second Amendment is about people being able to have a good time, protect their families and it should not prevent those things."

What are the problems?

28

Dialogue Systems

- Task-oriented dialogue
  - Assistive (e.g. customer service, giving recommendations, question answering…)
  - Co-operative (two agents solve a task together through dialogue)
  - Adversarial (two agents compete in a task through dialogue)

- Social dialogue
  - Chit-chat (for fun or company)
  - Therapy / mental wellbeing

[source: Stanford cs224 slides]

Question Generation

- Input: **Computational complexity theory** is a branch of the theory of computation in theoretical computer science that focuses on classifying computational problems according to their inherent difficulty, and relating those classes to each other. …

- Question: What branch of theoretical computer science deals with broadly classifying computational problems by difficulty and class of relationship?

Evaluations for NLG Systems

- The same meaning can be expressed in different ways.
- How can we reliably evaluate the system outputs?
- What we should care about in addition to relevance?

Bias and Ethics

Bias and Ethics

Bias and Ethics

- *"triggers cause…the GPT-2 language model to spew racist output even when conditioned on non-racial contexts"* [Wallace et al., EMNLP 2019]

Outlines

- Selected topics for this semester
- How to present a paper
- How to write a paper critique
- How to choose a project topic

How to present a paper

- It's a 15 minutes presentation (so it's like a summary, and facilitates the upcoming discussions)
- **Problem**: what does this paper study? (2-3 minutes)
- **Model**: which model is proposed and why it's suitable for the problem? (5-7 minutes)
- **Datasets and experiments**: which datasets are used for evaluation? And what are the evaluation metrics? (1-2 minutes)
- **Results and conclusion**: what are the results? Do the results support the conclusion? (3-5 minutes)

- Send your draft slides two days before the class (e.g. if you present on Sep 9, please email the instructor by Sep 7 to get feedback)

Outlines

- Selected topics for this semester
- How to present a paper
- How to write a paper critique
- How to choose a project topic

How to write a critique for an NLP paper?

- Many NLP papers are based on empirical methods
- Throughout the semester, one major goal is learning to evaluate:
  - The significance of a research topic
  - What is considered as novel methods
  - The importance of model/result evaluation, especially for NLG
  - How to come up with a new method to solve the problem

How to write a critique for an NLP paper?

- Many NLP papers are based on empirical methods
- Throughout the semester, one major goal is learning to evaluate:
  - The significance of a research topic
  - What is considered as novel methods
  - The importance of model/result evaluation, especially for NLG
  - How to come up with a new method to solve the problem

- Using critique template:
  https://web.eecs.umich.edu/~wangluxy/courses/eecs598_fa2020/material_eecs598_fa20/critique_template.pdf
  - List out two weaknesses and how to improve upon them
  - List out two questions that can be discussed in class

Outlines

- Selected topics for this semester
- How to present a paper
- How to write a paper critique
→ - How to choose a project topic

How to choose a course project

- NLP-related project, especially on the topics discussed in this course

- 2-3 students as a team

What is a reasonable course project

- The problem needs to be well-defined, useful, and practical.

- Can be an extension from one of the papers we read

- Can be a new problem

Sample Ideas

- Stanford NLP with deep learning class
  - http://web.stanford.edu/class/cs224n

- Feel free to explore anything you find interesting

Course Project

- You will need to talk to the instructor on project topics!
  - Zoom meetings (10-15 minutes) will be arranged on September 14 and 16, 5-6:30pm (see piazza)

- How to find teammates?
  - Talk to your classmates and see if you share interests!
  - How to do it online: Post on piazza with your background (programming language and skills) + potential project ideas + your email contact, other students should feel free to reach out

Course Project

- Three reports
  - One-page proposal (5%), due on Sep 18 at 11:59pm.
  - Progress report (10%)
  - Final report (15%)

- Two presentations
  - In class (5% each)

Submission and Late Policy

- All submissions are in electronic format.
  - Due on Canvas.

Submission and Late Policy

- Submissions turned in late will be charged 20 points (out of 100 points) off for each late day (i.e. 24 hours).

- Each student has a budget of **5 days in total** throughout the semester before a late penalty is applied.

- Late days are not applicable to presentations.

- Each group member is charged with the same number of late days, if any, for their submission.

Get in touch!

- All materials and schedule can be found on the course webpage:
  - https://web.eecs.umich.edu/~wangluxy/courses/eecs598_fa2020/eecs598_fa2020.html
- Office hours
  - Prof. Lu Wang: Wednesdays, from 5pm to 6pm (Zoom link is provided on Piazza)
- Piazza
  - http://piazza.com/umich/fall2020/eecs598017/home, please sign up.
  - All course relevant questions should go here!

Some Preliminaries on Sequence-to-sequence (seq2seq) Models for NLG

Some Preliminaries on Sequence-to-sequence (seq2seq) Models for NLG

["Sequence to Sequence Learning with Neural Networks", Sutskever et al., 2014]

Attention mechanism: decoder focuses on relevant words [Bahdanau et al., 2015]
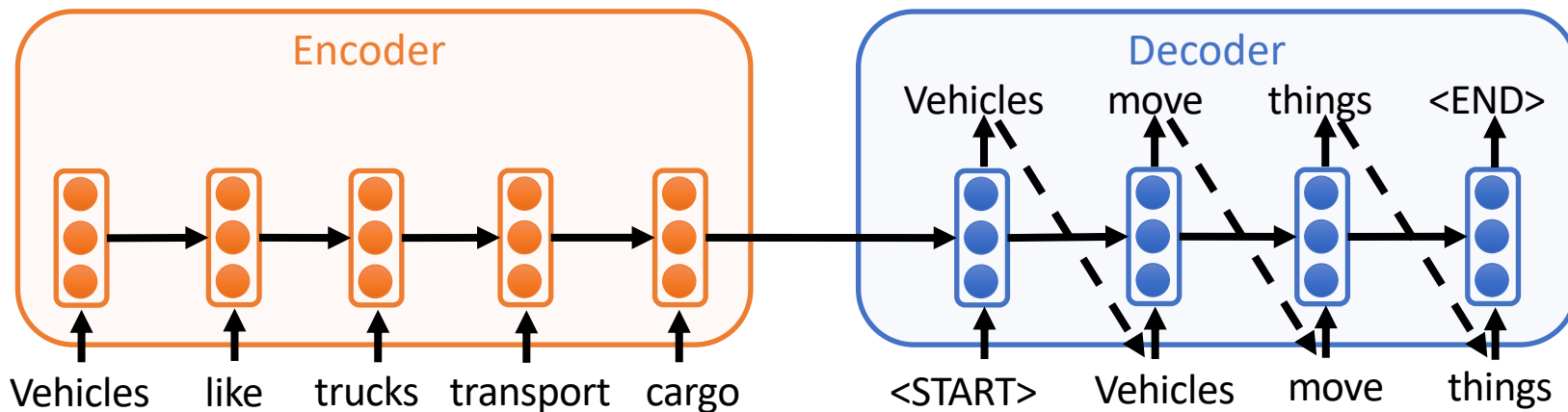
Encoder and decoder are often implemented with recurrent neural networks, e.g. long short-term memory (LSTM) networks.

