

## Active Feature Acquisition for Opinion Stream Classification under Drift

Active stream learning is frequently used to acquire labels for instances and less frequently to determine which features should be considered as the stream evolves. We introduce a framework for active feature selection, intended to adapt the feature space of a polarity learner over a stream of opinionated documents. We report on the first results of our framework on substreams of reviews on different product categories.

M.Sc. Ranjith Shivakumaraswamy, Otto-von-Guericke-University

M.Sc. Christian Beyer, Otto-von-Guericke-University

M.Sc. Vishnu Unnikrishnan, Otto-von-Guericke-University

Prof. Dr. Eirini Ntoutsi, Leibniz University Hannover

Prof. Dr. Myra Spiliopoulou, Otto-von-Guericke-University

## Workflow Over the Document Stream

Our framework slides a window  $W$  of  $n$  epochs (here: weeks) over the stream, learning on  $n$  epochs and testing on the epoch  $n+1$ .

**Module ALF for Feature Ranking:** Our active feature selector ALF ranks features on importance. Feature ranking methods include mutual information, information gain, document frequency thresholding, chi-square and document frequency thresholding (DFT) as discussed by Basu et al [3], Distinguishing Feature Selector (DFS), Odds Ratio and Normalized Difference Measure (NDM) as studied in [1], Gini-index, signed chi-square and signed information gain [10], the stratified feature ranking method of [4] and the approach proposed by [6]. We opted for the Distinguishing Feature Selector (ALF-DFS) and the Gini (ALF-Gini) because they were found to have the most competitive performance [14].

**Module RALF for Feature Subspace Recommendation:** The recommender takes as input the size  $M$  of the subspace to be replaced and invokes ALF for feature ranking. Currently we use  $M = \text{FeatureSpaceSize}/2$ .

We have four variants of RALF:

- **Baseline:** invokes ALF-Gini on the data inside the current window
- **Oracle-Random:** picks randomly  $M$  features from the feature space of the next epoch (the epoch  $n+1$ , i.e. the first epoch in the future)
- **Oracle-Gini:** invokes ALF-Gini on epoch  $n+1$  and returns the top- $M$  features
- **Oracle-DFS:** similar to ALF-Gini, but invokes ALF-DFS on epoch  $n+1$

Hence, the Oracle variants simulate an expert who knows which features will become important in the immediate future. We use the top- $M$  of these features to replace the least important ones of the current feature space, thus preserving the presently informative features still.

**Stream Classification Core:** The opinion stream learner replaces the least in-formative features (according to ALF's ranking) with the features suggested by RALF. It re-learns on the current window and uses the next epoch for testing. Then, the window shifts by one epoch, forgetting the least recent one.

**Drift-driven Feature Space Update:** Drift monitor that invokes RALF if and only if drift occurs. For drift detection we use the method of Gama et al [7]

## Acknowledgement

This work is partially funded by the German Research Foundation, project OSCAR "Opinion Stream Classification with Ensembles and Active Learners".

We further thank Elson Serrao who made the basic components of opinion stream mining available.

## References

- [1] Barddal, J.P., Gomes, H.M., Enembreck, F., Pfahringer, B.: A survey on feature drift adaptation: Denition, benchmark, challenges and future directions. *Journal of Systems and Software* 127, 278{294 (2017)
- [2] Basu, T., Murthy, C.: Eective text classification by a supervised feature selection approach. In: 12th IEEE ICDM, Workshops Volume. pp. 918{925. IEEE (2012)
- [3] Chen, R., Sun, N., Chen, X., Yang, M., Wu, Q.: Supervised feature selection with a stratied feature weighting method. *IEEE Access* 6, 15087{15098 (2018)
- [4] Contardo, G., Denoyer, L., Artieres, T.: Sequential cost-sensitive feature acquisition. In: *Int. Symp. on Intelligent Data Analysis*. pp. 284-294. Springer (2016)
- [5] Fattah, M.A.: A novel statistical feature selection approach for text categorization. *Journal of Information Processing Systems* 13(5) (2017)
- [6] Gama, J., Medas, P., Castillo, G., Rodrigues, P.: Learning with drift detection. In: *Brazilian Symposium on Artificial Intelligence*. pp. 286-295. Springer (2004)
- [7] Huang, S.J., Xu, M., Xie, M.K., Sugiyama, M., Niu, G., Chen, S.: Active feature acquisition with supervised matrix completion. In: 24th ACM SIGKDD Int. Conf. on Knowledge Discovery & Data Mining. pp. 1571-1579. ACM (2018)
- [8] McAuley, J., Targett, C., Shi, Q., Van Den Hengel, A.: Image-based recommendations on styles and substitutes. In: 38th Int ACM SIGIR Conf on Research and Development in Information Retrieval. pp. 43-52. ACM (2015)
- [9] Ogura, H., Amano, H., Kondo, M.: Comparison of metrics for feature selection in imbalanced text classification. *Expert Systems with Applications* 38(5), 4978-4989(2011)
- [10] Serrao, E., Spiliopoulou, M.: Active stream learning with an oracle of unknown availability for sentiment prediction. In: *IAL@ECML PKDD*. pp. 36{47 (2018)
- [11] Shim, H., Hwang, S.J., Yang, E.: Joint active feature acquisition and classification with variable-size set encoding. In: *Advances in Neural Information Processing Systems*. pp. 1368{1378 (2018)
- [12] Shivakumaraswamy, R.: Active learning over text streams. Tech. rep., Otto-von-Guericke-University Magdeburg Department of Computer Science (2019)
- [13] Uysal, A.K.: An improved global feature selection scheme for text classification. *Expert systems with Applications* 43, 82-92 (2016)
- [14] Asim, M.N., Wasim, M., Ali, M.S., Rehman, A.: Comparison of feature selection methods in text classification on highly skewed datasets. In: 1st Int. Conf. on Latest Trends in Electrical Engineering and Computing Technologies (INTELLECT). pp. 1-8. IEEE (2017)

## Introduction

Opinion stream classification algorithms assign a polarity label to each arriving opinionated document. The feature space over the stream may change though, e.g. when new product appear and the words/phrasing used by customers who reviewed them changes. Feature space adaption can benefit from an active learning approach, where a human expert specifies the features of importance.

