|  |  |  |
| --- | --- | --- |
|  | **Analyzing Concept Drift: A Case Study in Country Tweeting** |  |
| Jordan Limperis†  Engineering for Professionals  Johns Hopkins  Baltimore, Maryland  [Jlimper3@jhu.edu](mailto:Jlimper3@jhu.edu) |  |  |

# 

# ABSTRACT

# In this paper, we investigate utilizing linguistic attributes for detecting the sentiment of Twitter messages directed at tagged countries. To accomplish this, we evaluate existing lexical tools as well as the creative language used in microblogging. We investigate the presence of concept drift in the underlying distributions. Specifically, countries tagged over the Corona Virus Pandemic, Summer 2020, are used to build training data to detect. Temporal elements of the data are used to evaluate how the data is trained and whether concept drift affects model accuracy.

# CCS CONCEPTS

• Sentiment Analysis • Twitter • Concept Drift

# KEYWORDS

Sentiment Analysis, Twitter, Concept Drift, Country

**ACM Reference format:**

Jordan Limperis. 2021. Analyzing Concept Drift: A Case Study in Country Tweeting. In Proceedings of Johns Hopkins University AI Systems Class. Baltimore, Maryland*,* USA.

# Introduction

Social Media data has become a common measure to access public opinion on various topics. While the fast-paced nature of social media posts provides a method to capture trends quickly, it also poses challenges involving concept drift. When the historical data no longer predicts modern data, we know that concept drift has occurred; there exist covariates that altered the relationship of the variables the model is evaluating (Jeffrey C Schlimmer and Richard H Granger). This becomes particularly problematic when large concept drifts occur quickly resulting in an immediate need to address model performance. In general, Twitter has been described as a medium with high degrees of concept drift (Joana Costa et al). Here we examine the impact of concept drift by examining country sentiments expressed via Twitter during the COVID-19 pandemic. Country sentiment is important in the pandemic age as it provides an indicator toward citizen outlook on a country public policy. Through concept drift analysis, we hope to identify trends in country sentiment, possibly specific to individual countries, and to identify possible pitfalls of not evaluating concept drift when creating country sentiment models.

## Literature Review

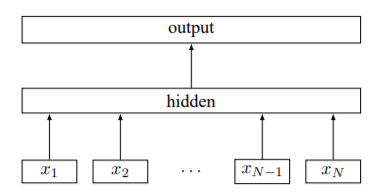
It is first important to distinguish concept drift from other reasons for decreased performance, like random noise or data preprocessing (Indrė Žliobaitė, Geoffrey I Webb et al). The classic example of concept drift is a change in the meaning of the classes, which is called real concept drift. Also, an observed performance change can occur due to a change in underlying distribution, which is virtual drift (Alexey Tsymbal, Gerhard Widmer). Many times, the complexity of the analysis prevents identification of specific types of drift, so they are treated equally (Indrė Žliobaitė). Due to the substantial amount of work dedicated to concept drift, three methods have been devised to deal with it: (1) a time window, (2) and incremental model, and (3) an ensemble model (Joana Costa et al). Martin Müller, and Marcel Salathé’s work on COVID-19 era vaccine sentiment as well as Joana Costa et al’s work on twitter streams show that the incremental method provides the most accurate results. Because the direct impact of concept drift is task, environment, and model dependent (Indrė Žliobaitė, Mykola Pechenizkiy, and Joao Gama), it is important to address concept drift in the specific case of country stance classification. To best of our knowledge, this will be the first paper to address concept drift in the over encompassing subject area of tagged countries in Twitter messages.

Sentiment Analysis over Twitter has been had a growing interest in the research community, so many related works have been published. Receiving widespread media attention, J. Bollen and H. Mao’s work in using sentiment analysis to predict the stock market is one such example of illustrating how linguistic sentiment can be used to accurately predict social-economic shifts. With the continued work of Mittal, A., & Goel, A., they found that public mood can indeed be captured from large-scale Twitter feeds and mapped against historical socio-economic events. There are problems though with this approach. For example, Silva S. et all found that updating models on the fly prevents inaccuracies in prediction, as it accounts for shift in the underlying behaviors of the Twitter users. This underlying shift is commonly referred to as concept drift and has been studied in various forms of artificial intelligence analysis (Tsymbal, A). Silva S. et all also illustrated that performing concept drift requires a large amount of manual work to account for labeling data. These challenges of identifying and prevent concept drift as well as requiring large amounts of manual work are two challenges that we look to address in this paper. Many tools have been released to evaluate sentiment analysis over Twitter. For example, we highlight production tools like Twendz, Tweetfeel, and Twitrratr. The proposed work in this paper differs from work with these tools, because it utilizes novel approaches illustrated by Silva et a, who pioneered adapting classification models in Twitter to concept drift without human intervention.

