Public Sentiment Drift Analysis Based on Hierarchical Variational Auto-encoder

Wenyue Zhang[1], Xiaoli Li[2], Yang Li[1], Suge Wang[1], Deyu Li[1], Liao Jian[1], Jianxing Zheng[1]

[1]School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China
[2]Institute for Infocomm Research, Singapore

wsg@sxu.edu.cn

Abstract

Detecting public sentiment drift is a challenging task due to sentiment change over time. Existing methods first build a classification model using historical data and subsequently detect drift if the model performs much worse on new data. In this paper, we focus on distribution learning by proposing a novel Hierarchical Variational Auto-Encoder (HVAE) model to learn better distribution representation, and design a new drift measure to directly evaluate distribution changes between historical data and new data. Our experimental results demonstrate that our proposed model achieves better results than three existing state-of-the-art methods.

1 Introduction

Public sentiments, whose information hidden in a temporal sequence of documents, become increasingly valuable for real-world applications. Especially, *identifying when public drifts occur* is of great importance to different stakeholders, such as government agencies, companies and news agencies, where they can take proactive actions to avoid damages and pay close attentions to new topics/sentiments etc (Hu et al., 2017). However, the dynamic nature of drifts/changes makes it a challenging problem, although concept drift analysis can be applied to focus on detecting *variation of data distributions* over time. Given historical data and new incoming data, how to *accurately detect drift based on distributional change* is a critical issue.

While existing methods are proposed for sentiment analysis Wu et al. (2019); Fu et al. (2019); Kong et al. (2019); Hoang et al. (2019); Li et al. (2019), etc., many of them are not designed for sentiment *drift* detection task. Xia et al. (2016) focuses on polarity shift detection, but it is at document-level instead of multi-document (or public) level. Recently, some research have been conducted for stream sentiment classification Iosifidis et al. (2017). In particular, statistic processing control (SPC) Ross et al. (2012); Raza et al. (2015) built a detection mechanism which accumulates statistic information of drift indications, and Zhou et al. (2018) applied the mechanism in sentiment drift detection. In addition, Bifet and Gavalda (2007) proposed ADWIN method, which used the upper bound of Hoeffding's inequality to mark drifts. Additionally, Nguyen et al. (2018) combined ADWIN with variational inference and built an online classification system. Wang et al. (2013) proposed an opinion drift detection method which is threshold-based, restricting its applications. A novel framework, proposed by Liu et al. (2016), contains an opinion shift detector based on KL-divergence, while its detection performance is affected by its labeling results. Tsytsarau and Palpanas (2016) defined a novel concept of opinion contradictions and used it in a sentiment change detection experiment. However, the pair-wise method does not involve much history information.

We observe most of the above methods *indirectly* detect drifts instead of *directly evaluate distribution difference*, leading to less effective results. The Variational Auto-Encoder (VAE) model, proposed by Kingma and Welling (2014), is capable to learn latent distributions of inputs, and has better generalization performance (Zhao et al., 2018). As such, we propose a novel *Hierarchical* Variational Auto-Encoder (**HVAE**) model to tackle sentiment drift problem. In particular, we take sentiment distribution changes as the drifts. Practically, sentiments are represented into a 2D vector, whose dimensions and values correspond to polarities (positive or negative) and corresponding intensities respectively. Our main contributions can be summarized as

3762

follows:

1. We propose a HVAE model, which designs 3-level meta-distributions to extend VAE over *hierarchical* structure, enabling effective learning latent distribution representations of input sentiments.

2. We propose a new drift detection measure to compare historical and new data distributions learned by the proposed HVAE.

3. Extensive experimental results on real-world data demonstrate our proposed model is significantly better than state-of-the-arts for sentiment drift detection.

2 Methodology

We now introduce our proposed methodology, including HVAE model and drift measure.

2.1 HVAE Model

Fig. 1 shows a novel three-level hierarchical structure, where W is a length N moving window, each time period i in W contains a set of historical documents $s_{1:L_i}^{(i)}$, and $s_{1:L_{new}}$ contains a set of newly incoming documents. More detailed annotations are described in Tab. 1.
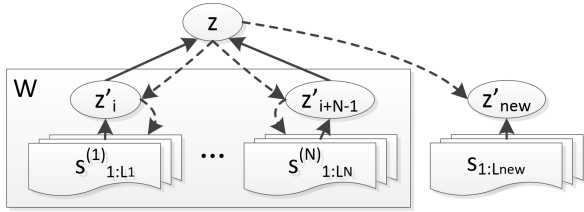

Figure 1: The illustration of HVAE model.

