Overview of the Multi-Task Mutual Learning Technique: A Comparative Analysis of Different Models for Sentiment Analysis and Topic Detection

Matteo Posenato

December 12, 2023

Università degli Studi di Padova

Outline

Introduction

Università degli Studi di Padova

- Digital Era Impact: Surge in social media and communication platforms, e.g., Twitter (467 million users, 175 million daily tweets).
- Significance of Data Influx: Overwhelming volume of data highlights the pivotal role of Natural Language Processing (NLP).
- Presentation Focus: Explore how NLP extracts insights from substantial textual data in the digital era.

Introduction

- NLP Definition: Computational techniques for analyzing and representing natural language, applied to specific tasks.
- Significance of Topic Detection: A key application of NLP, automatically extracts relevant themes from extensive datasets.
- Sentiment Analysis in the Era of Big Data: Opinion Mining and Sentiment Analysis categorize opinions, providing valuable insights into the public mood.

MTL with Mutual learning

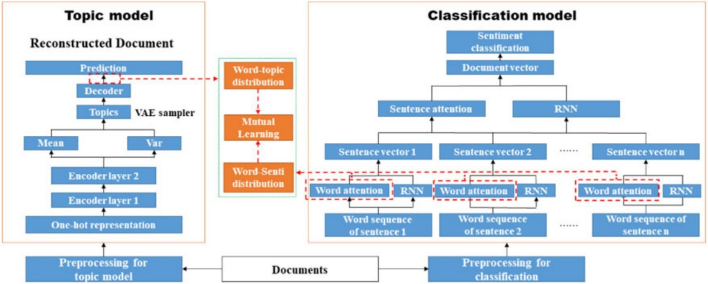


Fig. 4. Our proposed MTL with mutual learning for sentiment classification and topic detection.

- two different models are trained simultaneously
- the models share knowledge and their prediction during the training phase

Objective of this thesis

- Understand how the multi-task with mutual learning works and its behaviour with respect with different datasets
- See if changing the model inside of this technique can improve the performances

Datasets

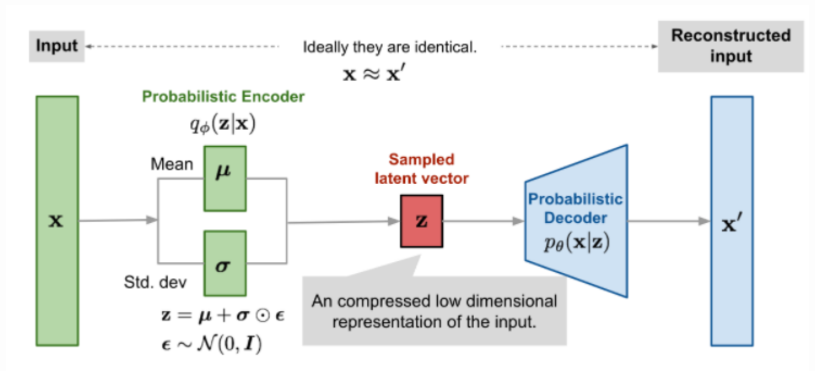
The two datasets are YELP and IMDB, the first one contains reviews of businesses, the second one contains movie's reviews

	#Class	#docs	Vocab.Size
YELP	5	39,923	53,823
IMDB	10	15,000	55,819

Table: Datasets

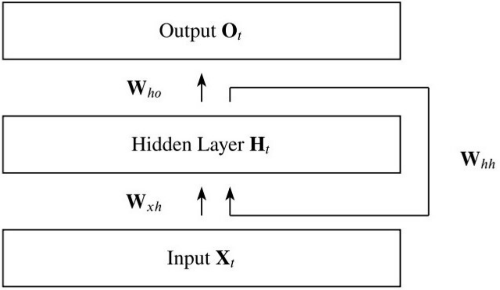
Variational Autoencoder

Variational autoencoders (VAEs) are a type of autoencoder that introduces regularization during training to ensure a regular and interpretable latent space.



Recurrent Neural Network

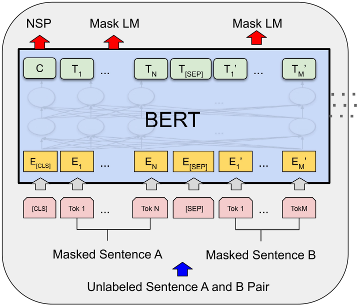
RNNs represent a fascinating class of neural network architectures designed primarily for identifying patterns within sequential data



Recurrent Neural Network

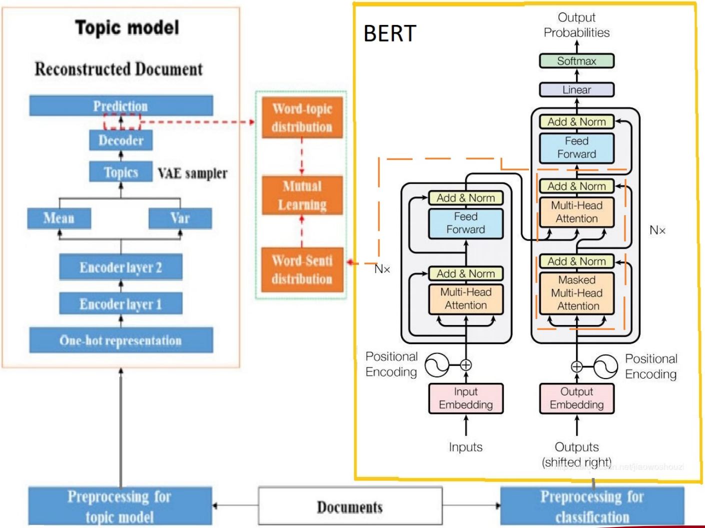
BERT

- Developed by Google
- It utilizes a transformer architecture
- Enabling bidirectional contextual understanding of words in a given text



