Modeling concept drift: A probabilistic graphical model based approach

Hanen Borchani⋆[1], Ana M. Martínez⋆[1], Andrés R. Masegosa⋆[2]⋆⋆, Helge Langseth[2], Thomas D. Nielsen[1], Antonio Salmerón[3], Antonio Fernández[4], Anders L. Madsen[1,5], Ramón Sáez[4]

[1] Department of Computer Science, Aalborg University, Denmark
[2] Department of Computer and Information Science, The Norwegian University of Science and Technology, Norway
[3] Department of Mathematics, University of Almería, Spain
[4] Banco de Crédito Cooperativo, Spain
[5] HUGIN EXPERT A/S, Aalborg, Denmark

Abstract. An often used approach for detecting and adapting to concept drift when doing classification is to treat the data as *i.i.d.* and use changes in classification accuracy as an indication of concept drift. In this paper, we take a different perspective and propose a framework, based on probabilistic graphical models, that explicitly represents concept drift using latent variables. To ensure efficient inference and learning, we resort to a variational Bayes inference scheme. As a proof of concept, we demonstrate and analyze the proposed framework using synthetic data sets as well as a real financial data set from a Spanish bank.

1 Introduction

Classification, which is the task of predicting the class, Y, of an object based on a set of attributes, $\boldsymbol{X}$, describing that object, has been studied extensively in the machine learning community (see, e.g., [1]). A special instance of this general task is the classification of objects in a streaming context, which amounts to observing objects at different points in time $t = t_1, t_2, \ldots$, and at each time-point t classifying the object based on the information collected up to and including time t, $\bigcup_{j:t_j \leq t} \boldsymbol{x}_{t_j}$.

As pointed out in, e.g., [2], doing classification in the context of data streams raises several issues. Among the challenges is that data in a streaming context should not be assumed to be *i.i.d.* First of all, the objects in the stream may not be independent, and, secondly, concept drift [3–5], where the underlying distribution generating the data changes over time, should be anticipated. The main contribution of this paper is a principled approach based on probabilistic graphical models [6] for modeling concept drift using latent (i.e., unobserved) variables. This should be contrasted to what is currently the most commonly

⋆ These authors are considered as first authors and contributed equally to this work.
⋆⋆ Corresponding author: Andrés R. Masegosa. E-mail: andres.masegosa@idi.ntnu.no

used technique to accommodate concept drift, namely to learn a classifier as if the data was *i.i.d.*, monitor classification accuracy, and then restart the learning process as soon as accuracy drops significantly (see, e.g., [4]).

We will exemplify the use of our modeling framework by analyzing the economic status of the customers of a Spanish bank over the period from 2007 to 2014. To keep the analysis as simple as possible, we use the *Naïve Bayes* classifier [7] as our base model, even if other classifiers with better dynamic properties (e.g., [8]) could also have been employed. The analysis is thus a proof of concept for the proposed modelling strategy, where we focus on the model's ability to detect and represent concept drift instead of its predictive performance. A related Bayesian approach to concept drift is studied in [9], where focus is on abrupt concept drift with independent drift regimes. This type of concept drift does, however, not fit with the financial domain considered in this paper, where we have a fixed customer base that exhibits a more gradual drift.

Classification in data streams also raises some computational problems [2], as data may arrive with high velocity and is unbounded in size (therefore requiring that old observations are "forgotten" to avoid running out of computer memory). To deal with this issue, our model analysis builds on the AMIDST toolbox[6]. This toolbox provides an efficient implementation of approximate inference and learning methods for streaming data by utilizing the *Bayesian network* modelling framework [6] complemented with variational Bayes inference and learning procedures [10]. Furthermore, the toolbox interfaces to MOA [11], thereby enabling us to directly draw on existing preprocessing and visualization functionality.

The remainder of this paper is organized as follows: In Section 2 we describe the real-life data set from the Spanish bank in detail, and discuss its most important dynamic features. Section 3 introduces our approach for explicitly modeling concept drift using latent variables, and in Section 4 we briefly sketch the inference machinery employed. In Section 5 we discuss the results obtained from synthetic data as well as the financial data set, and we conclude in Section 6.

