From Fully Supervised to Zero Shot Settings for Twitter Hashtag Recommendation

Abhay Kumar, Nishant Jain, Chirag Singh, Suraj Tripathi

Samsung R&D Institute India – Bangalore

Introduction

- We propose a comprehensive end-to-end pipeline for Twitter hashtags recommendation system including data collection, supervised training setting and zero shot training setting.
- Deep learning architectures for Supervised Setting → Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and Transformer Network.
- However, it is not feasible to collect data for all possible hashtag labels and train a classifier model on them.
- To overcome this limitation, we propose a Zero Shot Learning (ZSL) paradigm for predicting unseen hashtag labels by learning the relationship between the semantic space of tweets and the embedding space of hashtag labels.

Methods

Zero Shot Learning: In Zero-shot learning setting, we train a classifier from labelled training exemplars from seen classes and learn a mapping from input feature space to semantic embedding space. ZSL aims to classify class labels, which were never exposed during training pipeline.

Few Shot Learning: Few Shot Learning (FSL) paradigm is based on feeding a limited amount of training data. This is an extension of ZSL setting, where few examples of unseen (in ZSL setting) class labels are also exposed during the training process.

- **ConSE: Convex Combination of Semantic Embeddings:**
  - The ConSE [1] method extends the prediction probability beyond the seen hashtags, to a set of unseen hashtag labels. The embedding vector of the unseen hashtag for a test tweet x is predicted by a convex combination of seen hashtag embedding vectors weighted by their corresponding probabilities as shown in Eq. 1.

$$f(x) = \frac{1}{Z}\sum_{t=1}^{T} p_\theta(\hat{y}(x,t)|x).s(\hat{y}(x,t)) \quad (1)$$

  - where, $\hat{y}(x,t)$ denotes the t^{th} most probable training hashtag label for a tweet x, and $s(\hat{y}(x,t))$represents the embedding vector of $\hat{y}(x,t)$.
  - cosine similarity is used to find the most likely hashtag from unseen hashtag label as shown-

$$\hat{y}(x,1) = \underset{\acute{y} \in Y_U}{\operatorname{argmax}} \cos(f(x), s(\acute{y})) \quad (2)$$

- **ESZSL: Embarrassingly Simple Approach to Zero Shot Learning:**
  - ESZSL [2] approach is a general framework which models the relationship among features, attributes, and class labels by formulating it as a two linear layers network.
  - We minimize the following loss function to learn a linear predictor $\mathcal{W}$ from all m training examples.

$$\underset{\mathcal{W} \in \mathbb{R}^{p \times s}}{\text{minimize}} \ \mathcal{L}(\mathcal{X}^T \mathcal{W} \mathcal{A},\ \mathcal{Y}) + \ \Phi(\mathcal{W}) \quad (3)$$

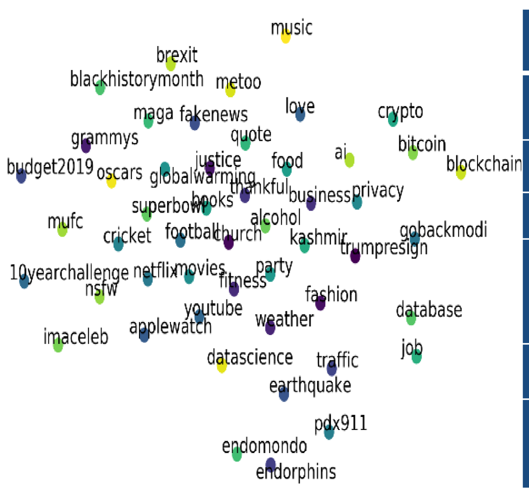
  - The closed form solution for *Eq.* 3 is expressed as-

$$\mathcal{W} = \ (\mathcal{X}\mathcal{X}^T + \gamma I)^{-1} \mathcal{X}\mathcal{Y}\mathcal{A}^T (\mathcal{A}\mathcal{A}^T + \gamma I)^{-1} \quad (4)$$

- **DEM-ZSL: Deep Embedding Model for Zero Shot Learning:**

DEM-ZSL [3] is end-to-end learning of deep embedding model with two branches. One branch learns the semantic space of the tweet. The other branch learns the semantic representation of the hashtag class labels. These 150 dimension embedding vectors of hashtag labels are mapped into 1024-dimensional semantic space. Finally, we minimize the least square error to reduce the discrepancy between tweet semantic features and the mapped 1024-dimensional semantic vector of hashtag labels.