Attention Mechanism
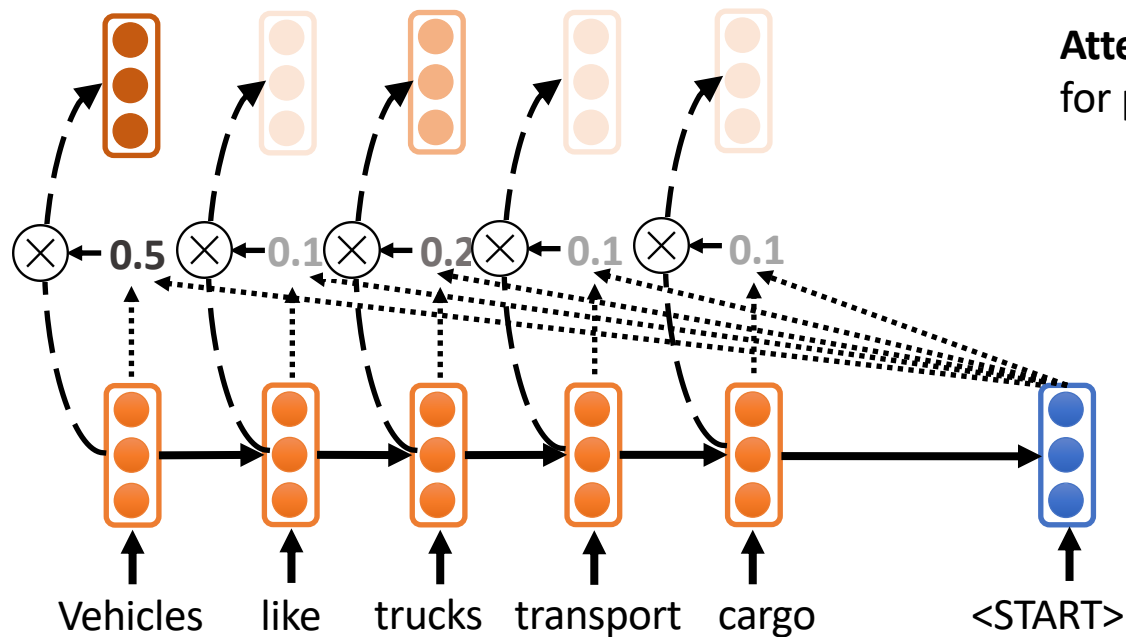
Attentions: which word(s) to focus on for prediction [Bahdanau et al., 2015]

0.5   0.1   0.2   0.1   0.1

Vehicles   like   trucks   transport   cargo   <START>

Attention Mechanism

Attentions: which word(s) to focus on for prediction   [Bahdanau et al., 2015]