2 The financial data set

2.1 Description of the data set

The data set, which was provided by Banco de Crédito Cooperativo (BCC), contains monthly aggregated information for a set of clients of BCC for the period from April 2007 to March 2014. Only "active" clients are considered, meaning that we restrict our attention to individuals between 18 and 65 years of age, who have at least one automatic bill payment or direct debit in the bank. To make the data set as homogeneous as possible, we only retained clients residing in the Almería region (a largely agricultural area in the south-east of Spain), and excluded BCC employees, since they have special conditions. The resulting number of clients is close to 50 000. We note that the number of clients who are active

[6] AMIDST is an open source toolbox available at `http://amidst.github.io/toolbox/` under the Apache Software License.

varies from month to month: clients with missing values for any of the variables for a given month are removed from the data set for that particular month (this amounts to roughly 25% of the clients). These missing values mainly occur in relation to the income and expense variables, and represent an absence of movements for the account in that period. Consequently, the customer population may vary across months. These clients are removed to support the subsequent analysis, and not because of limitations of the inference/learning engine.

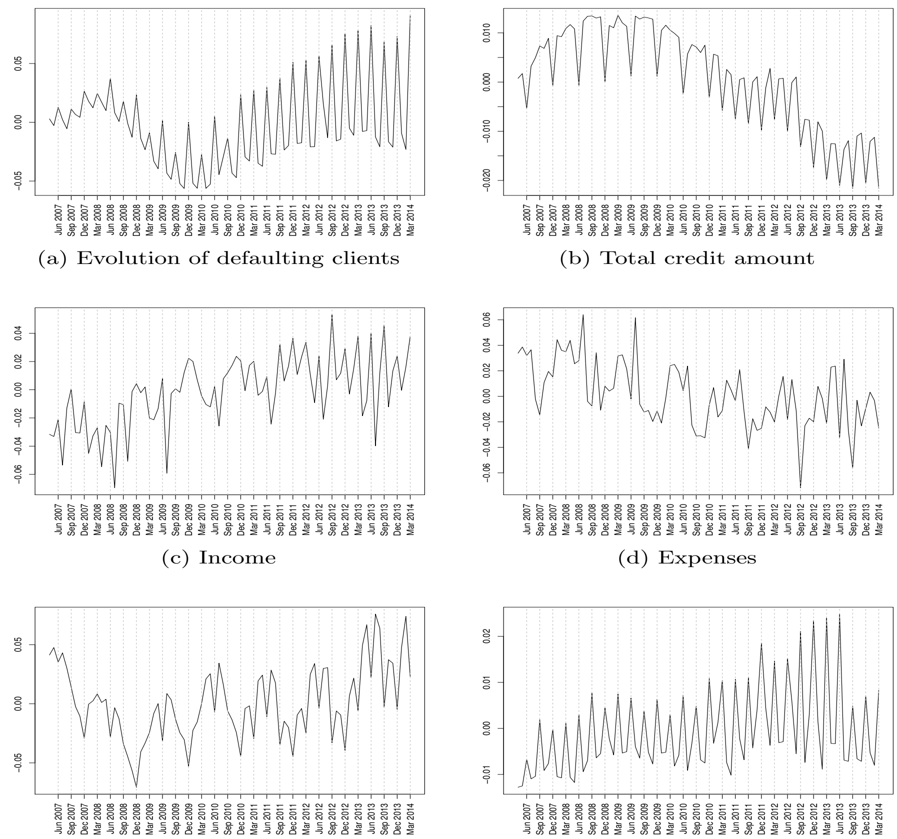
We extracted 11 quantitative attributes, each of which encodes monthly aggregated information for each of the clients. These attributes include, among others, the income, expenses and account balance, the client's total credit amount in all Spanish financial institutions, outstanding payments in mortgages, credit cards, and other personal loans. Each client has an associated class variable, which indicates if that particular client will default during the following 12 months. Fig. 1 (a) shows how the fraction of clients who default increases at the beginning of the period, then decreases for a period of almost two years. Next, the fraction increases again, before it eventually stabilizes; the semester/trimester fluctuations are (partly) a consequence of the changes in the customer base over the period, and will be further discussed in Section 5. We note here that the values in this and in the following figures have been linearly scaled (e.g., we do not report z_t for a particular variable z but rather $\alpha_z + \beta_z z_t$ where α_z and β_z are not disclosed in the paper). The transformation is performed to withhold business-critical information, while at the same time convey meaningful information about the data.

