Measuring, Predicting and Visualizing Short-Term Change in Word Representation and Usage in VKontakte Social Network

Ian Stewart
Georgia Institute of Technology
801 Atlantic Dr NW
Atlanta, GA 30332

Dustin Arendt, Eric Bell, Svitlana Volkova
Pacific Northwest National Laboratory
902 Battelle Blvd
Richland, WA 99354

Abstract

Language in social media is extremely dynamic: new words emerge, trend and disappear, while the meaning of existing words can fluctuate over time. This work addresses several important tasks of visualizing and predicting short term text representation shift, i.e. the change in a word's contextual semantics. We study the relationship between short-term concept drift and representation shift on a large social media corpus – VKontakte collected during the Russia-Ukraine crisis in 2014 – 2015. We visualize short-term representation shift for example keywords and build predictive models to forecast short-term shifts in meaning from previous meaning as well as from concept drift. We show that short-term representation shift can be accurately predicted up to several weeks in advance and that visualization provides insight into meaning change. Our approach can be used to explore and characterize specific aspects of the streaming corpus during crisis events and potentially improve other downstream classification tasks including real-time event forecasting in social media.

Introduction

Social media have been widely studied as sensors of human behavior to track unusual or novel activities in real time all over the globe (Alsaedi, Burnap, and Rana 2016; Asur and Huberman 2010). Much analysis of social media language focuses on surface-level features and patterns, like word frequency, to improve real-time event detection and tracking e.g., during crisis events (Bruno 2011; Crooks et al. 2013). These surface features provide a shallow signal into human behavior but miss some of the more subtle variations. Tracking emerging words only based on their virality and frequency trends (Mathioudakis and Koudas 2010; Weng, Menczer, and Ahn 2013) would miss the change in word meaning for existing words, or the meaning of newly emerging terms in social media.

For example, during the Russian-Ukrainian crisis the word *ukrop*, meaning *dill*, changed its meaning to *Ukrainian patriot* and developed a more negative connotation over time. Recent work has effectively tracked word meaning over time at scale, but it often examines long-term meaning shift within formal written text such as the Google Books corpus (Gulordava and Baroni 2011), rather than short-term shift within more informal contexts such as social media (Kulkarni et al. 2015).

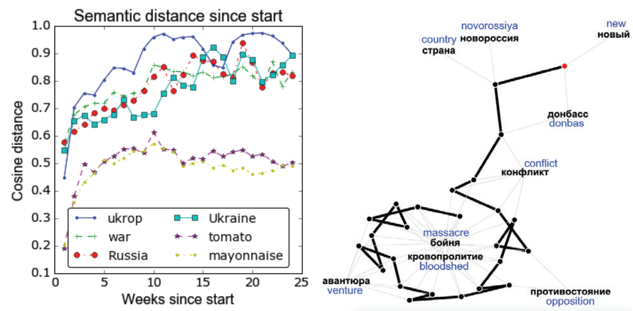

Figure 1: Representation shift between each word's current representation and its original representation (left). Semantic trajectory of the word *war* over time, projected in 2D, with two most similar words at each timestamp (right).

This study explores short-term representation shift using a corpus of VKontakte posts collected during the Russia-Ukraine crisis, which has been noted as a source of political instability (Volkova et al. 2016) and thus linguistic unpredictability. We develop an approach for predicting and tracking short-term shifts in word meaning or *representation*, utilizing a word's previous meaning as well as its change in frequency, or *concept drift*. Systems that can automatically assist analysts during dynamic events, such as the Russia-Ukraine crisis, could bring us much closer to understanding issues as they happen on the ground.

Our study makes the following novel contributions:

- We show that short-term representation shift can be predicted from prior shift and concept drift.
- We propose novel visual representations to track the development of word semantics in social media.

To motivate the study with the earlier *ukrop* example, we present in Figure 1 (left) an example of representation shift in a set of keywords drawn from our data. The y-axis measures each word's cosine distance from its initial representation. First, note the split between the semantically stable food words (i.e. minimal cosine distance) and the more dynamic conflict words that become increasingly distant from their original representation. Moreover, we see that some of the upper words such as *ukrop* exhibit especially dynamic behavior, alternatively growing farther and closer to its original meaning from weeks 11 to 17. This shift was likely the result of a split in meaning as *ukrop* lost its literal meaning in favor of its slang pejorative meaning.