In the *bottom* level of Fig. 1, each document s is viewed as a sample from the *middle* level distribution of z'; On the same principle, each z' distribution is sampled from its corresponding *top* level meta-distribution z.

The HVAE is composed of an encoder and a decoder, where the encoder infers latent distributions from inputs. In Eq. 1 and 2, the meta-distribution among input sentiments $s_{1:L_i}^{(i)}$ from i'th time period are learned by $EncodeModel$. Similarly, the meta-distribution among all period distributions are achieved as 3 and 4 show.

On the other hand, the decoder module is applied to generate inputs by making use of the learned latent variables. In Eq. 5 and 7, z and z_i' are sampled

Table 1: Annotations of our method

W	N length slide window.
s	A sentiment document, with its superscript indicating its time period.
L_i	Data quantity in i'th period.
z', z	Latent meta-distributions of period and window, respectively.
θ, ϕ	Parameters of decoder and encoder.
dash/solid line	Decode/Encode process.

$$(\mu'_{\phi_i}, \sigma'_{\phi_i}) = \text{EncoderModel}_\phi(s_{1:L_i}^{(i)}) \quad (1)$$

$$z'|s_{1:L_i}^{(i)} \sim N(\mu'_{\Phi_i}, \sigma'^2_{\Phi_i}) \quad (2)$$

$$(\mu_\Phi, \sigma_\phi) = \text{EncoderModel}_\phi(z'_{1:N}) \quad (3)$$

$$z|z'_{1:N} \sim N(\mu_\phi, \sigma_\phi^2) \quad (4)$$

from (μ_Φ, σ_ϕ) and $(\mu'_{\phi_i}, \sigma'_{\phi_i})$, respectively. Both $EncodeModel$ and $DecodeModel$ are neural networks. During the model training, two modules keep interacting and all parameters are updated through gradient descent, until inputs are well fitted and latent distributions are finally obtained.

The object function is shown in Eq. 9, where $S = \{s_{1:L_1}^{(1)}, s_{1:L_2}^{(2)}, \ldots, s_{1:L_N}^{(N)}\}$ are historical data in window, and $\log p(S)$ is the log-likelihood for fitting inputs. The Evidence Lower Bound (ELBO) of the log-likelihood is increased through training, and latent meta-distribution are learned at the same time.

$$(\mu_\theta, \sigma_\theta) = \text{DecoderModel}_\theta(z) \quad (5)$$

$$z_i'|z \sim N(\mu_\theta, \sigma_\theta^2) \quad (6)$$

$$(\mu'_{\theta_i}, \sigma'_{\theta_i}) = \text{DecoderModel}_\theta(z_i') \quad (7)$$

$$s_j^{(i)}|z_i' \sim N(\mu'_{\theta_i}, \sigma'^2_{\theta_i}) \quad (8)$$

Note at the bottom level of HVAE, input information of each time period are compressed to a less noisy condensed representation in middle level distribution form. In the same spirit, information from middle level time periods of a window are also compressed into a concise meta-distribution. Through training, the representativeness of all meta-distributions are increased. As such, the follow-

$$\log p(S) \geq$$

$$E_{q_\phi(z'_{1:N}, z|S)} \left[\overbrace{\log p(z) + \log p_\theta(z'_{1:N}|z) + \log p_\theta(S|z'_{1:N})}^{Decode} \atop \underbrace{-\log q_\phi(z|z'_{1:N}) - \log q_\phi(z'_{1:N}|S)}_{Encode} \right] \quad (9)$$

up drift comparison step can benefit significantly from the *better learned distribution representations*, where certain level of data chaotic issue that frequently occurs in sentiment drift will be more tolerated than existing methods.