2.2 Financial pre-analysis/context

Fig. 1 (b-f) shows the evolution of 5 of the 11 variables in the domain, namely the total credit amount, income, expenses, account balance, and credit cards. As mentioned above, the values on the y-axis have also been linearly transformed here. The plots reveal that both *seasonal* and *global* trends appear to be present in the data set.

The *seasonal trend* is particularly prominent for the credit reports (Fig. 1 (b)), where the values systematically drop after a period, then go up again. The period between drops is six months for the first half of the data set and three months in the second half. Experts at BCC identified this as the effect of fees being charged to accounts of clients that are normally inactive.

It is also possible to observe a *global ascending or descending trend*, which for this set of variables seems to be on-going until the third or fourth trimester of 2012. Other variables, like defaulted payments on credit cards (Fig. 1 (f)), also display a global trend, but do not follow the same pattern. This variable seems to drop around the third semester of 2013, something experts at BCC attribute to a sale of debt portfolios.



(a) Evolution of defaulting clients

(b) Total credit amount

(c) Income

(d) Expenses

(e) Account balance

(f) Credit cards

Fig. 1. Evolution of (a) defaulting clients and (b-f) 5 of the 11 variables in the financial data set.

2.3 Main challenges

There are two main factors that should be highlighted and which make concept drift detection in the financial data set more challenging than usual. Firstly, the class variable is highly imbalanced and, as shown in Figure 1 (a), the number of defaulting clients varies across time. Hence, monitoring classification accuracy as a way of detecting concept drift can be misleading. Secondly, the data samples arrive in batches of different sizes, i.e., aggregated information for the active customers in a given month. Monitoring concept drift within the samples of one of these batches will not be meaningful as concept drift can only happen from one month to another.

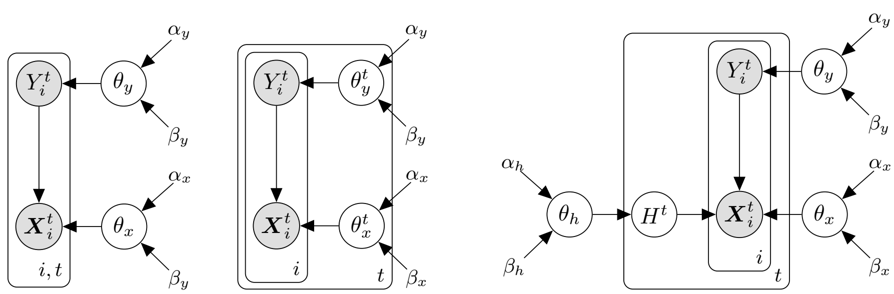
In order to successfully monitor concept drift in the financial data, both of these factors should be addressed.

3 Modeling concept drift using latent variables

In non-stationary domains, the distribution governing the data may change over time. This effect is known as *concept drift* [3–5]. In a classification model, where one wants to classify an instance described by its features $\boldsymbol{x} = (x_1, \ldots, x_n)$ wrt. a class variable y, Gama et al. [5, Eq. (2)] formally define concept drift as the existence of an instance $\boldsymbol{x}$ s.t. $P_{t_0}(\boldsymbol{x}, y) \neq P_{t_1}(\boldsymbol{x}, y)$, where $P_t(\boldsymbol{x}, y)$ denotes the joint distribution over $\boldsymbol{x}$ and y at time t. Concept drift situations can be further classified as either *real concept drift*, when $P_t(y|\boldsymbol{x})$ changes with time, or *virtual concept drift*, when $P_t(\boldsymbol{x})$ drifts while $P_t(y|\boldsymbol{x})$ is constant in t. In this paper the discussions relate to the general notion of concept drift as captured in the expression above, and we do therefore not distinguish between real and virtual concept drift. Concept drift may also appear in many forms, with changes happening abruptly, gradually, incrementally, or with reoccurring behaviour [5].