# Methods

## Algorithm

To evaluate concept drift, a temporal model is required to see model change over time. We have chosen to utilize fastText, which is based on classical machine learning models which much academic research has been performed on. Due its measured accuracy as well as being magnitudes faster than alternatives, fastText, provide value in analyzing large amounts of Twitter data with minimal human supervision (Armand Joulin et al). fastText uses the classical Bag of Word model, which discards word order to save immense computation cost. It instead utilizes a bad of n-grams as additional features to capture partial information about word order, which allows hashing to save computational cost. Figure 1 below illustrates the theoretical model conception.



**Figure 1**: Theoretical model architecture of fastText with N ngram features represented by x1, x2, …, xN. These features are averaged to form the hidden variable.

## Dataset

The data we are analyzing consists of any Twitter posts contained a tagged country (denoted by #) between the months of Jun 2020 to October 2020, which was during the height of the Corona Virus Pandemic. This data is filtered to specific countries tagged. For our initial analysis, we looked at the USA, China, Japan, Greece, and Italy. This dataset was archived by Jiang William who also created an initial library for VADER sentiment analysis. To prepare the data, HTML tags, Emoticons (Smilies), and punctuation were removed from the tweets. Contractions of words were also simplified to their base forms. We choose not to remove stop words as stop words can provide specific sentiment inversions (ex. “not happy” converted to “happy”).

## Sentiment Analysis

We take a novel approach to defining sentiment. Normally, to accomplish a fastText model training, specific tags are used to classify data as Neutral, Positive, or Negative (Armand Joulin et al). This method is highly intensive with tagging and is subject to human biases. Instead, we chose to utilize VADER to perform initial Tweet labelling. VADER is a lexical rule-based model that has been empirically validated against other sentiment analysis tools for social media. In fact, VADER outperforms individual human raters (Hutto, C., & Gilbert, E). This approach’s measured accuracy at labelling allows large amounts of data labeling to occur as well as a provides a discrete measure to evaluate FastText accuracy, and consequently concept drift. Any twitter post with a VADER sentiment score less than -.25 is labelled as negative, any twitter post with a VADER score above .25 is labelled positive, and any post with score between -.25 and .25 is labelled neutral. To account for possible class imbalance, which is another reason for decreased model performance, we chose to compare performance against input data that had been upsampled. By making the amount of neutral, negative, and positive tweets in a training set, we can ensure that class imbalance is not causing issues.

# Results and Discussion

In this section, we evaluate the performance of the fastText model across the months of July to October 2020 during the Corona Virus Pandemic. The goal is to alter the trained month to observe the effect of using the fastText model on subsequent months.

## Overall Precision

The below tables represent the validation over months 08, 09, 10 in 2020. Each row represents utilizing a different month (07, 08, 09) to train the model. Each column represents the month the trained model was validated over. In the below data, each number corresponds to the precision of the model as determined across all three classes: neutral, negative, and positive. This precision corresponds to the model’s overall accuracy in predicting the sentiment of the validated data set.

|  |  |  |  |
| --- | --- | --- | --- |
| USA | Precision: |  |  |
| Month | 08 | 09 | 10 |
| 07 | 0.65500 | 0.66055 | 0.65131 |
| 08 |  | 0.76877 | 0.75749 |
| 09 |  |  | 0.75624 |
|  |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Japan | Precision: |  |  |
| Month | 08 | 09 | 10 |
| 07 | 0.71314 | 0.71324 | 0.73570 |
| 08 |  | 0.79499 | 0.81345 |
| 09 |  |  | 0.80499 |
|  |  |  |  |
| China | Precision: |  |  |
| Month | 08 | 09 | 10 |
| 07 | 0.58289 | 0.55245 | 0.55546 |
| 08 |  | 0.68688 | 0.68982 |
| 09 |  |  | 0.65048 |
|  |  |  |  |
| Italy | Precision: |  |  |
| Month | 08 | 09 | 10 |
| 07 | 0.72996 | 0.72319 | 0.70594 |
| 08 |  | 0.79465 | 0.77192 |
| 09 |  |  | 0.72416 |
|  |  |  |  |
| Greece | Precision: |  |  |
| Month | 08 | 09 | 10 |
| 07 | 0.62718 | 0.59218 | 0.61417 |
| 08 |  | 0.71179 | 0.72967 |
| 09 |  |  | 0.68990 |
|  |  |  |  |

**Figure 2**: Precision of fastText model trained on twitter data with country tags from July to September 2020, represented by the y-axis. The x-axis corresponds to the month (August to October 2020) the fastText model was validated against.

