02/13 Group Study - Lecture 2: Language Modeling

Attendance: Yingjun Guan, Xiaoliang Jiang, Ziqi Jiang, Lan Li, Liri Fang, Yiren Liu **Notification:**

- 1. Scheduled zoom link https://zoom.us/j/400853409 以后都会使用此链接。
- Google calendar
 https://calendar.google.com/event?action=TEMPLATE&tmeid=aXE1MWE5aDlkb2EyNG swa2g0cHRzMmJiM3MgZ3VhbnlpbmdqdW4uZGF0anVAbQ&tmsrc=guanyingjun.datju% 40gmail.com
- 3. Other Useful links: [data mining read group]

Course material:

第二课: language modeling.

- 视频学习
- 浏览课件
- 阅读资料 GoldBerg's NNM4NLP chap8-9
- 如果你对language modeling, feed-forward NN, optimization, language model evaluation的基础 知识不熟悉的话,请移步这里的其他资源:

Discussion Notes:

Lan:

potential related work: [Attention is all you need]

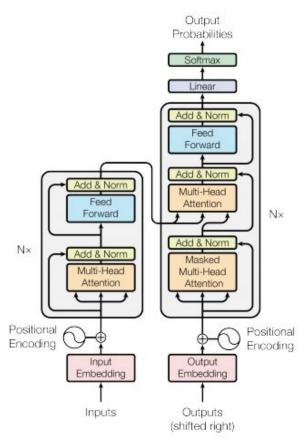


Figure 1: The Transformer - model architecture.

Add on positional parameter

Improve computation efficiency: parallelisation

Page 6: self-attention could help with long-range dependencies

The third is the path length between long-range dependencies in the network. Learning long-range dependencies is a key challenge in many sequence transduction tasks. One key factor affecting the ability to learn such dependencies is the length of the paths forward and backward signals have to traverse in the network. The shorter these paths between any combination of positions in the input and output sequences, the easier it is to learn long-range dependencies [12]. Hence we also compare the maximum path length between any two input and output positions in networks composed of the different layer types.

Xiaoliang:

- 1. Count-based language modeling
- 2. Bow, sequence/sequential data: the word after A and B vs the third word after 1st 2nd and before 4th 5th... -> speed up
- 3. Generative model &&
- 4. Smoothing https://en.wikipedia.org/wiki/Kneser%E2%80%93Ney_smoothing
- 5. Back-off smoothing https://en.wikipedia.org/wiki/Katz%27s back-off model

6. Log加和概率乘积

7. Loglikelihood: 整句, Per-word Log Likelihood: 除词个数, 词,

Ziqi

P26: What are Input Words, Output Words, context?

Liri:

同时生成第3词和第5词 paper:?

Momentum: https://www.youtube.com/watch?v=k8fTYJPd3 I

02/12 CBOW discussion

CBOW: page 117-118 [chap8-9]

CBOW is very similar to the traditional bag-of-words representation in which we discard order information, and works by either summing or averaging the embedding vectors of the corresponding features:

CBOW
$$(f_1, ..., f_k) = \frac{1}{k} \sum_{i=1}^{k} v(f_i).$$
 (8.1)

A simple variation on the CBOW representation is weighted CBOW, in which different vectors receive different weights:

WCBOW
$$(f_1, ..., f_k) = \frac{1}{\sum_{i=1}^k a_i} \sum_{i=1}^k a_i v(f_i).$$
 (8.2)

Here, each feature f_i has an associated weight a_i , indicating the relative importance of the feature. For example, in a document classification task, a feature f_i may correspond to a word in the document, and the associated weight a_i could be the word's TF-IDF score.

02/11 Group study.

Attendence: Yingjun Guan, Liri Fang, Mengfei Lan, Yixing Hu, Yiren Liu, Xiaoliang Jiang, Lan Li

- 1. Self introduction.
- 2. http://phontron.com/class/nn4nlp2020/description.html
- 3. L1 revision. (http://phontron.com/class/nn4nlp2020/assets/slides/nn4nlp-01-intro.pdf)
- 4. L2 revision.
- 5. Phenomena to Handle
 - a. Morphology eg. go/goed

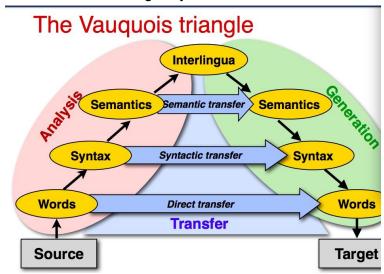
- b. Syntax 句法?
 - i. 主谓宾顺序?
 - ii. 主谓宾缺失?
- c. **Semantics/World Knowledge**: eg. The store went to Jane.
- d. Discourse:
 - i. 1.远距离?答非所问?Q:吃了啥 A:管老师好
 - ii. 2. 代词?

e. Pragmatics

Pragmatic Analysis is part of the process of extracting information from text. Specifically, it's the portion that focuses on taking structures set of text and figuring out what the actual meaning was. It actually comes from the field of linguistics (as a lot of **NLP** does), where the context is considered from the text. \rightarrow google search

[dictionary]: the branch of linguistics dealing with language in use and the contexts in which it is used, including such matters as deixis, the taking of turns in conversation, text organization, presupposition, and implicature.

f. Multilinguality



Lanjie 提到的三角形

Liri Fang: dialect \rightarrow ?

Xiaoliang: 我就来记几个专业名词

- 1. Sentence Classification
- 2. Computation Graphs
- 3. Forward propagation; backward propagation.
- 4. Keras
- 5. AutoGrad

6.

BOW vs CBOW vs deepCBOW

6. Study plan in future:

Time: Tuesday, Thursday 3:30-4:40pm

https://github.com/uiuc-dm-group/DMRG-20SP

第三课: CNN for text

- 视频学习
- 浏览<u>课件</u>
- 阅读材料 GoldBerg's NNM4NLP Chap13
- 如果你对n-gram, CNN, structured convolution, CNN visualization的基础知识不很了解,请移步这里的其他资源