In what follows we shall consider a new modeling technique for capturing concept drift. The modeling technique will address the general situation, where we, at each time point t, have a collection $(\boldsymbol{x}_i^t, y_i^t)$, for $i = 1 : N_t$, of instances (a.k.a. a window)[7]. We shall assume that concept drift only happens across time steps and not within a collection of instances captured at the same time-point, i.e., the model can only drift every N_t samples.

In a Bayesian paradigm, where the probability distributions are parameterized using latent variables, a simple Bayesian network-based generative model for classification is shown in Fig. 2 (a) using plate notation. In this model the parameters are shared for all points in time t and across all instances, and the model does therefore not provide an explicit representation of concept drift. In Fig. 2 (b) the model is extended to support a simple form of concept drift by duplicating the parameters over time, and thereby allowing them to change.



(a) No concept drift (b) Concept drift (c) Modeling concept drift with a latent variable

Fig. 2. Modeling concept drift through parameter duplication. In all figures, $\alpha_{(\cdot)}$ and $\beta_{(\cdot)}$ are hyper-parameters for the distributions over the parameters θ_x and θ_y.

Alternatively, concept drift can be modeled explicitly using *latent variables*. For simplicity, assume that only the probability distribution $P(\boldsymbol{x}|y)$ drifts. We

[7] For now, we shall assume that the total number of instances N_t does not vary with time t; this assumption is lifted in Section 5 when we consider the financial data set.

can model this using a latent variable H^t, which contributes to the conditional distribution for $\boldsymbol{X}_i$ as illustrated in Fig. 2 (c). The semantics of the H^t-variable is that it determines the "situation" at time t. For example, if the j'th feature $X^t_{i,j}$ follows a conditional normal distribution, we may use H^t to define a time-dependent component of the mean vector:

$$X^t_{i,j}|\{H^t = h^t, Y^t_i = y\} \sim \mathcal{N}(\delta_{j,y} + \gamma_{j,y}h^t, \sigma^2_{j,y}),$$

where $\delta_{j,y}$, $\gamma_{j,y}$, and $\sigma^2_{j,y}$ are elements of θ_x.

The a priori expected level of concept drift can be expressed through the prior distribution for H^t, i.e., using the hyper-parameters α_h and β_h. All observations inside one point in time share the same instance of the H^t-variable, thus concept drift is modelled as a population-wide effect, as desired. Note also that depending on the nature of the variable H^t, this model allows us to represent both *gradual* (H^t continuous) and *abrupt* (H^t discrete) concept drift [5]. Furthermore, the model can easily be extended to model multiple concepts drifts by introducing multiple latent variables, each representing a different drift regime.

Conditioning on the model parameters the concept drift variables are assumed independent across time with no 'memory effect'. If we, on the other hand, expect a gradual form of concept drift, we may wish to capture the drift across time. The model in Fig. 3 reflects this scenario through the dependence relations among the latent H^t variables.

The latent variable models considered so far provide seamless representations of both gradual and abrupt concept drift relating to continuous features. Similar model types are also applicable when modeling abrupt concept drift for discrete features, but when dealing with gradual concept drift we need to move outside the standard class of conjugate Conditional Linear Gaussian (CLG) models. We shall not consider these types of models further in this paper, however, will instead focus on the case where both the feature variables and the latent variable H^t are continuous, using CLG distributions, and where Y^t_i is discrete with a Dirichlet distribution over its parameters.

We would like to reemphasize that the main element of the proposed framework is the use of latent variables for modeling concept drift. In the models presented in this section, these concept drift variables are used to account for concept drift relative to a simple Naïve Bayes classifier. These types of classifiers could in principle be replaced by other types of more expressive probabilistic classifiers, such as dynamic Naïve Bayes models [12] or general Bayesian networks. However, since the main goal of the present paper is to provide a proof of concept for the proposed modeling framework, we will in the remainder of the paper rely on these simpler models.