2.2 Drift Measure

The drift measure evaluates the *distributional difference* between historical and newly arrived data. Note this is different from existing methods which are typically based on classifier performance degradation; here we directly compare data distribution and thus more effective. In particular, we choose two distributions, namely $z'|z$ and $z'_{new}|S_{new}$, for drift measuring. Difference between two distributions are illustrated as shading area in Fig. 2. Obviously, the bigger the size of shading area, the smaller the similarity between the new and history data distribution, which has become a part of built-in drift detection algorithm of HVAE model. Mathematically, the irregular shading area can be computed as the integration of distribution difference. Correspondingly, we propose a new measure, namely, *Accumulation of Distribution Differences* (ADD) as Eq. 10. The parameters of latent distribution over newly arrived z'_{new} are $(\mu'_{\phi_{new}}, \sigma'_{\phi_{new}})$, and the parameters of latent distribution within latest historical data window $z'|z$ are $(\mu_\theta, \sigma_\theta)$. The p is the drift indicator, whose value is larger when the drift is more significant and vice versa.

$$p = \int \left| N(x; \mu'_{\phi_{new}}, \sigma'^2_{\phi_{new}}) - N(x; \mu_\theta, \sigma_\theta^2) \right| dx \quad (10)$$

More specifically, we compute the *intersections* of the two distribution curves, which are named as x_1 and x_2 ($x_1 = x_2$ when there is only one intersection, i.e. both distributions are same, and make $x_1 \leq x_2$ valid). The intersections split the curves into segments whose probability difference accumulations are normalized to [0,1] as our final drift score.

$$p = \left\{ {|F_1(x_1) - F_2(x_1)| + \atop |F_1(x_2) - F_1(x_1) - [F_2(x_2) - F_2(x_1)]| \atop + |F_1(x_2) - F_2(x_2)|} \right\} \Big/ 2 \quad (11)$$
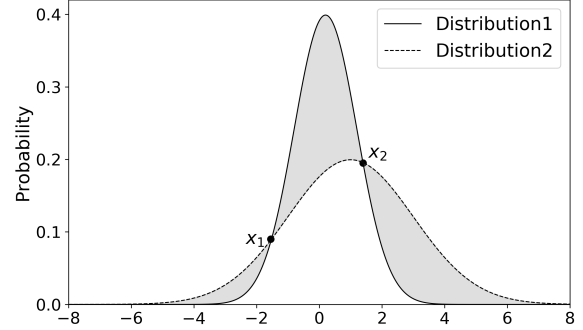

Figure 2: The difference between two gaussian distributions, which is represented as the shading area, the x_1 and x_2 are intersections of distributions. The reason of gaussian distribution assumption is that mean sentiment of across multiple documents is more likely to perform as gaussian according to the Central Limit Theorem.

Under the condition of processing extremely unstable inputs (big drifts), many of the values will be very close to 1, which decreases system performance. Hence, for increasing sparsity, the score can be squared to be the final drift score, i.e., p^2 from Eq. 11, named as ADD_2.

Through the ADD, all sentiment drifts of each time period is collected to current window W. Whether to update the HVAE with new data in the window can be viewed as a Bernoulli experiment, which applies the parameter $\bar{p}_i = \text{mean}(p_1, p_2, \cdots, p_N)$ and deviation $\sigma_i = \sqrt{\bar{p}_i(1 - \bar{p}_i)/i}$. The parameters are inputted to SPC method (Bouchachia, 2011) for drift detection, and retrain/update model with next window once it alarms for drift occurring. If SPC does not incur alarm, the window moves one time period and meanwhile obtains new parameters as new data arrives.

3 Experiments

We have conducted extensive experiments to evaluate our proposed HVAE model.

3.1 Datasets & Baselines

We employ two datasets for our experiments, including Twitter data and CIRCLES data. Twitter

3764

Sentiment[1], containing 1.6m tweets created from 2009-04-06 to 2009-06-25. The dataset is split by hours (time periods), and we delete those periods with empty categories, resulting in 432 periods. Note inputs of tweets are sentiment labels whose format is one-hot.

The second dataset is CIRCLES, which is taken from Gama et al. (2004) and sampled from a uniform distribution in which $x \in [0, 1.2], y \in [0, 1]$. It contains 40 data blocks, each of them contains 10 time periods which apply one of four kind category boundaries (in Tab. 2). Each time period contains 100 2-dimensions numeric vectors. The CIRCLES data is noise-free and used for simulating gradual drift scenario. Both datasets are used to validate model performance, representing both noise-free (ideal) and real-world scenario.

Table 2: The four circles category boundaries

Center	[0.2,0.5]	[0.4,0.5]	[0.6,0.5]	[0.8,0.5]
Radius	0.15	0.2	0.25	0.3

We compared with three state-of-the-art systems, including: 1) Nguyen et al. (2018), the VAE cooperated with a built-in drift detection method, 2) Zhou et al. (2018), an improved EWMA algorithm (Raza et al., 2015) based on statistic chart, 3) Iosifidis et al. (2017), stream sentiment classification based method.