52
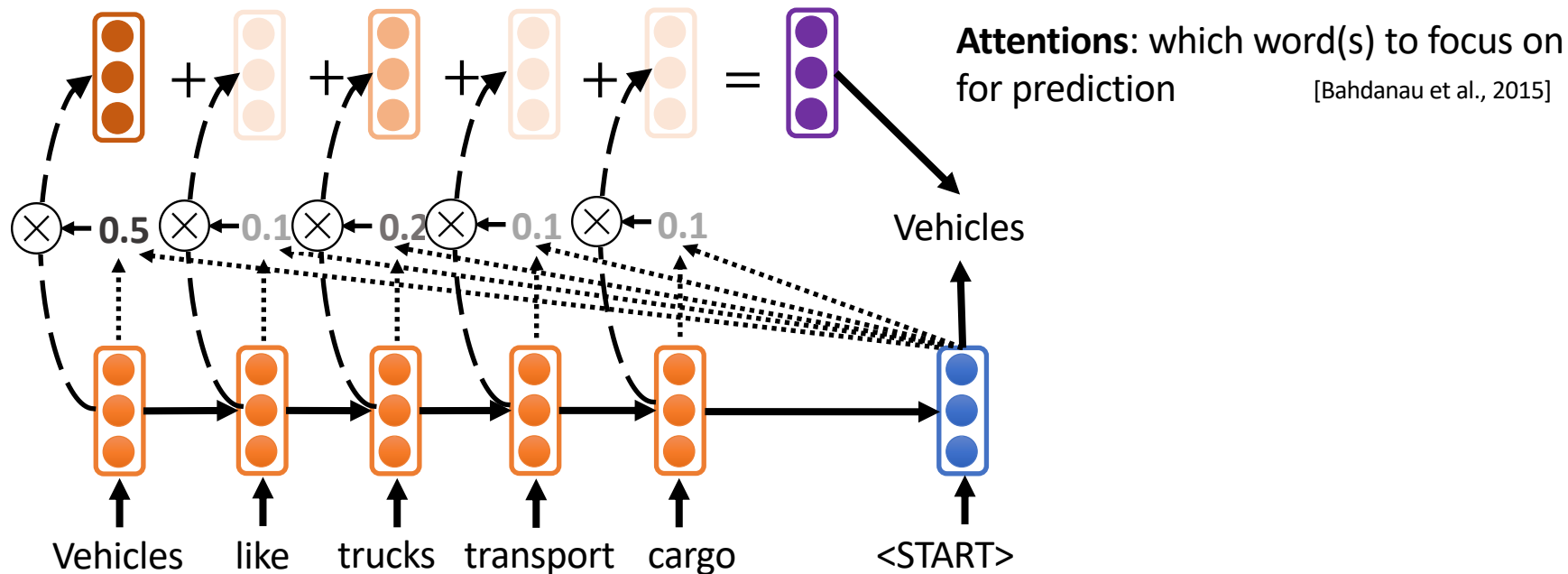
Attention Mechanism
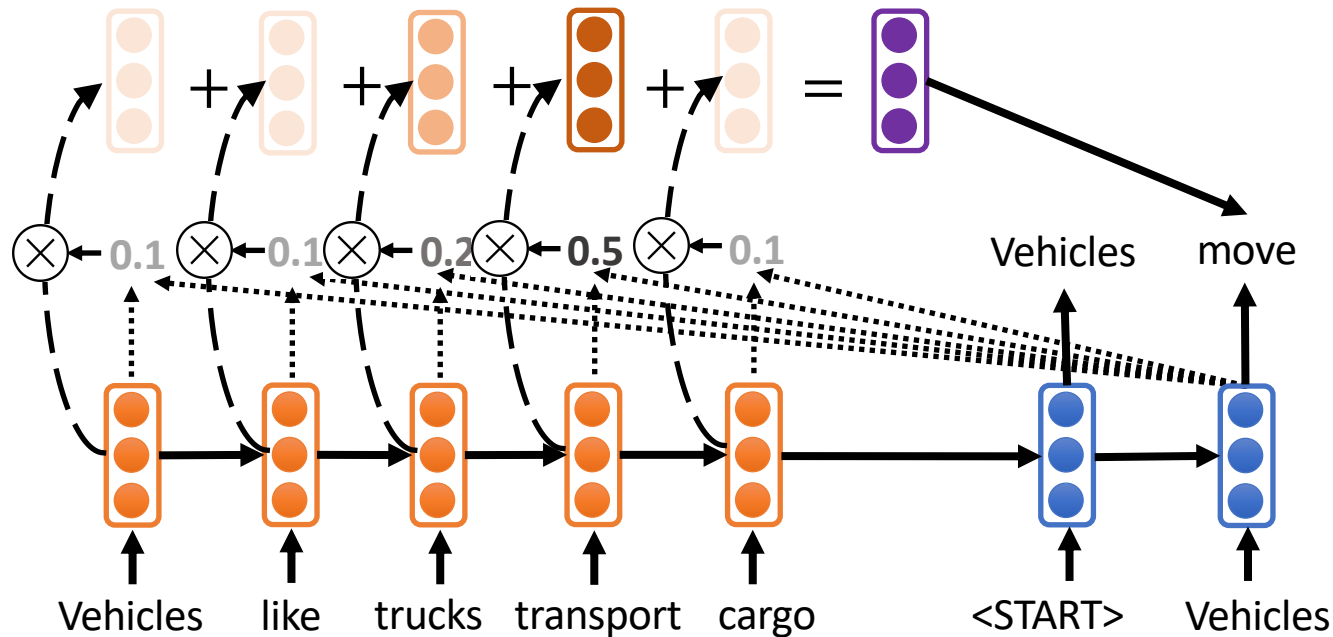
Attention Mechanism
