

cases	doc_1		doc_2		decision	id
	authors	<ul style="list-style-type: none">J GaneshManish GuptaVasudeva Varma	authors	<ul style="list-style-type: none">Ganesh JawaharManish GuptaVasudeva Varma	DUPLICATES	371
	title	Interpretation of Semantic Tweet Representations	title	Interpretation of Semantic Tweet Representations		
	publication_date	2017-04-04 07:24:15+00:00	publication_date	2017-04-04 00:00:00		
	source	SupportedSources.ARXIV	source	SupportedSources.SEMANTIC_SCHOLAR		
	journal	None	journal			
	volume		volume			
	doi		doi	10.1145/3110025.3110083		
	urls	<ul style="list-style-type: none">http://arxiv.org/pdf/1704.00898v2http://arxiv.org/abs/1704.00898v2http://arxiv.org/pdf/1704.00898v2	urls	<ul style="list-style-type: none">https://www.semanticscholar.org/paper/af52ac2e85d190f2c80cc24f6cbf03b2f4fd6caa		
	id	id-2477424712836775733	id	id-2607072339537773600		
	abstract	Research in analysis of microblogging platforms is experiencing a renewed surge with a large number of works applying representation learning models for applications like sentiment analysis, semantic textual similarity computation, hashtag prediction, etc. Although the performance of the representation learning models has been better than the traditional baselines for such tasks, little is known about the elementary properties of a tweet encoded within these representations, or why particular representations work better for certain tasks. Our work presented here constitutes the first step in opening the black-box of vector embeddings for tweets. Traditional feature engineering methods for high-level applications have exploited various elementary properties of tweets. We believe that a tweet representation is effective for an application because it meticulously encodes the application-specific elementary properties of tweets. To understand the elementary properties encoded in a tweet representation, we evaluate the representations on the accuracy to which they can model each of those properties such as tweet length, presence of particular words, hashtags, mentions, capitalization, etc. Our systematic extensive study of nine supervised and four unsupervised tweet representations against most popular eight textual and five social elementary properties reveal that Bi-directional LSTMs (BLSTMs) and Skip-Thought Vectors (STV) best encode the textual and social properties of tweets respectively. FastText is the best model for low resource settings, providing very little degradation with reduction in embedding size. Finally, we draw interesting insights by correlating the model performance obtained for elementary property prediction tasks with the highlevel downstream applications.	abstract	Research in analysis of microblogging platforms is experiencing a renewed surge with a large number of works applying representation learning models for applications like sentiment analysis, semantic textual similarity computation, hashtag prediction, etc. Although the performance of the representation learning models has been better than the traditional baselines for such tasks, little is known about the elementary properties of a tweet encoded within these representations, or why particular representations work better for certain tasks. Our work presented here constitutes the first step in opening the black-box of vector embeddings for tweets. Traditional feature engineering methods for high-level applications have exploited various elementary properties of tweets. We believe that a tweet representation is effective for an application because it meticulously encodes the application-specific elementary properties of tweets. To understand the elementary properties encoded in a tweet representation, we evaluate the representations on the accuracy to which they can model each of those properties such as tweet length, presence of particular words, hashtags, mentions, capitalization, etc.Our systematic extensive study of nine supervised and four unsupervised tweet representations against most popular eight textual and five social elementary properties reveal that Bi-directional LSTMs (BLSTMs) and Skip-Thought Vectors (STV) best encode the textual and social properties of tweets respectively. FastText is the best model for low resource settings, providing very little degradation with reduction in embedding size. Finally, we draw interesting insights by correlating the model performance obtained for elementary property prediction tasks with the high- level downstream applications.		
	versions		versions			