3.2 Experimental Settings

Given a sequence of tweets, we will detect sentiment drifts across different time periods (cutting points). As such, drift detection is treated as sequence segmentation task, and the better segmentations, the higher the overall accuracy. Therefore, experiment results of our model and all baselines are compared with the same metric, i.e., overall accuracy. For running Nguyen et al. (2018) and Zhou et al. (2018) on the two datasets, their drift detection components are implemented and tested. Our experimental settings are the same with existing methods Iosifidis et al. (2017). Accumulative multinomial Naive Bayes (Accumulative MNB) is employed as the sentiment classifier. Data is processed from a new time period according to the principle of prequential evaluation, and drift adaption is done in an *rebuild* way. Specifically, if drift does not occur when a new period arrives, labels of data are predicted and then appended to the training set to retrain the classifier. Otherwise, the

current training set is abandoned, and the classifier is updated with a new window. In the case of the CIRCLES, each detected drift is viewed as a category boundary switch and data before the drift are evaluated by the previous boundary.

For ablation experiment, several variations of HVAE models are generated: No_D does not apply decoder module, while No_E does not apply encoder module. Finally $Plain$ has only one level meta-distribution.

3.3 Experimental Results & Analysis

Table 3 shows HVAE with ADD_2 setting achieves the best accuracy, i.e. 0.55% better than HVAE with ADD. In addition, HVAE is 3.97%, 4.33%, 8.75% better than Nguyen et al. (2018), Zhou et al. (2018) and Iosifidis et al. (2017) respectively, indicating it is extremely effective for sentiment drift detection.

Table 3: Accuracy comparison on Twitter data.

Mea. Mod.		ADD	ADD_2
HVAE		0.827	**0.833**
Nguyen et al. (2018)		0.793	
Zhou et al. (2018)		0.789	
Iosifidis et al. (2017)		0.745	
HVAE Ablation	Plain	0.798	0.769
	No_D	0.8011	0.800
	No_E	0.826	0.829

Table 4 shows HVAE with ADD_2 once again achieves best result and is 10.5% and 13.5% better than two existing methods. Note we did not compare with Iosifidis et al. (2017), as it cannot be applied to numeric data. For ablation study, we can clearly see the importance of our proposed 3-level hierarchical structure for meta-distribution learning, encoder and decode modules respectively.

Table 4: Accuracy comparison on CIRCLES data.

Mea. Mod.		ADD	ADD_2
HVAE		0.315	**0.350**
Nguyen et al. (2018)		0.245	
Zhou et al. (2018)		0.215	
HVAE Ablation	Plain	0.265	0.310
	No_D	0.262	0.270
	No_E	0.207	0.235

According to results, HVAE achieves better performance than all baselines and ablation models,

---

[1]http://help.sentiment140.com/

which validate the effectiveness of several novelties in our proposed model. It is obvious that all models have much better results on tweets than CIRCLES, as gradual drift detection is more difficult. Generally, the resutls of ADD_2 measure algorithm is superior than the ADD, except for results which apply Plain model with the Twitter dataset. Since the sentiment fluctuation of tweets are stable in most of time periods, the disadvantage of ADD may not be so obvious. Moreover, the Plain manner without latent meta-distributions is lack of generalization capability, which performs much worse than the superior ADD_2. In the ablation experiment on CIRCLES, different from results with Twitter data, No_E performance worst and Plain is best. This is because the CIRCLES data are sampled from uniform distribution, but the No_E model is based on Gaussian distribution assumption and it lacks of Encoder part for input fitting. The Plain suffers less performance loss from distribution assumption since it has no meta-distribution structure. The results indicate the importance of HVAE's all innovative components where we need to extract latent distributions from inputs (encoder) as well as fit them with the distributions (decoder).

4   Conclusions

To tackle challenges in sentiment drift detection, we have proposed a novel HVAE model, which is 3-level meta-distributions to extend VAE over *hierarchical* structure, leading to effective learning latent distribution representations of input sentiments. In addition, a new drift measure is designed to effectively measure distribution difference between historical and newly arrived data. Different from existing classifier and threshold based models, the proposed method *directly measures the distribution differences* and thus is more effective. Finally, extensive experimental results demonstrate that HVAE performs significantly better than three state-of-the-art techniques across two benchmark datasets, indicating that it can be effectively used for real-world public sentiment drift analysis.