4 Bayesian inference with streaming data

In the Bayesian paradigm, model learning can be considered an inference process. Given the data seen so far, denoted by D^t, the learning task reduces to computing the posterior distribution over the quantities of interest, i.e., $P(\theta_x, \theta_y, H^t|D^t)$

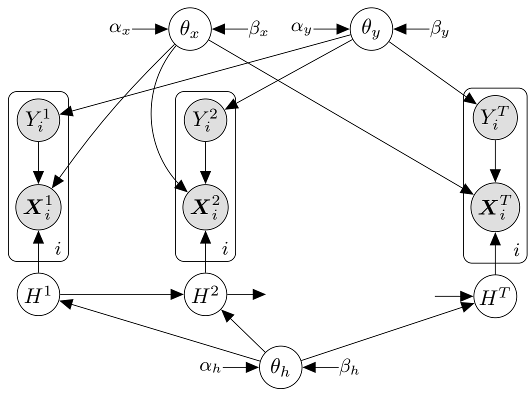


Fig. 3. Concept drift is preserved over time.

for the models described in Section 3. This approach can also naturally be applied when dealing with streaming data. A new data sample $(\boldsymbol{x}_{t+1}, y_{t+1})$ in the stream is included by simply updating the above posterior using Bayes' rule, $P(\theta_x, \theta_y, H^{t+1}|D^{t+1})$.

Inference in Bayesian networks is, however, NP-hard in general [13], and given the size of the data sets we are currently considering, exact inference in the underlying models is not feasible. For the models considered in this paper, we will therefore rely on the *variational Bayes* [14] framework for doing approximate inference and learning; a general introduction to the variational Bayes procedure can be found in [10].

In its general form, one considers the random variables $(\boldsymbol{X}, \boldsymbol{Z})$, where $\boldsymbol{X} = \boldsymbol{x}$ is observed and we want to approximate $f(\boldsymbol{z}|\boldsymbol{x})$. We call the approximation $q(\boldsymbol{z})$, where we for simplicity of notation suppress that $q(\boldsymbol{z})$ depends on the observation $\boldsymbol{x}$. We measure the quality of the approximation by the KL distance from q to f. One popular strategy for minimizing this distance is to assume that $q(\boldsymbol{z})$ factorizes into smaller factors, like for instance its separate variables, $q(\boldsymbol{z}) = \prod_i q_i(z_i)$. This approach is commonly known as the *mean-field* approximation.

The calculations can be structured efficiently in conjugate exponential models using a *message passing* scheme [15]. In this scheme, messages are sent along the edges in the graph based on the (expected) natural parameters of the distributions in the model. The message passing scheme outlined above has been implemented for the model classes presented in Section 3, and forms the basis for the experimental results presented in the following section.

5 Results

The experimental study is divided into two parts. First, we analyse two synthetic data sets, widely employed as benchmarks in the concept drift literature.

Next, we present the results from analysing the financial data set. All the experiments have been performed using MOA [11], where the developed AMIDST model (in Fig. 3) has been integrated as a new Bayesian streaming classifier, named *bayes.amidstModels*. The Java code to reproduce the experiments can be downloaded from `http://amidst.github.io/toolbox/`.

5.1 Synthetic data sets

We first analyse the `SEA` data set [16] containing 60 000 samples, with 3 attributes (x_1, x_2, x_3) and 2 classes ($y = 0$ and $y = 1$). The attributes are numerical and uniformly distributed between 0 and 10. Only two of the attributes are relevant for the class label, y, which is defined as $y^t = 1$ if $x_1^t + x_2^t \leq \epsilon^t$ and $y^t = 0$ otherwise. Concept drift has been created by changing the threshold ϵ^t as a function of t. The data set covers four "phases", each with a duration of 15 000 samples, and with different ϵ^t (9, 8, 7, and 9.5 for the four phases, respectively). Fig. 4 (left) shows the results of this analysis for batches of size N_t equal to 1000. The plot illustrates the progress of the expected value of the latent variable (denoted H^t) as well as the prequential accuracies computed using a sliding windows of size 1000 for a simple Naïve Bayes model (NB) and the adaptive Hoeffding tree model (AHT). As can be observed, the output of our model (i.e., the expected value of H^t) detects the drift points and clearly identifies the occurrences of the four different phases in the data, whereas those phases are less easily detected based on the accuracy results.