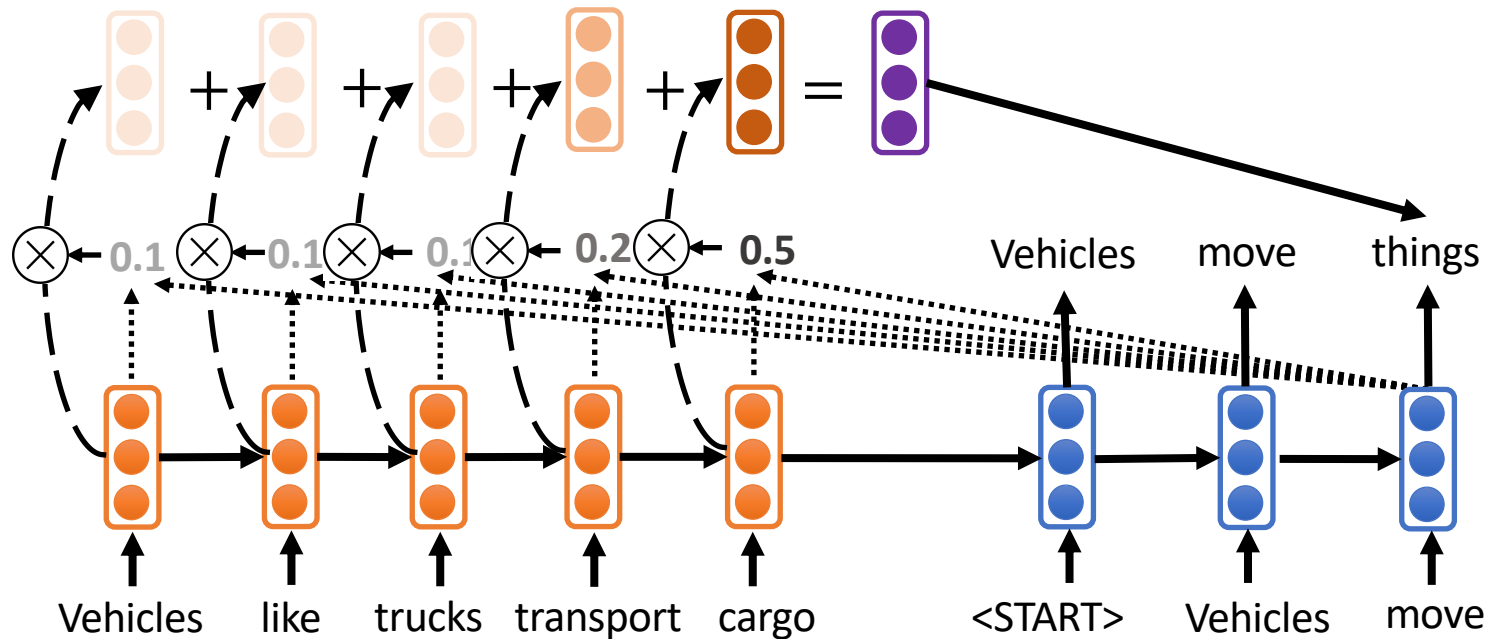
Attention Mechanism
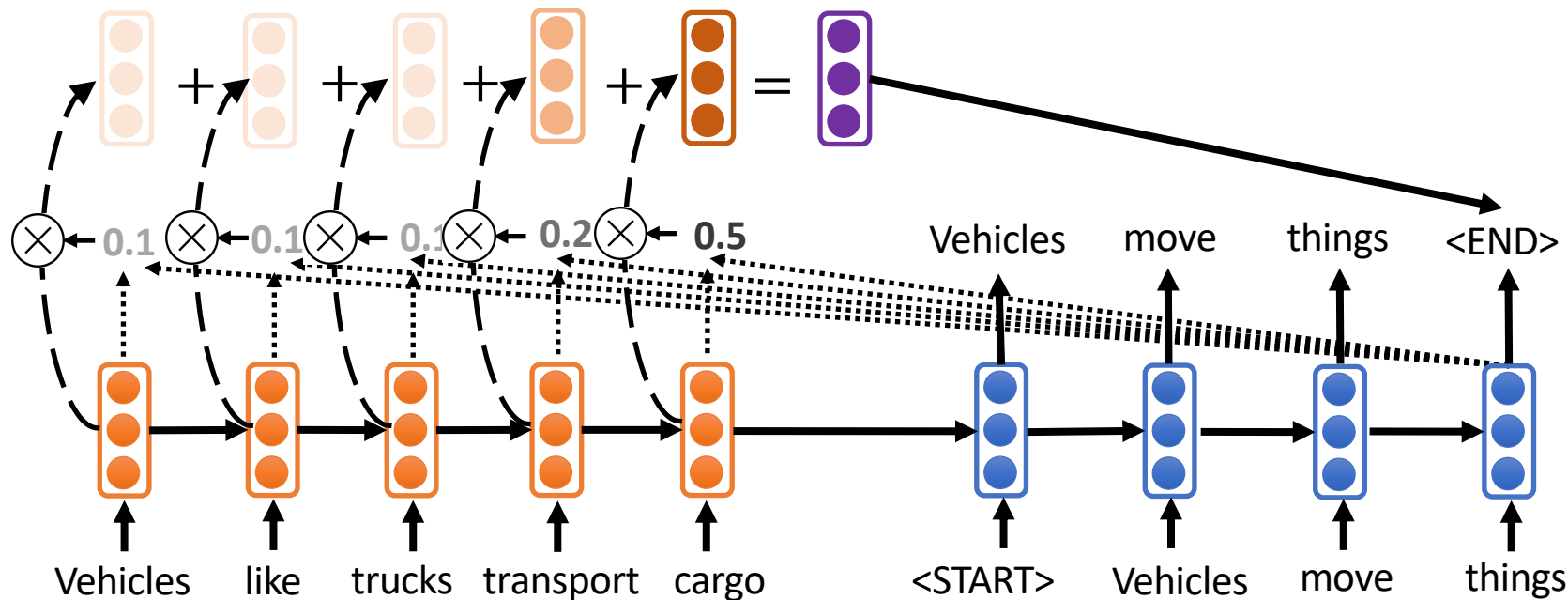
Transformer