Acknowledgements

References

Albert Bifet and Ricard Gavalda. 2007. Learning from time-changing data with adaptive windowing. In *Proceedings of the 2007 SIAM international conference on data mining*, pages 443–448. SIAM.

Abdelhamid Bouchachia. 2011. Fuzzy classification in dynamic environments. *Soft Computing*, 15(5):1009–1022.

Xianghua Fu, Yanzhi Wei, Fan Xu, Ting Wang, Yu Lu, Jianqiang Li, and Joshua Zhexue Huang. 2019. Semi-supervised aspect-level sentiment classification model based on variational autoencoder. *Knowledge-Based Systems*, 171:81–92.

Joao Gama, Pedro Medas, Gladys Castillo, and Pedro Rodrigues. 2004. Learning with drift detection. In *Brazilian symposium on artificial intelligence*, pages 286–295. Springer.

Tai Hoang, Huy Le, and Tho Quan. 2019. Towards autoencoding variational inference for aspect-based opinion summary. *Applied Artificial Intelligence*, pages 1–21.

Linmei Hu, Juanzi Li, Liqiang Nie, Xiao-Li Li, and Chao Shao. 2017. What happens next? future subevent prediction using contextual hierarchical lstm.

Vasileios Iosifidis, Annina Oelschlager, and Eirini Ntoutsi. 2017. Sentiment classification over opinionated data streams through informed model adaptation. In *International Conference on Theory and Practice of Digital Libraries*, pages 369–381. Springer.

Diederik P. Kingma and Max Welling. 2014. Auto-encoding variational bayes. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*.

Xiang Kong, Bohan Li, Graham Neubig, Eduard Hovy, and Yiming Yang. 2019. An adversarial approach to high-quality, sentiment-controlled neural dialogue generation. *arXiv preprint arXiv:1901.07129*.

Yang Li, Ying Lv, Suge Wang, Jiye Liang, Juanzi Li, and Xiaoli Li. 2019. Cooperative hybrid semi-supervised learning for text sentiment classification. *Symmetry*, 11(2):133.

Z. Liu, Y. Lin, M. Wang, and Z. Lu. 2016. Discovering opinion changes in online reviews via learning fine-grained sentiments. In *2016 IEEE 2nd International Conference on Collaboration and Internet Computing (CIC)*, pages 1–10.

Thi Thu Thuy Nguyen, Tien Thanh Nguyen, Alan Wee-Chung Liew, and Shi-Lin Wang. 2018. Variational inference based bayes online classifiers with concept drift adaptation. *Pattern Recognition*, 81:280–293.

Haider Raza, Girijesh Prasad, and Yuhua Li. 2015. Ewma model based shift-detection methods for detecting covariate shifts in non-stationary environments. *Pattern Recognition*, 48(3):659–669.

Gordon J Ross, Niall M Adams, Dimitris K Tasoulis, and David J Hand. 2012. Exponentially weighted moving average charts for detecting concept drift. *Pattern recognition letters*, 33(2):191–198.

M. Tsytsarau and T. Palpanas. 2016. Managing diverse sentiments at large scale. *IEEE Transactions on Knowledge and Data Engineering*, 28(11):3028–3040.

Daling Wang, Shi Feng, Dong Wang, and Ge Yu. 2013. Detecting opinion drift from chinese web comments based on sentiment distribution computing. pages 72–81, Berlin, Heidelberg. Springer Berlin Heidelberg.

Chuhan Wu, Fangzhao Wu, Sixing Wu, Zhigang Yuan, Junxin Liu, and Yongfeng Huang. 2019. Semi-supervised dimensional sentiment analysis with variational autoencoder. *Knowledge-Based Systems*, 165:30–39.

Rui Xia, Feng Xu, Jianfei Yu, Yong Qi, and Erik Cambria. 2016. Polarity shift detection, elimination and ensemble: A three-stage model for document-level sentiment analysis. *Information Processing & Management*, 52(1):36–45. Emotion and Sentiment in Social and Expressive Media.

Shengjia Zhao, Hongyu Ren, Arianna Yuan, Jiaming Song, Noah Goodman, and Stefano Ermon. 2018. Bias and generalization in deep generative models: An empirical study. In *Advances in Neural Information Processing Systems*, pages 10792–10801.

Dan Zhou, Liu Liu, and Xin Lai. 2018. The improved ewma chart for heteroscedasticity process. *Annals of Data Science*, 5(1):21–27.