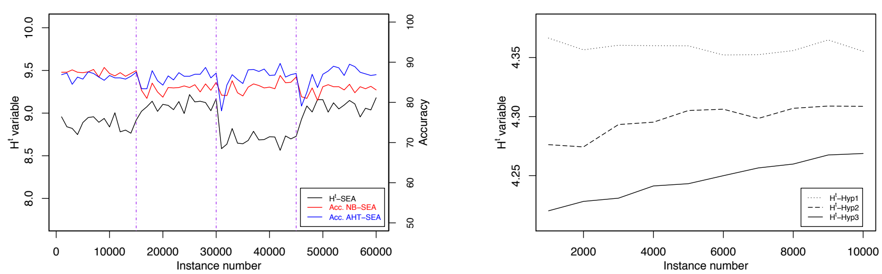


Fig. 4. Left: Results for the SEA data set. Right: Results for the hyperplane data sets

The second data set considered is the rotating hyperplane [17]. This benchmark data set is widely used to simulate "gradual" concept drift problems. We considered three versions of this data set, denoted `Hyp1`, `Hyp2`, and `Hyp3`, each including 10 000 instances. For each data set, 8 out of 9 attributes are drifting but with different magnitudes of change (i.e., 0.1, 0.5, and 1 for the three data sets, respectively), see [17] for details. Fig. 4 (right) shows the evolution of the latent variable H^t for each considered data set using a sliding window of size 1000. Here we see that the different drift magnitudes of the three data sets are directly reflected in the development trends of the latent variables. For instance, for the Hyp1 data, the curve of the H^t variable presents a stable behavior which correctly illustrates the very low change magnitude for this data set, i.e., 0.1.

5.2 Financial data set

In this section we analyze the financial data described in Section 2. Notice that for this data set the batch sizes, N_t, refer to the number of active customer in a given month and can vary from one month to another.

Fig. 5 (left) shows the evolution of the classification accuracy for the NB model using a latent variable. At first sight, the evolution of the accuracy may reflect some inherent trend in the data; however, a more careful analysis reveals that it simply reflects the evolution of defaulters as shown in Fig. 1 (a). This is basically due to the data being imbalanced as pointed out in Section 2, and in such settings the use of classification accuracy for detecting concept drift can be misleading. As an alternative performance measure, we may consider the area under the ROC curve as shown in Fig. 5 (right). The plot provides a more smooth behaviour with gradual performance improvements over time.

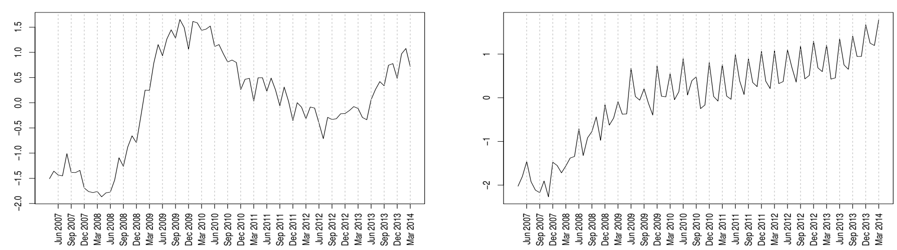


Fig. 5. Evolution of the accuracy (left) and the area under the ROC (right) for the financial data set. For confidentiality reason, the y-axis values have been linearly scaled.

In contrast, in Fig. 6 we plot the evolution of the latent variable H^t over time. Before discussing how this plot may provide insight into the financial data, recall first that in this model (cf. Fig. 3) a single scalar value tries to capture the global trend of 11 variables conditioned on a binary class variable. Moreover, as the vast majority of the clients in the data set are non-defaulters, the latent variable will mostly be influenced by this group of customers. With this basis, at least two observations can now be made about the evolution of the latent variable:

Observation 1: There are regular peaks in the time series. Before 2011, these peaks occur every June and December (6 months period); after 2011 the peaks appear every March, June, September, and December (3 months period). Fig. 6 thus seems to represent two time series, one containing the values at the peaks and one containing the remaining observations. The two underlying series evolve in parallel.

