

# From Fully Supervised to Zero Shot Settings for Twitter Hashtag Recommendation

**Abhay Kumar, Nishant Jain, Chirag Singh, Suraj Tripathi**

**Paper# 325**

**Presented by  
Abhay Kumar**

# 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.
- 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.

# Proposed Approach

- **Zero Shot Learning:** In Zero-shot learning setting, we train a classifier from la-beled training exemplars from seen classes and learn a mapping from input fea-ture space to semantic embedding space. ZSL aims to classify class labels, which were never exposed during training pipeline. ZSL methods used are-
  1. **ConSE: Convex Combination of Semantic Embeddings**
  2. **ESZSL: Embarrassingly Simple Approach to Zero Shot Learning**
  3. **DEM-ZSL: Deep Embedding Model for Zero Shot Learning**
- **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.

# Results

- Title & Authors
- Introduction
- Proposed Approach
- Results**
- Poster Screenshot

Cleaned Tweets	Expected Hashtag	Top 5 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

# Poster Screenshot

Title & Authors  
Introduction  
Proposed Approach  
Results  
Poster Screenshot

## 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



**SAMSUNG**

### Introduction

- We propose a comprehensive end-to-end pipeline for Twitter hashtag 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 producing 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 labeled training examples 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 finding 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.

**CoSE: Coarse Combination of Semantic Embeddings:**

- The CoSE [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 produced by a convex combination of seen hashtag embedding vectors weighted by their corresponding probabilities as shown in Eq. 1.

$$f(x) = \sum_{i=1}^k p_i \hat{e}_i(x, \mathcal{H}_s) \hat{e}_i^T(x, \mathcal{H}_s) \quad (1)$$

- where,  $\hat{y}(x, \mathcal{H}_s)$  denotes the  $i^{th}$  most probable training hashtag label for a tweet  $x$ , and  $\hat{e}_i^T(x, \mathcal{H}_s)$  represents the embedding vector of  $\hat{y}(x, \mathcal{H}_s)$ .
- $\hat{y}(x, \mathcal{H}_s)$  is used to find the most likely hashtag from unseen hashtag label as shown.

$$\hat{y}(x, \mathcal{H}_s) = \underset{y \in \mathcal{H}_u}{\text{argmax}} \{f(y, x, \mathcal{H}_s)\} \quad (2)$$

**ESZSL: Embarrassingly Simple Approach to Zero Shot Learning:**

- ESZSL [2] approach is a general framework which models the relationship among instance, attributes, and class labels by formulating it as a two-layer neural network.
- We minimize the following loss function to learn a linear predictor  $W$  from all our training examples.

$$\mathcal{L}(X, Y, W, \mathcal{H}_s, \mathcal{H}_u) = \mathcal{L}(W) \quad (3)$$

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

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

**DEEML: Deep Embedding Model for Zero Shot Learning:**

DEEML [3] is an 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-dimensional 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.

### Results

Model	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	57.8	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)		
		h@1	h@2	h@3	h@1	h@2	h@3
40/10	CoSE	49	43	38	34	36	39
	ESZSL	60	73	80	64	75	83
	DEEML	62	74	86	73	83	97
30/70	CoSE	27	36	39	43	50	67
	ESZSL	35	48	71	46	56	73
	DEEML	42	54	71	61	71	84
20/20	CoSE	24	30	45	39	47	62
	ESZSL	23	35	44	40	49	66
	DEEML	29	42	59	58	69	82

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

Closed Tweets	Expected Hashtag	Top 3 Predicted Hashtags
High class girls women revealed under wear should come out on there are loads of girls who should be in your coverage to show them to later up on line	sexwork	sexwork, #sexwork, #sexwork
Just: The uncertainty of a parameter estimate goes up as $n$ of the sample size approaches infinity. The variability of a parameter estimate does not	statistics	the, #statistics, #statistics, #statistics
So it's obvious the right age I want to see in this that goes what happens, like come change	superhero	superhero, #superhero, #superhero, #superhero
History of gay villages is deep & still unresolved & sensitive for many visitors & their visiting	gayvillage	gayvillage, #gayvillage, #gayvillage, #gayvillage
we who should we fill up their positions, which will be asking for much more? it is a direct result of jobs	hikemans	hikemans, #hikemans, #hikemans, #hikemans

Table 3. Hashtag recommendation results for few tweet examples

### Conclusions

- Experimented with various deep learning based models 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 in-pool in the training phase. These ZSL models learn the mapping from semantic space of tweets to hashtag embedding space.
- For determining the semantic space of tweets, we have experimented CNN and RNN based models and Hashtag embedding is learned by fine-tuning word2vec model on twitter data.

### References

1. Nishant, M., Mishra, T., Zang, Z., Zang, T., Shukla, J. et al. Zero-shot learning by neural embeddings of semantic embeddings. *arXiv preprint arXiv:1812.03161*, 2018.
2. Kumar, A., Jain, N., and Singh, C. An embarrassingly simple approach to zero-shot learning in deep neural networks. In *Proceedings of the Conference on Machine Learning*, pp. 2152-2161, 2015.
3. Zhang, L., Jiang, J., and Gong, J. Learning a deep embedding model for zero-shot learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 2012-2016, 2017.

**SAMSUNG** Research