We draw further motivation in Figure 1 (right) which illustrates the trajectory of representation shift in the word *war* starting from the upper-right red dot and progressing toward the end of the line in the lower-left. Projected into 2 dimensions, the embedding of *war* begins close to the embeddings of situation-specific words such as *Donbas* (location of conflict) and cycles toward more violent words such as *massacre*. This context shift appears to be a kind of semantic narrowing (Sagi, Kaufmann, and Clark 2009) toward more negative or pejorative words that is captured in the decreasing shift distances in the later timesteps.

Background

Our work leverages techniques from distributional semantics (Harris 1954) to approximate the change in word representation over time. Kim et al. (2014) propose a novel approach by measuring change in English word semantics across the 20th century by comparing each word's initial meaning (measured by its embedding) with its meanings in later years. Further studies adopt this methodology to propose laws of semantic change relating concept drift to representation shift (Hamilton, Leskovec, and Jurafsky 2016b) as well as to separate insignificant from significant linguistic change across domains (Kulkarni et al. 2015).

Our work builds on prior studies by first tracking semantic change within a non-English language and in the noisy domain of social media rather than non-social corpora like Google Books (Hamilton, Leskovec, and Jurafsky 2016b; Kim et al. 2014). In addition, we look to highlight more subtle, short-term changes in connotation, such as the added pejorative connotation of *ukrop* (*dill*) as it became a negative descriptor for *Ukrainian patriot*. Lastly, our study is among the first to build predictive models to forecast representation shift rather than characterizing it (Kulkarni et al. 2015).

Data

We rely on public data from the VKontakte (VK) social network, a popular website in Russia and Ukraine similar to Facebook. The data was collected[1] over a period of 25 weeks between Sept 2014 and Mar 2015, and comprise over 600K posts, with an average of 167 words per post.

The VKontakte data provides an ideal testbed for our study of representation shift because it was collected during a volatile period in Russia and Ukraine that led to sudden language change, such as the adoption of new meaning for words like *ukrop*. Nonetheless, our methods can apply to any active social media platform that uses primarily text data for user interaction, such as Twitter or Facebook.

Following standard practices, we first stem all words in the data using the Russian morphology package PyMorph[2] and lower-case the words to match social media's irregular capitalization practices. We then collect all unigrams with a post frequency of 5 and above to avoid the long tail of misspellings and irrelevant words, leaving us with a vocabulary V of 60,000 words. We remove stop-words from frequency counts but retain them in the word vector vocabulary in order to preserve the context afforded by stop-words.

Approach

Word Dynamics Frequency-based methods can capture linguistic shift because changes in word frequency often correspond to words gaining or losing senses (Hamilton, Leskovec, and Jurafsky 2016b). Thus, we first extract tf-idf score χ, to capture changes in word usage in social media without over-representing words with consistently high frequency: $\chi_{w,t} = log(count(w,t)) \times log\frac{|P_t|}{|p \in P_t : w \in p|}$ where $|P|$ is the total number of posts and $|p_t \in P : w \in p_t|$ is the number of posts at time t where the word w appears.

The tf-idf scores for each word w are concatenated chronologically to form time series $\tau_\chi(w)$, which represents our measure of concept drift.

Temporal Embeddings To learn temporal word embeddings, we applied word2vec models (Mikolov et al. 2013) implemented in gensim[3] (Řehůřek and Sojka 2010) that have been used for a variety of applications related to tracking semantic change (Hamilton, Leskovec, and Jurafsky 2016a; 2016b; Kim et al. 2014; Kulkarni et al. 2015). We chose word embeddings for their simple mathematical representation and well-understood usage in encoding meaning.

We initialize a word2vec model with vocabulary V, then train the model with tokenized posts for each timestep (i.e. week), using as a baseline the embeddings trained at the previous timestep. This guarantees that the dimensionality remains consistent across weeks and allows us to reliably track a word's representation shift through time (Kim et al. 2014).