Observation 2: Both underlying series increase rapidly until the second or third trimester of 2012 (the highest points in the two series are reached in

June/July 2012). Afterwards, the series seem to gradually decrease, and this is particularly evident from the third trimester of 2013.

Our interpretation of these observations relies on the figures presented in Section 2, where the temporal evolution of the monthly average of each variable is depicted. To gain some insight into the first observation, we may recall that clients with missing values for *Expenses* or *Income* are discarded when analysing a particular month. We also previously commented that these clients are assumed to be less active than the remainder of the population, and they are consequently not present in the data in the majority of the months; they only appear when fees are deducted from their accounts every semester/trimester (this amounts to approximately 20% of the customers). From the variables in Fig. 1 we can see that these customers have a quite particular profile, which introduces a seasonal effect in the data set. We believe that this is what the latent variable is also capturing. Furthermore, we attribute the fact that the two underlying series are approximately equidistant over time to indicate that they probably represent groups of clients that are similarly affected by the national economical climate.

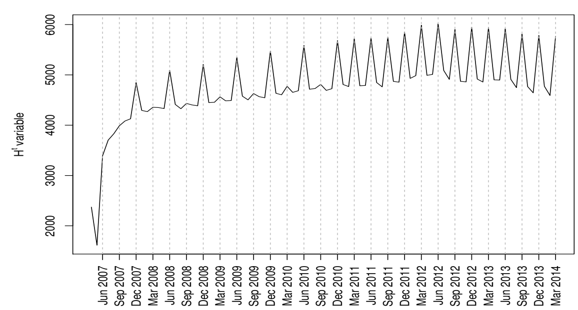


Fig. 6. Evolution of the H^t variable for the financial data set.

Regarding the interpretation of Observation 2, it appears evident that the expected value of the latent variable moves from the very beginning of the period until the second/third trimester of 2012 (the peak is in July 2012), before it remains stable until the second trimester of 2013. Thereafter it moves slightly in the reverse direction. The movements in the latent variable are used to facilitate the evolution of the attributes' distributions, but looking at each variable in Fig. 1 separately, we cannot pinpoint a direct and simple explanation of the above behaviour. For example, both *Expenses* and *Income* continuously move until the first/second trimester of 2012, after which they become more stable. On the other hand, the *Account balance* has a different trend. Thus, when looking at each variable in isolation, it is hard to find a common evolution pattern.

Motivated by the fact that the provided data is expected to reflect the socio-economical status of a significant part of the population in the province of Almería, we now relate the global trend of the latent variable to the history of the financial crisis in Spain during the studied period. In that light, the evolution of the latent variable in Fig. 6 should tell us that the economic climate gets worse from the beginning of the period until the second/third trimester of 2012, then stabilizes until the second trimester of 2013, before it starts recovering slightly. If this interpretation is correct, we should see a correlation between our latent variable and relevant economic factors influencing the socio-economical status of the population during this period of time. Fig. 7 (left) shows the unemployment rate in the province of Almería, which increases from the beginning of the period. We notice some peaks associated with the seasonality of the tourism and agriculture professions, which are two of the main economic drivers of this region. Taking this seasonality into account, the unemployment rate reaches its maximum value around the turn of year 2012/2013 before it slowly improves. Fig. 7 (right) shows the relationship between the unemployment rate and the expected value of the latent variable. From the figure we see a close correlation between these two entities.

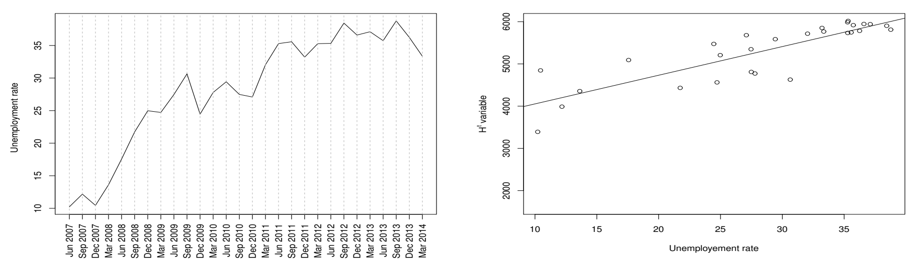


Fig. 7. Economic indicators. Left: Unemployment rate in Almería. Right: Scatter plot of the unemployment rate and the expected value of the latent variable (Spearman's rank correlation coefficient is 0.85).