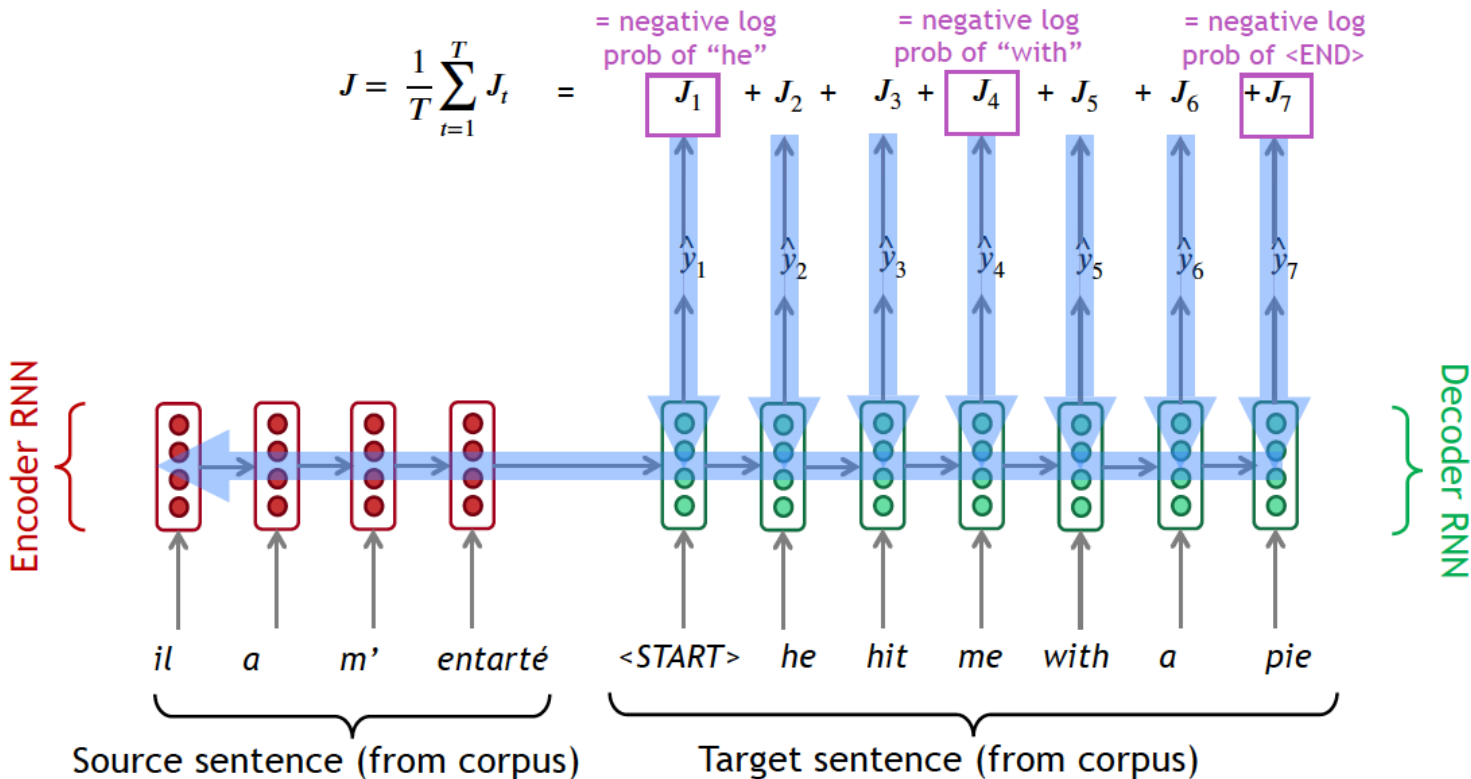["Attention Is All You Need", Vaswani et al, 2017]
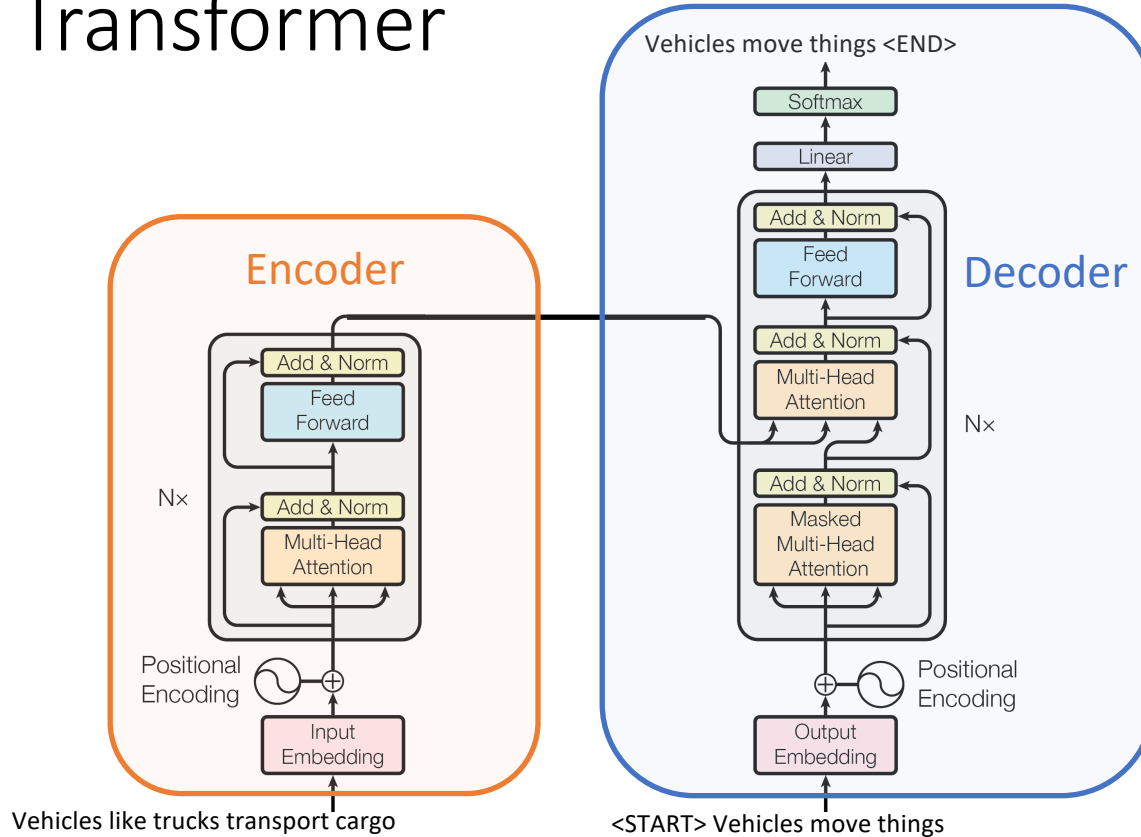
Seq2seq is used not only in NLG tasks