For each word w in the vocabulary, at each week t between 1 and 25, we generate an embedding vector to yield a time series $\tau_e(w)$: $\tau_e(w) = e_{t_0}(w), e_{t_1}(w), \ldots, e_T(w)$.

To build the embeddings, we chose a dimensionality of 30 for a fairly coarse-grained representation, avoiding data sparsity. We used standard training hyperparameters (e.g., window size 5), following prior experiments in building word vectors from social media data (Kulkarni et al. 2015).

Differencing Statistics Our predictive tests require us to compare representation shift with concept drift, and we therefore compute the first-order differences for all statistics. For each statistic s in $\tau_\chi(w)$ and $\tau_e(w)$ over all words in vocabulary of size N, over the course of T timesteps, we calculate a vector: $\Delta\tau_s(w) = \Delta s_{t_0,t_1}(w) \ldots \Delta s_{T-1,T}(w)$, where we compute $\Delta\tau_\chi(w)$ with subtraction and $\Delta\tau_e(w)$ with cosine distance between embeddings.

Results

We present example visualizations of representation shift in existing words, and then demonstrate the results of predict-

[1]The data was collected when one of the authors was affiliated with Johns Hopkins University.

[2]https://pymorphy2.readthedocs.io/en/latest/

[3]https://pypi.python.org/pypi/gensim

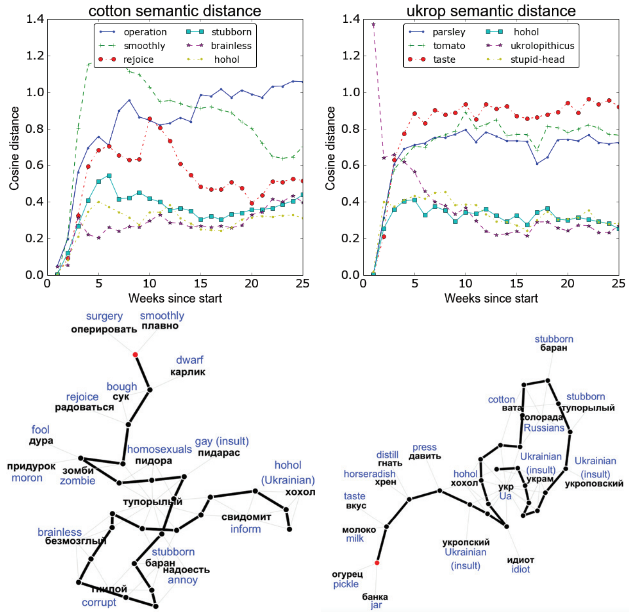

Figure 2: Representation shift between each word and its nearest neighbors (top) and semantic trajectories of representation shift, including nearest neighbors (bottom).

ing representation shift.

Visualizing Representation Shift

To explore the shape of representation shift, we can visualize both a word's semantic distance from the previous timestep and the distance to its semantic neighbors. We show the trajectory of the keywords *cotton* and *ukrop* (*"dill"*) through representation space, in two different ways, in Figure 2. The top images show the cosine distance from each keyword's embedding to six of its nearest neighbors: three from the beginning of the data and three from the end of the data. The bottom images show the movement of the keyword through a 2-D projection of the embedding space, as well as the keyword's relative distance to its two nearest neighbors at each week, where the red point marks $t = 0$.

For both *cotton* and *ukrop*, in the bottom graphs we see a trend toward negative language: for example, the nearest neighbors to *cotton* are initially medical words such as *surgery*, but over time the neighbors become more negative terms such as *brainless*. However, in the top graphs the words have somewhat different trajectories: while *ukrop* shows a more clean separation between literal (food) and negative (non-food) meaning, *cotton* appears to retain some similarity to its non-negative neighbors such as *rejoice*. This complication could be further investigated by an analyst and would not have been detected through frequency alone.