Pre-training

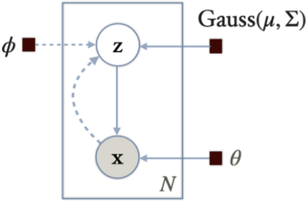
MTL with BERT



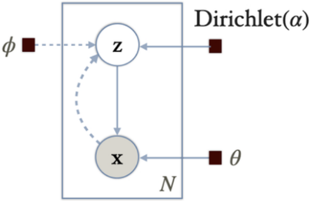
Dirchlet Variational Autoencoder

Università degli Studi di Padova

Dirichlet Variational Autoencoders (DirVAE) is an extension of the traditional Variational Autoencoder (VAE) that incorporates Dirichlet priors to model latent variable distributions.

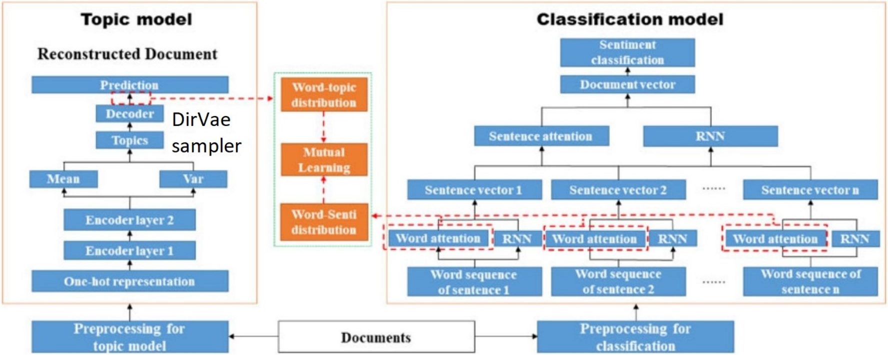


(a) standard VAE



(b) DirVAE

MTL with DirVae



complexity

VAE + RNN								
	IMDB					YELP		
batch size	10	25	50	100	500	25	50	100
time(h)	6h	5h	4h	5h 30 min.	9h 15 min.	7h	5h 50 min	6h

Table: Complexity VAE + RNN

IMDB		
model	VAE + BERT	DirVAE + RNN
batch size	50	50
time(h)	4h	4h 10 min

Table: BERT-DirVAE

VAE + RNN						
	IMDB			YELP		
vocab size	2000	3000	5000	2000	3000	5000
batch size	50	50	50	50	50	50
accuracy	0,17	0,192	0,127	0,229	0.227	0,227

Table: Accuracy VAE + RNN

YELP		
model	VAE + BERT	DirVAE + RNN
batch size	50	50
vocab size	3000	3000
accuracy	0,209	0,274

Table: BERT-DirVAE

topic detection loss

VAE + RNN						
	IMDB			YELP		
vocab size	2000	3000	5000	2000	3000	5000
batch size	50	50	50	50	50	50
topic loss	97,2639	34,4885	83,8601	12,3715	3,344	12,6933

Table: Topic Loss VAE + RNN

YELP		
model	VAE + BERT	DirVAE + RNN
batch size	50	50
vocab size	3000	3000
topic loss	64,6515	20,3235

Table: BERT-DirVAE

classification loss

VAE + RNN						
	IMDB			YELP		
vocab size	2000	3000	5000	2000	3000	5000
batch size	50	50	50	50	50	50
classification loss	2,2078	2,1559	2,1086	1,6148	1,6305	1,6096

Table: Classification Loss VAE + RNN

IMDB		
model	VAE + BERT	DirVAE + RNN
batch size	50	50
vocab size	3000	3000
classification loss	2,3015	2,1459

Table: BERT-DirVAE

KLD

VAE + RNN						
	IMDB			YELP		
vocab size	2000	3000	5000	2000	3000	5000
batch size	50	50	50	50	50	50
kld	0,0031	0,0011	0,0009	0,0006	0,0002	0,0002

Table: KLD VAE + RNN

YELP		
model	VAE + BERT	DirVAE + RNN
batch size	50	50
vocab size	3000	3000
kld	0,0060	0,0001

Table: BERT-DirVAE

Conclusion

- Summary of Performance:
  - All performance metrics are conditioned by our computing power.
  - The impact of vocabulary size is most evident in smaller datasets.
  - The standard model and its variants have similar results
- Conclusions:
  - Enormous computing power is required
  - there is no evidence to say that one model is better than others
  - this technique is very adaptable, being able to change the models within it according to our needs
- Further works:
  - Implement some metrics to help better understand, such as the NMI measuring the robustness of topics
  - Try to apply this technique to other NLP applications