- Syntax and semantic parsing
- Program code generation

Two Important Aspect

- What is the training objective?
- How to do decoding? (i.e. generating a sequence of words)

What is the training objective? (Using translation as an example)

$$J = \frac{1}{T}\sum_{t=1}^{T} J_t \quad = \quad J_1 + J_2 + J_3 + J_4 + J_5 + J_6 + J_7$$

= negative log prob of "he"

= negative log prob of "with"

= negative log prob of <END>

$\hat{y}_1 \quad \hat{y}_2 \quad \hat{y}_3 \quad \hat{y}_4 \quad \hat{y}_5 \quad \hat{y}_6 \quad \hat{y}_7$

Encoder RNN

Decoder RNN

il   a   m'   entarté   <START>   he   hit   me   with   a   pie

Source sentence (from corpus)   Target sentence (from corpus)

59

How to do decoding?

- Greedy decoding: for each decoding step, choose the one that has the highest probability
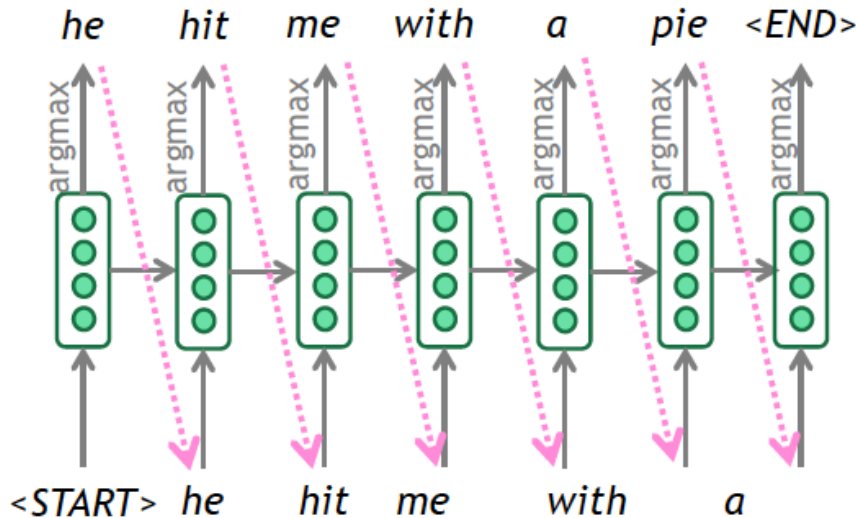
How to do decoding?

