Fake News Detection System using XLNet model with Topic Distributions

Makhin Artyom
317

Lomonosov Moscow State University
CMC MMF

March 2, 2021

COVID-19 Fake News Dataset [1]

Fake news

Public fact verification websites and social media.
The posts are manually verified with the original documents.

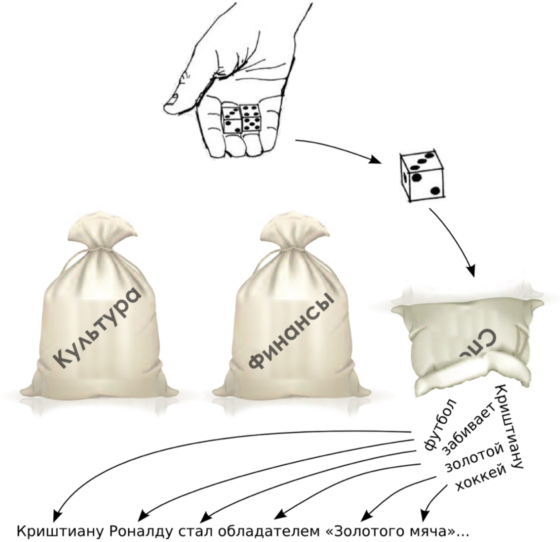
Real news

Official and verified twitter handles of the relevant sources
Each tweets is red by a human and is marked as real news if it contains useful information on COVID-19

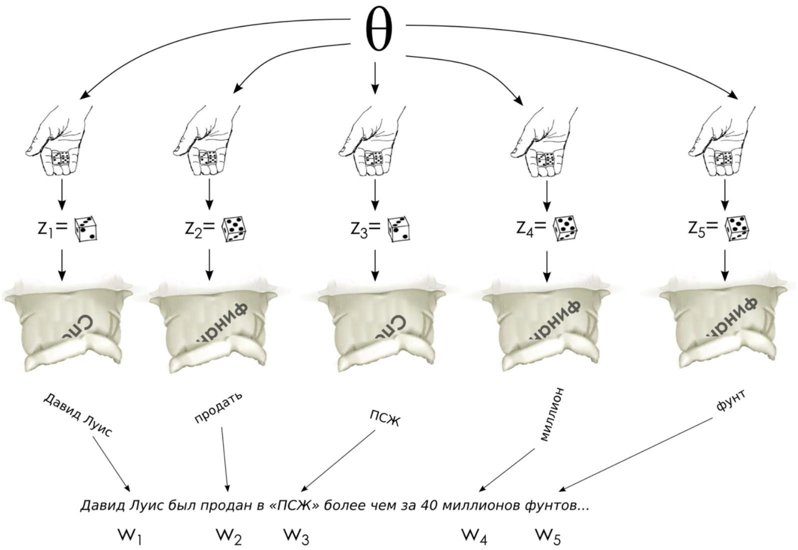
Split	**Real**	**Fake**	**Total**
Train	3360	3060	6420
Validation	1120	1020	2140
Test	1120	1020	2140
Total	5600	5100	10700

[1] https://arxiv.org/pdf/2011.03327.pdf

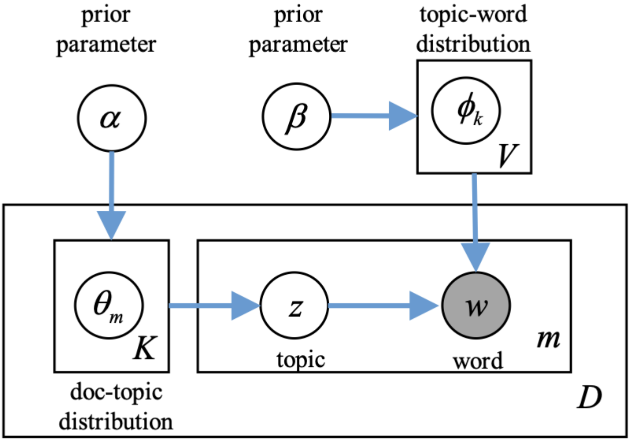
Latent Dirichlet allocation



Latent Dirichlet allocation



LDA Graph Model



Latent Dirichlet allocation

$$p(\mathcal{D} \mid \alpha, \beta) = \prod_{D \in \mathcal{D}} p(D \mid \alpha, \beta) = \prod_{D \in \mathcal{D}} \left(p(c \mid \alpha) \prod_{w \in D} p(w \mid \beta_c) \right)$$

Latent Dirichlet allocation

$$p(\mathcal{D} \mid \alpha, \beta) = \prod_{D \in \mathcal{D}} p(D \mid \alpha, \beta) = \prod_{D \in \mathcal{D}} \left(p(c \mid \alpha) \prod_{w \in D} p(w \mid \beta_c) \right)$$

$$\alpha, \beta = \arg\max_{\alpha, \beta} \prod_{D \in \mathcal{D}} \mathbb{E}_{c|\alpha} \left[p(c \mid \alpha) \prod_{w \in D} p(w \mid \beta_c) \right]$$

Latent Dirichlet allocation

$$p(\mathcal{D} \mid \alpha, \beta) = \prod_{D \in \mathcal{D}} p(D \mid \alpha, \beta) = \prod_{D \in \mathcal{D}} \left(p(c \mid \alpha) \prod_{w \in D} p(w \mid \beta_c) \right)$$

$$\alpha, \beta = \arg\max_{\alpha, \beta} \prod_{D \in \mathcal{D}} \mathbb{E}_{c|\alpha} \left[p(c \mid \alpha) \prod_{w \in D} p(w \mid \beta_c) \right]$$

$$c_i(D) = p(c = i \mid D) = \frac{p(D \mid c = i) p(c = i)}{\sum_j p(D \mid c = j) p(c = j)} = \frac{\alpha_i \prod_{w \in D} \beta_i(w)}{\sum_j \alpha_j \prod_{w \in D} \beta_j(w)}$$