6 Conclusions

In this paper we have developed a classification model for data streams that is compatible with concept drift scenarios. Our approach distinguishes itself from traditional alternatives by explicitly including the effect of the concept drift in the model using latent variables. We have shown through analysis of both synthetic and real-life data that the model is able to capture and handle both abrupt and gradual concept drift scenarios.

The analysis is a proof of concept for the proposed model class, and the opportunities for future research are manifold: Firstly, we will consider more sophisticated base-classifiers that are better suited for dynamic domains (e.g., the dynamic Naïve Bayes model), which we expect will improve the classification accuracy of the model. Next, we will look at extensions of the concept drift modelling itself, e.g., by using more than one latent variable and thereby being able to represent concept drift that behave differently for different subsets of the variables. Finally, as our development is motivated by the financial dataset, we will look deeper into socio-economical indicators from Spain to understand even better the mechanisms driving the concept drift in this domain.

Acknowledgments

This work was performed as part of the AMIDST project. AMIDST has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 619209. The data set has been provided by Banco de Crédito Cooperativo.

References

1. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer New York Inc., New York, NY, USA (2001)
2. Gaber, M.M., Zaslavsky, A.B., Krishnaswamy, S.: A survey of classification methods in data streams. In Aggarwal, C.C., ed.: Data Streams - Models and Algorithms. Volume 31 of Advances in Database Systems. Springer (2007) 39–59
3. Schlimmer, J.C., Granger, R.H.: Incremental learning from noisy data. Machine Learning **1** (1986) 317–354
4. Widmer, G., Kubat, M.: Learning in the presence of concept drift and hidden contexts. Machine Learning **23(1)** (1996) 69–101
5. Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., Bouchachia, A.: A survey on concept drift adaptation. ACM Computing Surveys **46** (2014) 44:1–44:37
6. Jensen, F.V., Nielsen, T.D.: Bayesian Networks and Decision Graphs. Springer-Verlag, Berlin, Germany (2007)
7. Duda, R.O., Hart, P.E.: Pattern Classification and Scene Analysis. John Wiley and Sons, New York (1973)
8. Zhong, S., Langseth, H., Nielsen, T.D.: A classification-based approach to monitoring the safety of dynamic systems. Reliability Engineering and System Safety **121** (2014) 61–71
9. Bach, S., Maloof, M.: A Bayesian approach to concept drift. In: Advances in Neural Information Processing Systems. (2010) 127–135
10. Jordan, M.I., Ghahramani, Z., Jaakkola, T.S., Saul, L.K.: An introduction to variational methods for graphical models. Machine Learning **37** (1999) 183–233
11. Bifet, A., Holmes, G., Kirkby, R., Pfahringer, B.: MOA: Massive Online Analysis. Journal of Machine Learning Research **11** (2010) 1601–1604
12. Martínez, M., Sucar, L.E.: Learning dynamic naive Bayesian classifiers. In: Proceedings of the Twenty-first International Florida Artificial Intelligence Research Symposium Conference. (2008) 655–659
13. Cooper, G.F.: The computational complexity of probabilistic inference using Bayesian belief networks. Artificial Intelligence **42** (1990) 393–405
14. Beal, M.J.: Variational algorithms for approximate Bayesian inference. PhD thesis, Gatsby Computational Neuroscience Unit, University College London (2003)
15. Winn, J.M., Bishop, C.M.: Variational message passing. Journal of Machine Learning Research **6** (2005) 661–694
16. Street, N., Kim, Y.: A streaming ensemble algorithm (SEA) for large-scale classification. In: 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD). (2001) 377–382
17. Hulten, G., Spencer, L., Domingos, P.: Mining time changing data streams. In: Proceedings of the Seventh International Conference on Knowledge Discovery and Data Mining. (2001) 97–106