- Greedy decoding has no way to undo decisions!
  - Input: *il a m'entarté* *(he hit me with a pie)*
  - → *he* ____
  - → *he hit* ____
  - → *he hit a* ____   (whoops! no going back now…)
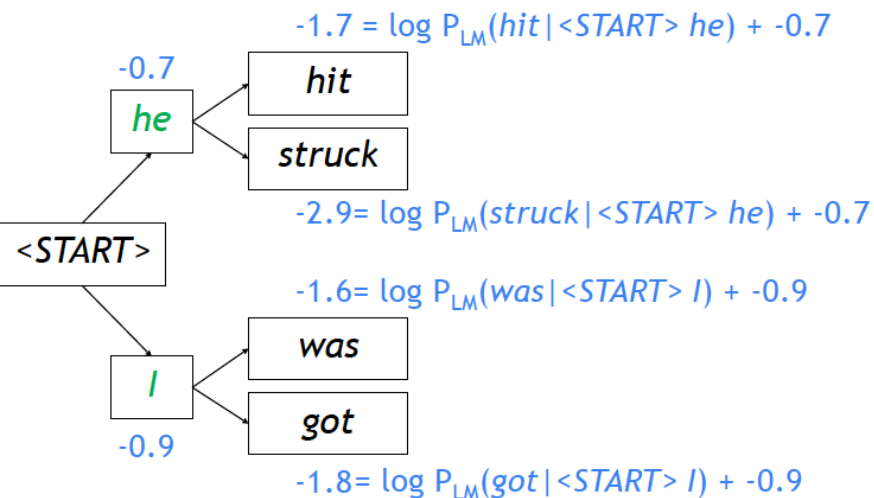
- How to fix this?

Can we try all possible sequences?

How to do decoding?

- Beam search decoding
- Core idea: On each step of decoder, keep track of the k most probable partial translations (which we call hypotheses) k is the beam size (in practice around 5 to 10)
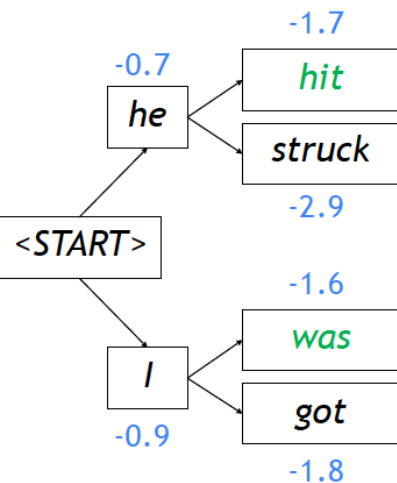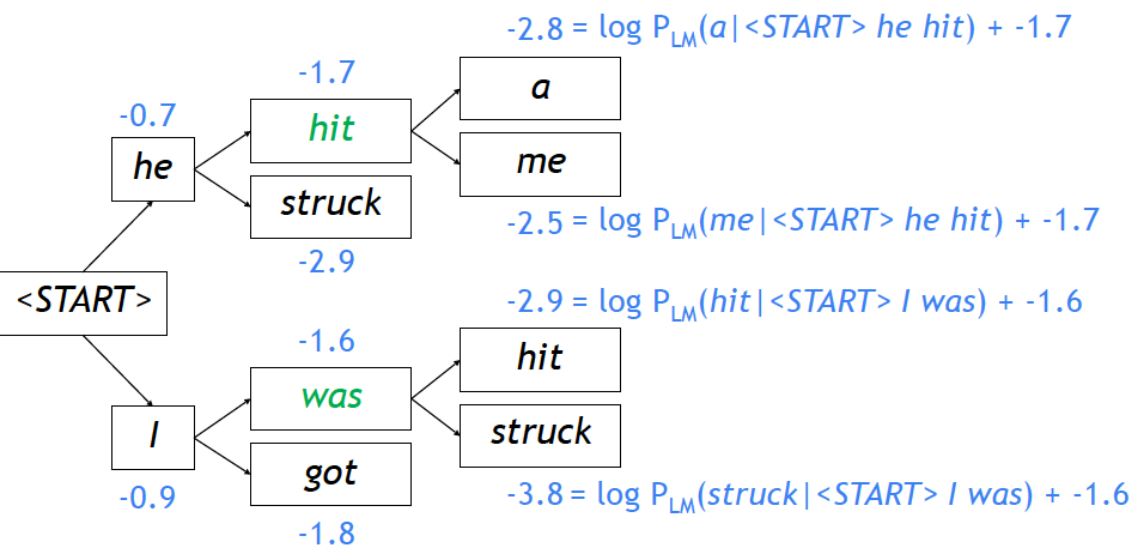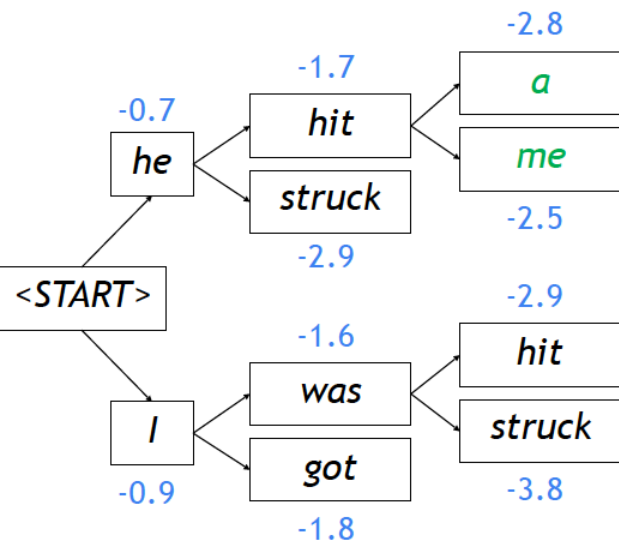
Beam search example: beam size k=2
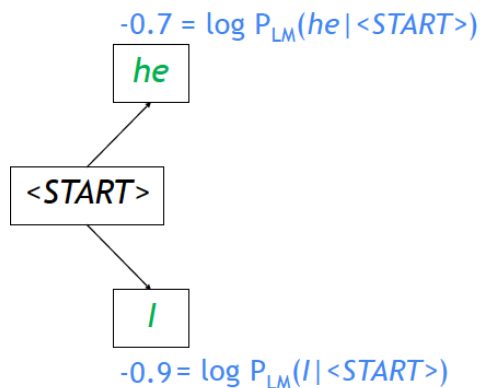
-0.7 = log $P_{LM}(he|<START>)$