Dataset



Criteria	Count
Total Collected tweets	805,031
Cleaned sampled tweets	321,385
Tweets containing hashtags	31,249
Tweets containing more than one hashtag	3,301
Unique hashtags	956
Hashtags with at least 200 unique tweets	50

Results

Models	Accuracy (%)	Precision	Recall	f1 score
ANN (Baseline)	40.7	0.41	0.41	0.41
CNN	43.7	0.44	0.44	0.44
RNN	46.9	0.47	0.47	0.47
RNN with Attention	47.0	0.47	0.47	0.47
Transformer Network	57.4	0.57	0.57	0.57

Table 1. Experimental results for hashtag recommendation in fully supervised setting

Seen/Unseen Hashtags	ZSL Model	Zero Shot Setting (in percentage)			Few Shot Setting (in percentage)		
		hit@1	hit@2	hit@5	hit@1	hit@2	hit@5
40/10	ConSE	49	65	88	54	66	89
	ESZSL	60	73	93	64	75	93
	DEM	62	74	96	73	83	97
30/20	ConSE	27	36	59	43	50	67
	ESZSL	35	48	71	46	56	73
	DEM	42	54	71	61	71	84
25/25	ConSE	24	30	45	38	47	62
	ESZSL	23	35	61	40	49	66
	DEM	29	42	59	58	69	82

Table 2. Experimental results for hashtag recommendation in ZSL and FSL settings

Cleaned Tweets	Expected Hashtag	Top 5 Predicted Hashtags
high time girls women troubled under user should come out on there are cases of girls exploited need to give courage to those souls to take up on him	#metoo	#metoo, #justice, #netflix, #privacy, #fakenews
fact: The uncertainty of a parameter estimate goes to zero as the sample size approaches infinity. The variability of a parameter estimate does not.	#datascience	#ai, #datascience, #church, #privacy, #bitcoin
So it's obvious the refs don't want us to win, but guess what bitches, here we come champs	#superbowl	#superbowl, #church, #fitness, #movies, #thankful
History of gun violence is long & still unresolved & unsolved for many victims & those seeking	#justice	#trumpresign, #justice, #fakenews, #metoo, #church
wtf why should we fill up their pockets, anybody who is paying to watch these 2 is a direct victim of msm	#fakenews	#fakenews, #movies, #metoo, #privacy, #trumpresign

Table 3. Hashtag recommendation results for few tweet examples

Conclusions

- Experimented with various deep learning based model for Twitter hashtag recommendation both in the supervised and zero-shot setting.
- The ZSL models can predict unseen hashtags, even if those hashtags are not ex-posed in the training phase. These ZSL models learn the mapping from semantic space of tweets to hashtags embedding space.
- For determining the semantic space of tweets, we have implemented CNN and RNN based encoders and Hashtag embedding is learnt by fine-tuning word2vec model on twitter data.

References

1. *Norouzi, M., Mikolov, T., Bengio, S., Singer, Y., Shlens, J. et. al.: Zero-shot learning by convex combination of semantic embeddings. arXiv preprint arXiv:1312.5650, 2013.*
2. *Romera-Paredes, B. and Torr, P.: An embarrassingly simple approach to zero-shot learning. In: International Conference on Machine Learning, pp. 2152-2161, 2015.*
3. *Zhang, L., Xiang, T. and Gong, S.: Learning a deep embedding model for zero-shot learning. In: Proceedings of the IEEE CVPR, pp. 2021-2030, 2017.*

SAMSUNG Research