To account for possible class imbalance in the trained data, we performed another experiment to ensure there is not significant change in precision from ensuring equal class balance, or desampling. Overall, we saw that most models saw less than a 1% change in performance, with the Greece model seeing an overall 4% change in accuracy after desampling training data. It is worth nothing that desampling consistently increased the accuracy of the trained models in all scenarios. Overall, there was not significant change in the data from desampling, except for Greece, which would mean any changes in precision over training validations would in part correspond to concept drift. The possible impact of Greeces class imbalance affecting concept drift analysis caused for its exclusion.

|  |  |
| --- | --- |
| Country | Model Performance Change Average after Desampling |
| USA | .06% |
| Japan | .74% |
| China | 2.39% |
| Italy | .65% |
| Greece | 4.80% |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |

**Figure 3**: Depictions in percentile increases in precision after desampling or ensuring equal classes in the trained data. The percent change was averaged over all months validated for an overall effect.

Except Greece, all the other countries show unexplained variability in the precision depending on input month. This indicates the presence of other cofounding variable that have possibility changed. We evaluate the impact of this in the Conclusion, section 4.

## Class Precision

To better pinpoint where there is changes in cofounding variables, we also chose to find individual class precision as measured across each month from using different months as a training data set. By focusing on neutral, positive, or negative sentiment outlooks in tweets, we could see an accentuated change in performance in specific areas. Italy was one such example, which saw a high degree of precision change when training on August 2020 and September 2020 twitter data. These changes are much more drastic than the foreseen overall precision scores for Italy.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Negative Sentiment – Italy | Precision: |  |  |  |
| Month | 08 | 09 | 10 |  |
| 07 | 0.55875 | 0.61321 | 0.63512 |
| 08 |  | 0.71006 | 0.70710 |
| 09 |  |  | 0.87456 |

|  |  |  |  |
| --- | --- | --- | --- |
| Positive Sentiment – Italy | Precision: |  |  |
| Month | 08 | 09 | 10 |
| 07 | 0.72897 | 0.72649 | 0.70825 |
| 08 |  | 0.82250 | 0.80209 |
| 09 |  |  | 0.69663 |

|  |  |  |  |
| --- | --- | --- | --- |
| Neutral Sentiment – Italy | Precision: |  |  |
| Month | 08 | 09 | 10 |
| 07 | 0.74432 | 0.72817 | 0.70970 |
| 08 |  | 0.78245 | 0.75743 |
| 09 |  |  | 0.74379 |

**Figure 4:** Class-specific precision for Italy. As can be seen, there is a high increase in negative sentiment identification and a decrease in positive sentiment identification

From the above data of Italy, we can see definitive alteration in the core data set with a loss in precision for positive sentiment analysis and an increase in negative sentiment analysis as we train on subsequent months. This points to the presence of concept drift that we evaluate in conclusion section. The other country’s sentiment scores are provided in the Index in section 5

# Conclusion

Overall, the variability of the precision score based on trained month indicates the presence of concept drift. This alteration can be clearly seen in many of the countries. Despite the difference in ability of some the fastText model to evaluate certain countries (range of precision is 50-80%), all countries see an increase in precision when trained on August 2020, as compared to all other months, even September 2020. This indicates the presence of some event in the time range of August 2020 that changed the underlying variables in the model. This is known as concept drift. During the months of July to August 2020, the Corona Virus pandemic was undergoing it’s second spike starting in July 2020 resulting in public health policy changes. It is probable that this change in public health relations caused a change in narrative of social media data that altered underlying structure of how tagged countries were discussed in social media, like twitter. This correlative relation between how the Corona Virus pandemic affected twitter narratives leading to change in country sentiments, or country outlooks, is worth further analysis.

Even though there was variability in the twitter country-tagged data, the type of concept drift varies across individual countries. This variance in sentiment and possibly country outlook is seen in the specific class precision changes across individual countries. For example, the growth in precision of positive sentiment prediction in the USA based on trained month contrasts with the growth of negative sentiment prediction in Italy. With the overall trend of increased sentiment analysis prediction in the August month, this difference in positive/negative sentiment accuracy points to differences in how the concept drift is affecting each individual country distribution. For example, Italy and the USA twitter users might be reacting differently to the overall effects of the Corona Virus Pandemic and how the public outlook is in each country. The increase in accuracy to predict negative sentiment could be in relation to the highly negative and publicized effect of the Corona Virus on Italy. These types of correlative relations between the Corona Virus Pandemic and country sentiment is worth further research analysis.