he

<START>

I

-0.9 = log $P_{LM}(I|<START>)$

Beam search example: beam size k=2

-1.7 = log P_{LM}(*hit*|*<START> he*) + -0.7

-0.7
he

hit

struck

-2.9= log P_{LM}(*struck*|*<START> he*) + -0.7

<START>

-1.6= log P_{LM}(*was*|*<START> I*) + -0.9

was

I

got

-0.9

-1.8= log P_{LM}(*got*|*<START> I*) + -0.9

Beam search example: beam size k=2

Beam search example: beam size k=2

-2.8 = log $P_{LM}(a|<START> he\ hit)$ + -1.7

-2.5 = log $P_{LM}(me|<START> he\ hit)$ + -1.7

-2.9 = log $P_{LM}(hit|<START> I\ was)$ + -1.6

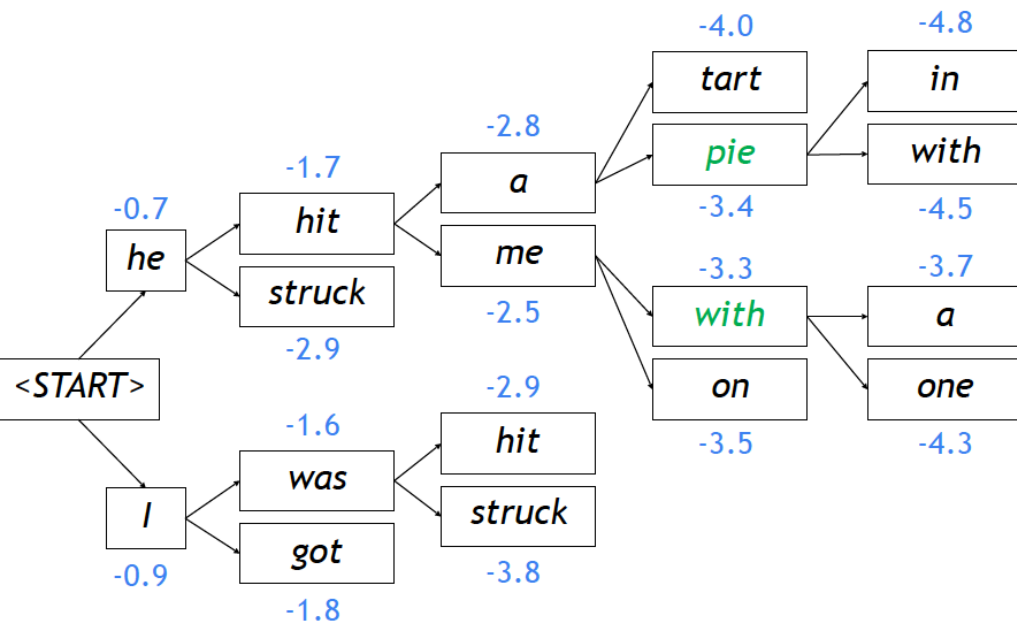-3.8 = log $P_{LM}(struck|<START> I\ was)$ + -1.6
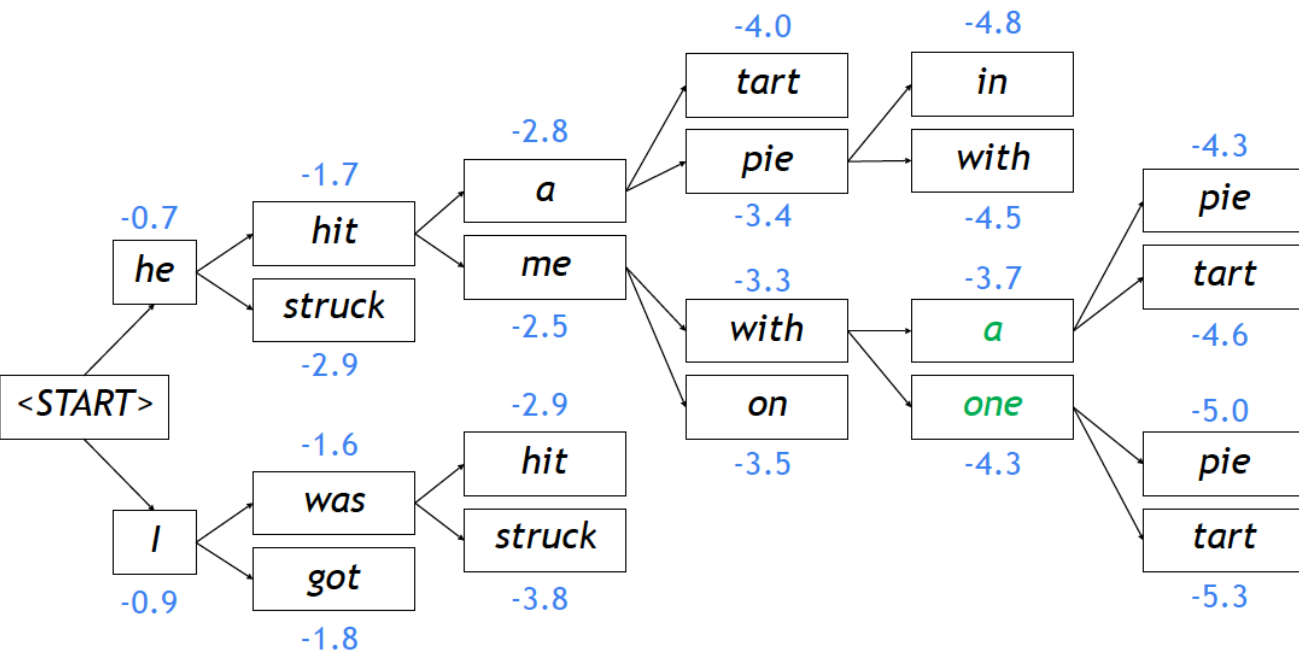
Beam search example: beam size k=2
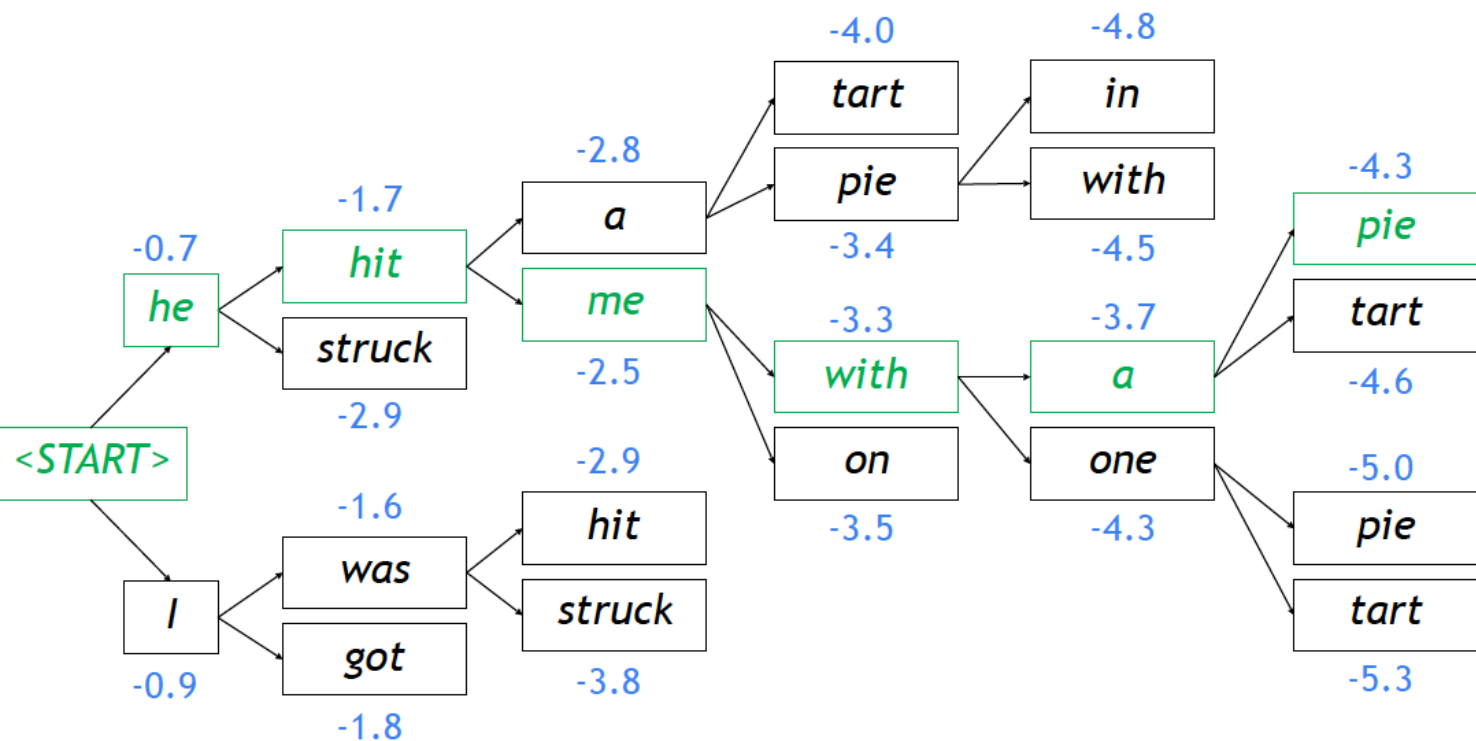
Beam search example: beam size k=2
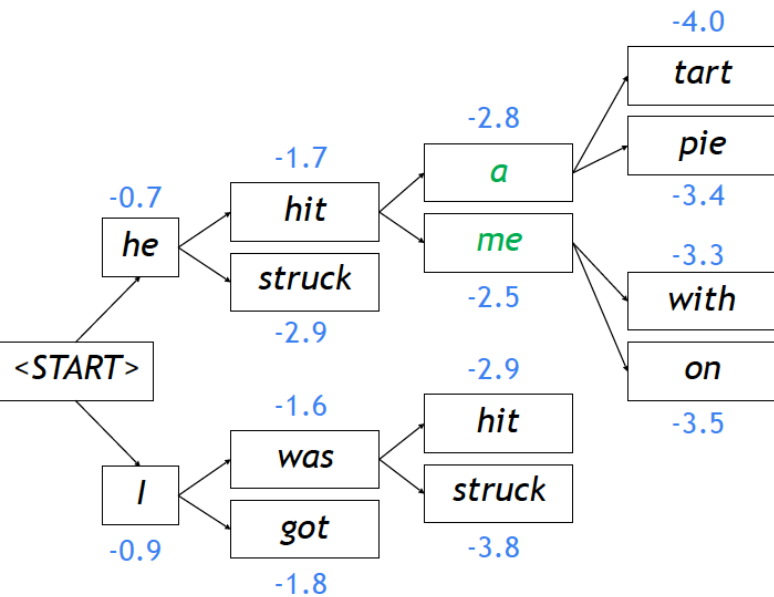
Beam search example: beam size k=2

Beam search example: beam size k=2

Beam search example: beam size k=2

TODOs

- Start thinking about project ideas, and looking for teammates
- First set of two papers' critiques (neural abstractive summarization) are **due on Sep 1, 11:59pm**
  - Half a page to one page using template (https://web.eecs.umich.edu/~wangluxy/courses/eecs598_fa2020/material_eecs598_fa20/critique_template.pdf)
  - If you're on waiting list and don't get enrolled in Canvas on time, you can email me the paper critiques and upload them afterwards.

- The presentation order will be based on random assignment and made available on course website soon.