Latent Dirichlet allocation

$$p(\mathcal{D} \mid \alpha, \beta) = \prod_{D \in \mathcal{D}} p(D \mid \alpha, \beta) = \prod_{D \in \mathcal{D}} \left(p(c \mid \alpha) \prod_{w \in D} p(w \mid \beta_c) \right)$$

$$\alpha, \beta = \arg\max_{\alpha, \beta} \prod_{D \in \mathcal{D}} \mathbb{E}_{c|\alpha} \left[p(c \mid \alpha) \prod_{w \in D} p(w \mid \beta_c) \right]$$

$$c_i(D) = p(c = i \mid D) = \frac{p(D \mid c = i) p(c = i)}{\sum_j p(D \mid c = j) p(c = j)} = \frac{\alpha_i \prod_{w \in D} \beta_i(w)}{\sum_j \alpha_j \prod_{w \in D} \beta_j(w)}$$

$$\alpha_i = \frac{\sum_{D \in \mathcal{D}} c_i(D) + \alpha_0}{|\mathcal{D}| + |C| \alpha_0}, \qquad \beta_j(w) = \frac{\sum_{D \in \mathcal{D}} c_j(D) \#\{w \in D\} + \beta_0}{\sum_{D \in \mathcal{D}} c_j(D) + \beta_0 |W|}$$

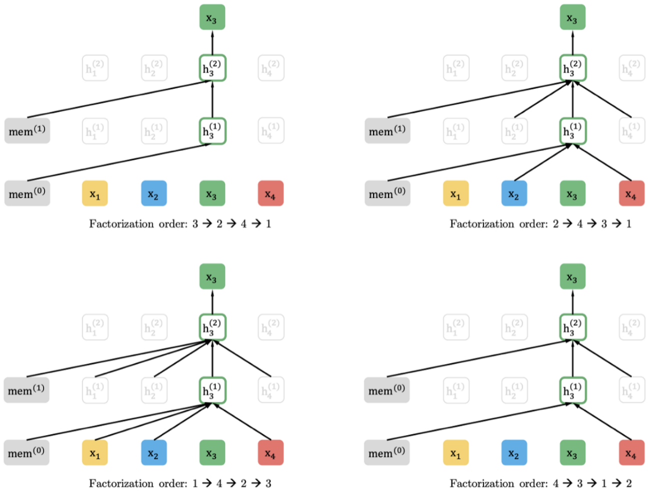


Figure 1: Illustration of the permutation language modeling objective for predicting x_3 given the same input sequence $\mathbf{x}$ but with different factorization orders.

Consider the line "New York is a city" and that we need to predict "New York"

Assume that the current permutation is

XLNET would compute:

log P(New | is a city) + log P(York | New, is a city)

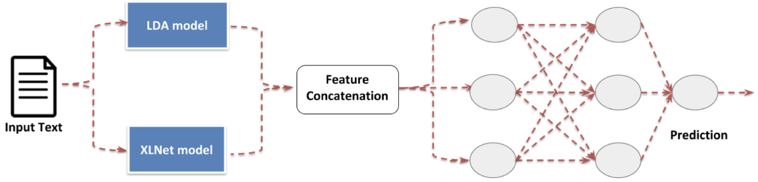
BERT would compute:

log P(New | is a city) + log P(York | is a city)

For each tokenized input text, we construct the following:

- **input ids:** a sequence of integers identifying each input token to its index number in the XLNet tokenizer vocabulary
- **attention mask:** a sequence of 1s and 0s, with 1s for all input tokens and 0s for all padding tokens
- **topic embeddings:** a sequence of probabilities signifies the likelihood of a word in conjunction with a given topic using LDA model
- **labels:** a single value of 1 or 0. In our task, 1 means "Real News," and 0 means "Fake News."

News article is passed through XLNet model to obtain contextualized representations (denoted as CE(·))
The LDA model leveraged to compute the document-topic embeddings (denoted as TE(·))



The concatenated feature representation is passed through 2-fully connected layers followed by a Softmax Layer

$$IE(a_i) = \big[[CE(t), TE(t)]\big| t \in a_i\big]$$

$$y_i = Softmax(IE(a_i))$$

All from sklearn

Model	**Acc**	**P**	**R**	**F1**
DT	85.23	85.31	85.23	85.25
LR	92.76	92.79	92.76	92.75
SVM	**93.46**	**93.48**	**93.46**	**93.46**
GDBT	86.82	87.08	86.82	86.82

Other methods

Method	**Precision**	**Recall**	**F1-score**
Baseline method [14]	0.935	0.935	0.935
USE + SVM	0.92	0.92	0.92
BERT with Topic Distributions	0.949	0.948	0.948
XLNet	0.949	0.948	0.948
Ensemble Approach: BERT and BERT + topic	0.966	0.966	0.966
XLNet with Topic Distributions (Proposed method)	**0.968**	**0.967**	**0.967**

Fake News Detection System using XLNet model with Topic Distributions:

- https://arxiv.org/pdf/2101.11425.pdf

Dataset:

- https://arxiv.org/pdf/2011.03327.pdf
- https://competitions.codalab.org/competitions/26655results (download dataset)

LDA:

- https://arxiv.org/pdf/2010.04391.pdf
- https://habr.com/ru/company/surfingbird/blog/230103/
- https://habr.com/ru/company/surfingbird/blog/228249/

XLNet:

- https://towardsdatascience.com/xlnet-explained-in-simple-terms-255b9fb2c97c
- https://habr.com/ru/post/536692/

USE:

- https://tfhub.dev/google/universal-sentence-encoder-large/3