A more intensive research study on the differences in concept drift patterns is needed, especially in the context of Twitter analysis. With the presence of concept drift in time-series Twitter Data commonly used in sentiment analysis, it is important to account for and test for it. It is also worth noting that the relationship between public events and social media sentiment is rich contextual information that can be evaluated further, and possibly used to predict how the public might react to certain events.

REFERENCES

1. Alexey Tsymbal. 2004. “The problem of concept drift: definitions and related work”. In: Computer Science Department, Trinity College Dublin 106.2 , p. 58.
2. Armand Joulin et al. 2016. “Bag of tricks for efficient text classification”. In: arXiv preprint arXiv:1607.01759.
3. Charles Malafosse 2019. “FastText sentiment analysis for tweets: A straightforward guide”. In: towarddatascience.com
4. Geoffrey I Webb et al. 2016. “Characterizing concept drift”. In: Data Mining and Knowledge Discovery 30.4, pp. 964–994
5. Gerhard Widmer and Miroslav Kubat. 1996. “Learning in the presence of concept drift and hidden contexts”. In: Machine learning 23.1, pp. 69–101.
6. Hutto, C., & Gilbert, E. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the International AAAI Conference on Web and Social Media ,Vol. 8, No. 1.
7. Indrė Žliobaitė. “Learning under concept drift: an overview”. In: arXiv preprint arXiv:1010.4784 (2010).
8. Indrė Žliobaitė, Mykola Pechenizkiy, and Joao Gama. “An overview of concept drift applications”. In: Big data analysis: new algorithms for a new society. Springer, 2016, pp. 91–114
9. Jeffrey C Schlimmer and Richard H Granger. “Incremental learning from noisy data”. In: Machine learning 1.3 (1986), pp. 317–354.
10. Jiang William. World - Twitter Sentiment By Country, Version 3. Retrieved 2/21/2021 from [www.kaggle.com/wjia26](http://www.kaggle.com/wjia26) (2020).
11. Joana Costa et al. “Concept drift awareness in twitter streams”. In: 2014 13th International Conference on Machine Learning and Applications. IEEE. 2014, pp. 294–299
12. J. Bollen and H. Mao. Twitter mood as a stock market predictor. IEEE Computer, 44(10):91–94
13. Martin Müller, Marcel Salathé Addressing machine learning concept drift reveals declining vaccine sentiment during the COVID-19 pandemic. arXiv:2012.02197 [cs.SI] (2020).Machinery, New York, NY, USA, 87–98. <https://doi.org/10.1145/244130.244151>
14. Mittal, A., & Goel, A. (2012). Stock prediction using twitter sentiment analysis. Standford University, CS229 (2011 http://cs229. stanford. edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis. pdf), 15.
15. Silva, I. S., Gomide, J., Barbosa, G. A., Santos, W., Veloso, A., Meira Jr, W., & Ferreira, R. (2011). Observatório da Dengue: surveillance based on twitter sentiment stream analysis. In Proceedings of the Brazilian Symposium on Databases, Demos Track. Florianópolis, Brazil (pp. 49-54).
16. Tsymbal, A. (2004). The problem of concept drift: definitions and related work. Computer Science Department, Trinity College Dublin, 106(2), 58.

# Index

The below values correspond to the class-specific (neutral, positive, negative) precisions depending on trained month for the fastText model (see the discussion in section 3.2).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Negative Sentiment – Italy | Precision: |  |  |  |
| Month | 08 | 09 | 10 |  |
| 07 | 0.55875 | 0.61321 | 0.63512 |
| 08 |  | 0.71006 | 0.70710 |
| 09 |  |  | 0.87456 |

|  |  |  |  |
| --- | --- | --- | --- |
| Positive Sentiment – Italy | Precision: |  |  |
| Month | 08 | 09 | 10 |
| 07 | 0.72897 | 0.72649 | 0.70825 |
| 08 |  | 0.82250 | 0.80209 |
| 09 |  |  | 0.69663 |

|  |  |  |  |
| --- | --- | --- | --- |
| Neutral Sentiment – Italy | Precision: |  |  |
| Month | 08 | 09 | 10 |
| 07 | 0.74432 | 0.72817 | 0.70970 |
| 08 |  | 0.78245 | 0.75743 |
| 09 |  |  | 0.74379 |