Predicting Representation Shift

We now frame representation shift as a prediction problem: can we use frequency-based measures to predict change in meaning? Contrasting with previous studies (Kulkarni et al. 2015), we look to predict representation shift $\Delta\tau_e(w)$ from either meaning shift or concept drift, or both:

1. $\Delta\tau_e(w) = \phi(\Delta\tau_e(w))$ (Table 1).
2. $\Delta\tau_e(w) = \phi(\Delta\tau_\chi(w))$ (Table 2).
3. $\Delta\tau_e(w) = \phi(\Delta\tau_\chi(w), \Delta\tau_e(w))$ (Table 3).

We predict the final value in the time series by training with all data prior to the final time step, using as predictors all data up to 1, 2 and 3 weeks before the final time step. We perform these experiments using 4-fold cross validation.

We use the following **evaluation metrics** for each word w, y_i is the observed value of representation shift $\Delta\tau_e(w)$ at time t, $\hat{y}_i$ is the predicted value at time t, $\bar{y}$ and $\bar{\hat{y}}$ denote the mean values over all words in the vocabulary. We first report Pearson correlation r:

$$r = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}}.$$

We also measure Root Mean Squared Error (RMSE) γ:

$$\gamma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \times 10^{-2}}.$$

We restrict our prediction to words with consistent ($\geq 50\%$) nonzero frequency to avoid learning from overly sparse time series. Although this cuts down our vocabulary to about 20,000 unique words, it also ensures that the estimators will not achieve artificially high accuracy by predicting zeros for the majority of the words.

Our primary model is a one-layer Long Short-Term Memory (LSTM) neural network[4] for regression (Oancea and Ciucu 2013). We initialize the network with one input node per timestep and a single output node, using the raw output as the predicted value for regression. We contrast the LSTM's performance with an AdaBoost regressor.[5] We compare these models' performances with a baseline model that predicts the observed value of representation shift $\Delta\tau_e(w)$ at time t using the value copied from time $t - 1$.

First, the results of predicting representation shift from previous representation shift are shown in Table 1. Comparing models' relative performance, we see clearly that the LSTM outperforms the AdaBoost model which outperforms the baseline, in both metrics. Furthermore, we see that predicting one week ahead clearly surpasses forecasting for two or more weeks in all metrics and that the performance drops only slightly as we increase the distance of forecasting. This suggests that the signal for representation shift immediately before the period of prediction is nontrivial, and thus that representation shift can occur in a short timeframe.

Next, we show the results of prediction for representation shift from previous concept drift in Table 2. We see an immediate decrease in performance as compared with the previous task as measured by the Pearson correlation r, demonstrating the lack of signal associated. However, we note that

[4] Theano via keras: https://keras.io/

[5] Scikit-learn: http://scikit-learn.org/stable/

	1 week		2 weeks		3 weeks	
	r	γ	r	γ	r	γ
Baseline	0.62	5.00	0.30	5.59	0.29	4.83
AdaBoost	0.69	5.73	0.40	6.56	0.39	5.97
LSTM	0.73	4.16	0.50	3.89	0.49	3.91

Table 1: Prediction results for representation shift from previous representation shift: $\Delta\tau_e(w) = \phi(\Delta\tau_e(w))$.

	1 week		2 weeks		3 weeks	
	r	γ	r	γ	r	γ
Baseline	0.17	47.4	0.28	59.5	0.29	43.3
AdaBoost	0.40	7.26	0.16	7.89	0.15	8.30
LSTM	0.44	5.56	0.21	4.37	0.18	4.39

Table 2: Prediction results for representation shift from previous concept drift: $\Delta\tau_e(w) = \phi(\Delta\tau_\chi(w))$.

	1 week		2 weeks		3 weeks	
	r	γ	r	γ	r	γ
Baseline	0.35	5.19	0.30	5.37	0.21	4.34
AdaBoost	0.47	5.24	0.49	6.43	0.40	4.46
LSTM	0.52	3.21	0.52	4.29	0.48	2.90

Table 3: Prediction results for representation shift from representation shift and concept drift:
$\Delta\tau_e(w) = \phi(\Delta\tau_\chi(w), \Delta\tau_e(w))$.

the RMSE γ increased only slightly for both AdaBoost and LSTM as compared with the previous prediction task, suggesting that concept drift can provide nontrivial signal for representation shift. Similar to before, the drop in performance between one and three weeks supports the short-term relationship between representation shift and concept drift.