Contardo et al. [5] use reinforcement learning to acquire features, and also consider feature acquisition cost. Huang et al. [8] take uncertainty into account. The "sequential feature acquisition framework" of Shim et al. [12] acquires one feature at a time until the desired model confidence is achieved. These approaches are for static data, though, which are processed in their entirety to build the model. In the stream context, Barddal et al. [2] survey methods that detect feature drift and select features for learning, under the assumption that all features are known in advance. We do not make this assumption. Rather, whenever drift is detected, we use words from recent documents and rebuild the feature space.

We propose a framework for active feature selection on a stream. It consists of: an active learner of features (ALF) that ranks features on importance; a recommender (RALF) that invokes ALF and then recommends a feature subspace to be replaced with the new features; a drift monitor that invokes RALF when model quality decreases.

## Experiment Setup

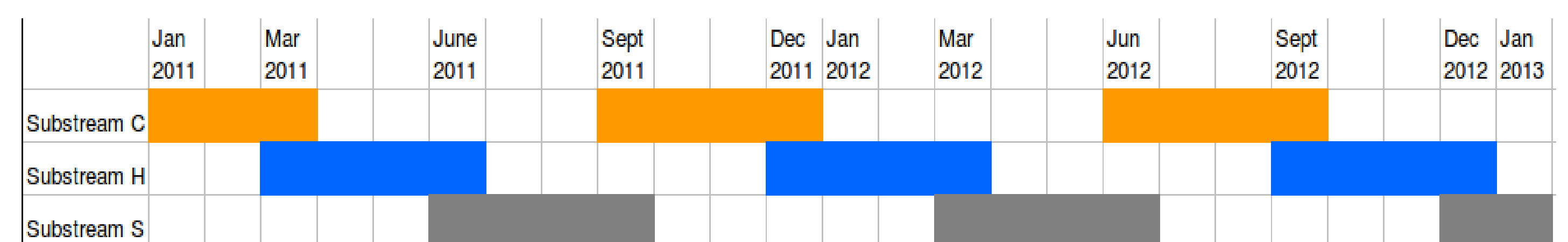
We compared the RALF variants to a default model that does not change the feature space. We performed prequential evaluation, aggregated the SGD log loss values every two months. We used Friedman test with Iman-Davenport modification, rejecting the  $H_0$  for  $p$ -values  $\leq 0.01$ , and then applied Nemenyi post-hoc test. All experiments and results are in [13].

**Data Setup:** We use the "clothing, shoes and jewelry" reviews (substream C), "health and personal care" (substream H) and "sports and outdoors" (S) from the Amazon data set of [9] (<http://jmcauley.ucsd.edu/data/amazon/>), from 01/2011 to 01/2013. There were very few reviews before 2011 and a steep in-increase of positive ones from 2013 on: this product-independent drift calls for conventional classifier adaption, which is beyond our scope. We map ratings 1 and 2 to "Negative", 4 and 5 to "Positive", and 3 to "Neutral".

**Feature Drift Imputation:** We start and stop the substream of each product category at specific time points (as shown in the below image). Hence, product-specific words appear only at given time intervals. We slide a window of 5 weeks in one-week steps over this stream. We build an initial model from the first three weeks, i.e. only from substream C. The first drift occurs when substream H starts.

**Setup of the Components:** As classification core we use Stochastic Gradient Descent (SGD) of scikit-learn ( $\alpha = 0.001$ , l2 penalty and hinge loss). For text preparation, we use the components of [11]. We build the feature space using bag-of-words ("words": 3-grams) and TFIDF, and invoke the dictionary vectorizer of scikit-learn. We vary the feature space size  $M_{full} = 500, 1000, 5000, 10000, 15000$ , so RALF replaces the  $M = M_{full}/2$  least important features.

**Results:** The default model always had inferior performance. Hence updating the feature space is beneficial as response to drift caused through the introduction of new products. Oracle-DFS performed best. Oracle-Gini was within the critical distance to it. Oracle-Random improved as the feature space size increased. The Baseline, which uses ALF-Gini without benefiting from an Oracle, is comparable to Oracle-Gini and Oracle-Random. It is better than the default model except for  $M_{full} = 500$  (where it is within the critical distance from the default model). Hence, ALF-Gini can improve model performance by replacing the least informative features in the current window, when feature drift occurs.



One substream per product category, shifted over time to simulate feature drift

## Performance Comparisons



Performance comparison of models with 500 (right) and 1000(left) model features

## Conclusions

We presented an active feature selection framework for a stream of opinionated documents. Upon drift detection, our framework re-ranks the features with help of the Oracle and replaces the least informative old features with the most in-formative new ones. We evaluated our framework by simulating topic drift. We found that replacing a feature subspace in the presence of drift is beneficial, even if there is no Oracle. We next plan to vary the size and position of the feature subspace to be replaced. Replacing the currently most informative features instead of the least informative ones might be better under concept shift.