|  |  |  |  |
| --- | --- | --- | --- |
| Negative Sentiment - USA |  |  |  |
| Month | 08 | 09 | 10 |
| 07 | 0.61165 | 0.60942 | 0.57194 |
| 08 |  | 0.74087 | 0.70243 |
| 09 |  |  | 0.69596 |

|  |  |  |  |
| --- | --- | --- | --- |
| Positive Sentiment - USA |  |  |  |
| Month | 08 | 09 | 10 |
| 07 | 0.70409 | 0.69983 | 0.69725 |
| 08 |  | 0.81074 | 0.80433 |
| 09 |  |  | 0.80800 |

|  |  |  |  |
| --- | --- | --- | --- |
| Neutral Sentiment- USA |  |  |  |
| Month | 08 | 09 | 10 |
| 07 | 0.64056 | 0.65352 | 0.65030 |
| 08 |  | 0.74661 | 0.74440 |
| 09 |  |  | 0.74393 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Negative Sentiment - Japan | | | | |  | | | | | | |  | | | | | | | |
| Month | | | | | 08 | | | | 09 | | | 10 | | | | | | | |
| 07 | | | | | 0.79942 | | | | 0.74795 | | | 0.77179 | | | | | | | |
| 08 | | | | |  | | | | 0.73675 | | | 0.76076 | | | | | | | |
| 09 | | | | |  | | | |  | | | 0.77176 | | | | | | | |
| Positive Sentiment – Japan | | | | |  | | | | | | |  | | | | |
| Month | | | | 08 | | | 09 | | | | | 10 | | | | |
| 07 | | | | 0.69525 | | | 0.72197 | | | | | 0.74440 | | | | |
| 08 | | | |  | | | 0.85088 | | | | | 0.85287 | | | | |
| 09 | | | |  | | |  | | | | | 0.84576 | | | | |
| Neutral Sentiment- Japan | |  | | | | | | | | | |  | | | | |
| Month | | | | 08 | | | | | 09 | | | 10 | | | | |
| 07 | | | | 0.71923 | | | | | 0.70445 | | | 0.72740 | | | | |
| 08 | | | |  | | | | | 0.76319 | | | 0.79115 | | | | |
| 09 | | | |  | | | | |  | | | 0.77870 | | | | |
| Negative Sentiment – China | | | | | | | | |  | | |  | | | | | | |
| Month | | | | 08 | | | | | 09 | | | 10 | | | | | | |
| 07 | | | | 0.59885 | | | | | 0.58956 | | | 0.60834 | | | | | | |
| 08 | | | |  | | | | | 0.71282 | | | 0.73133 | | | | | | |
| 09 | | | |  | | | | |  | | | 0.67763 | | | | | | |
| Positive Sentiment- China | | | | | |  | | | | | |  | |
| Month | | 08 | | | | 09 | | | | | | 10 | |
| 07 | | 0.57906 | | | | 0.53807 | | | | | | 0.54695 | |
| 08 | |  | | | | 0.71096 | | | | | | 0.70721 | |
| 09 | |  | | | |  | | | | | | 0.65966 | |
| Neutral Sentiment - China | | | | | |  | | | | | |  | | | |
| Month | | | 08 | | | 09 | | | | | | 10 | | | |
| 07 | | | 0.57913 | | | 0.54633 | | | | | | 0.53546 | | | |
| 08 | | |  | | | 0.65381 | | | | | | 0.65039 | | | |
| 09 | | |  | | |  | | | | | | 0.62719 | | | |
| Negative Sentiment- Greece | | | | | | | | | |  | | |  | | | | |
| Month | | | 08 | | | | | | | 09 | | | 10 | | | | |
| 07 | | | 0.57909 | | | | | | | 0.62758 | | | 0.62468 | | | | |
| 08 | | |  | | | | | | | 0.71827 | | | 0.72777 | | | | |
| 09 | | |  | | | | | | |  | | | 0.67940 | | | | |
| Positive Sentiment - Greece | | | | | | | |  | | |  | | | |
| Month | 08 | | | | | | | 09 | | | 10 | | | |
| 07 | 0.64034 | | | | | | | 0.55129 | | | 0.59822 | | | |
| 08 |  | | | | | | | 0.73347 | | | 0.75943 | | | |
| 09 |  | | | | | | |  | | | 0.71526 | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| Neutral Sentiment- Greece | |  |  |
| Month | 08 | 09 | 10 |
| 07 | 0.62977 | 0.60694 | 0.62240 |
| 08 |  | 0.69516 | 0.71087 |
| 09 |  |  | 0.67918 |

**Figure 5:** Class specific changes in precision. The y-axis corresponds to trained month (07-09), while the x-axis refers to the month validated against to achieve the precision score.