Lastly, we predict representation shift as a function of both concept drift and representation shift. The results in Table 3 show that this combined prediction performs somewhere between the other predictions, e.g. Pearson's correlation for combined prediction greater than $\Delta\tau_e(w) = \phi(\Delta\tau_\chi(w))$ prediction but less than $\Delta\tau_e(w) = \phi(\Delta\tau_e(w))$ prediction. The performance for two-three week predictions indicates that concept drift does contribute some signal to amplify the signal from representation shift, but the one week prediction results show lower performance due to concept drift noise. Note that the RMSE is comparable to the first prediction task (and lower for the LSTM), and thus the combined prediction has a competitive margin of error.

Conclusion

This work provides a generalizable proof of concept for future studies on short-term representation shift in social media – despite noisy data, the word vector representations generated are robust. Our prediction results show that by considering representation in addition to raw frequency, we are able not only to forecast meaning change for words over time but also to isolate interesting words, i.e. those with dynamic contexts. We propose representation shift as a novel metric to track unexpected changes during a crisis, showing the power of semantics in action.

References

Alsaedi, N.; Burnap, P.; and Rana, O. 2016. Sensing real-world events using Arabic Twitter posts. In *Proceedings of ICWSM*, 515–518.

Asur, S., and Huberman, B. A. 2010. Predicting the future with social media. In *Proceedings of WI-IAT*, 492–499.

Bruno, N. 2011. Tweet first, verify later? How real-time information is changing the coverage of worldwide crisis events. *Oxford: Reuters Institute for the Study of Journalism* 10(2011).

Crooks, A.; Croitoru, A.; Stefanidis, A.; and Radzikowski, J. 2013. #earthquake: Twitter as a distributed sensor system. *Transactions in GIS* 17(1):124–147.

Gulordava, K., and Baroni, M. 2011. A distributional similarity approach to the detection of semantic change in the Google Books Ngram corpus. In *Proceedings of GEMS*, 67–71.

Hamilton, W. L.; Leskovec, J.; and Jurafsky, D. 2016a. Cultural shift or linguistic drift? comparing two computational measures of semantic change. *Proceedings of EMNLP* 2116–2121.

Hamilton, W. L.; Leskovec, J.; and Jurafsky, D. 2016b. Diachronic word embeddings reveal statistical laws of semantic change. *Proceedings of ACL* 1489–1501.

Harris, Z. S. 1954. Distributional structure. *Word* 10(2-3):146–162.

Kim, Y.; Chiu, Y.-I.; Hanaki, K.; Hegde, D.; and Petrov, S. 2014. Temporal analysis of language through neural language models. *Proceedings of ACL Workshop on LTCSS* 61–65.

Kulkarni, V.; Al-Rfou, R.; Perozzi, B.; and Skiena, S. 2015. Statistically significant detection of linguistic change. In *Proceedings of WWW*, 625–635.

Mathioudakis, M., and Koudas, N. 2010. Twittermonitor: Trend detection over the Twitter stream. In *Proceedings of SIGMOD*, 1155–1158.

Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. In *Proceedings of NIPS*, 3111–3119.

Oancea, B., and Ciucu, Ş. C. 2013. Time series forecasting using neural networks. *Challenges of the Knowledge Society* 1402–1408.

Řehůřek, R., and Sojka, P. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC Workshop on New Challenges for NLP Frameworks*, 45–50.

Sagi, E.; Kaufmann, S.; and Clark, B. 2009. Semantic density analysis: Comparing word meaning across time and phonetic space. In *Proceedings of GEMS*, 104–111.

Volkova, S.; Chetviorkin, I.; Arendt, D.; and Van Durme, B. 2016. Contrasting public opinion dynamics and emotional response during crisis. In *Proceedings of SocInfo16*, 312–329.

Weng, L.; Menczer, F.; and Ahn, Y.-Y. 2013. Virality prediction and community structure in social networks. *Scientific reports* 3:1–6.