Recent Approaches to Drift Effects in Credit Rating Models

Rachael Chikoore[1], Okuthe P. Kogeda[2(✉)], and S. O. Ojo[1]

[1] Department of Computer Science, Faculty of Information Communication Technology, Tshwane University of Technology, Private Bag X680, Pretoria 0001, South Africa
maichicco@gmail.com, OjoSO@tut.ac.za

[2] Department of Computer Science and Informatics, Faculty of Natural and Agricultural Sciences, University of the Free State, P. O. Box 339, Bloemfontein 9300, South Africa
kogedapo@ufs.ac.za

Abstract. Credit Rating is the valuation of the credit worthiness of the borrowing entity, which gives an indication of the borrower's current credit position and the probability of default. A credit rating model must be very accurate in doing its predictions because critical decisions are made based on the classification that would have been made for the prospective borrower. Different changes occur in the environment that would have been used to come up with the initial model, which might not be applicable to the current sample population and this might have an effect on the prediction accuracy. Changes to the data stream, economic climate, social and cultural environment may cause a drift. Drift shows that there is a change in probability distribution of the concept under study. Population drift is an example of concept drift. Having a static credit rating model will bring challenges in future predictions, hence, there is the need for designing a dynamic credit rating system that caters for the changes that might occur to the initial population sample in order to maintain the prediction accuracy of the model. In this paper, a detailed literature study was conducted exploring recent solution approaches to drift effect in credit rating models. A comprehensive recent solutions is presented in this paper that could be a source of information of interested researchers.

Keywords: Credit rating · Drift effect · Concept drift · Population drift · Prediction · Approaches · Credit rating models · Probability distribution

1 Introduction

It is very common for researchers and model designers to assume that the probabilities of class membership, which are conditional on the feature vectors do not change. When model data changes over time, it can result in poor and degrading predictive performance in predictive models that assume a static relationship between input and output variables. This problem of the changing underlying relationships in the data is called concept drift in the field of machine learning.

The reality is that both population drift and concept drift occur and can affect the prediction accuracy of any given model. It is highly likely that there is a change in

R. Zitouni et al. (Eds.): AFRICOMM 2019, LNICST 311, pp. 237–253, 2020.
https://doi.org/10.1007/978-3-030-41593-8_18

population in a credit scoring model. The population may change responding to the change in the economic environment and other social changes that could affect an individual. There is a very thin line that separates population drift from concept drift. Population drift shows that there is a change in probability distribution of the concept under study yet concept drift is said to have occurred to this change in distribution and also other related changes hence concept drift is the broader change in the model environment.

Evolving data means that data distribution changes over time. This can happen in different ways, for example, Feature drift in which the distribution of input data X changes, p(X), Real concept drift in which the relationship between input X and target y changes, $p(y|X)$ and Changing prior distribution, for example, of the target p(y) when the arrival of new information occurs [1].

This paper aims at analyzing the different approaches to managing drift and its effects in credit rating. A detailed literature review was undertaken unearthing a number of recent solution approaches to drift effect in credit ratings.

The remainder of this paper is organized as follows: In Sect. 2, we provide the general credit rating approaches. In Sect. 3, we give the statistical approaches. In Sect. 4, we outline the data mining approaches. In Sect. 5, we explain the hybrid/ensemble approaches. In Sect. 6, we provide the general approaches to handling drift. In Sect. 7, we give the drift effects and proposed solutions. In Sect. 8, we provide the credit scoring models application areas and in Sect. 9, we conclude the paper.

2 General Credit Rating Approaches

Credit rating in different countries is managed by credit rating bureaus. The most common rating is using a range of values and the letters of the alphabet. Scoring can be done for a new applicant in order to estimate the credit risk. Behavioral scoring is based on the client's previous or current credit standing, it is believed that the way an individual handles the previous or the current loan might have a direct link on their future behavior. Collection scoring is used to categorize the clients into groups, depending on their behavior. Fraud detection classifies the applicants according to the probability that an applicant is guilty [2].

Credit rating classifies the methods of coming up with a rating into two categories, Statistical methods and Data mining methods. Examples of statistical methods include but not limited to simple Ordinary Least Square (OLS) and ordered probit model. Examples of data mining methods include but not limited to Decision trees, Multi Class Support Vector Machines and Multi Class Proxial Support Vector Machines (PSVM) [3].

3 Statistical Approaches

Statistical methods are usually used for feature selection and the most common and effective classifiers that have been used in credit scoring are Artificial Neural Networks (ANN) and Support Vector Machine (SVM) [4]. Linear Discriminate Analysis (LDA) is an example of a statistic approach which was first proposed by Fisher as a classification method. It uses a Linear Discriminate Function (LDF) which passes through the centroids of the two classes to classify customers. It is a commonly used approach but its disadvantage is that it requires linear relationships between depended and independent variables and its assumption is that input variables follow a normal distribution. Logistic Regression (LR) does not require normal distributed input variables. It has the ability to predict default probability of an applicant and pick on the variables related to his behavior. Multivariate Adaptive Regression Splines (MARS) is a non-linear and non-parametric regression method and is excellent in dealing with high dimensional data. It does not presume that there is a linear association amongst the dependent and independent variables. It has a short training time and also has a strong intelligibility.

4 Data Mining Approaches

A comparison on the most commonly used data mining methods in credit rating was done. The research aimed at bringing awareness to researchers on the various approaches that can be applied in credit scoring [5]. Data mining methods have been widely used due to their ability in discovering practical knowledge from the database and transforming it into meaningful information. Data mining approach to credit rating was classified as ranging from Neural Networks, Bayesian Classifier, Logic regression, K-Nearest Neighbor, Decision tree, Survival analysis, Fuzzy rule based system, Support Vector machine and also Hybrid methods. The research concluded that the Support Vector machine approach was commonly used.

Paleologo *et al.* [6] did a comparative study on the accuracy of classification models in order to reduce the credit risk. They used data mining of the enterprise software to come up with four classification models. The four models were decision tree, logic regression, neural networks and support vector machines. The conclusion indicated that the support vector machine models performed better than the decision tree, logic regression and neural network model.

Zhong and Li [7] gave a summary of the different approaches that are used in credit scoring and introduced a new approach called ensemble learning model. The main focus of their study was to show that it is important to move away from the static approach to credit scoring towards a dynamic approach. They summed up the approaches to credit scoring into Statistical models, Artificial Intelligence (AI) models, Hybrid models and also Ensemble methods.

Bayesian classifiers are white box in nature. They predict that a given sample belongs to a certain class. They are believed to be highly accurate when used in prediction. Decision trees use the recursive partitioning approach to prediction. The virtual subdivisions are done on customer data to make the homogeneity of default risk in the subset greater than the original set. Division continues until the new subsets meet the specifications of the end node. C4.5 and CART are the most popular approaches used in credit scoring. Markov models perform predictions based on past trends. It uses historical data to predict the distribution of population at any given time.

Apart from Statistical methods, Artificial intelligence methods such as Artificial Neural Networks (ANN), Genetic Algorithms (GA) and Support Vector Machines (SVM) have been use in credit scoring. ANN comprises a large number of nodes which are linked and they receive signals from the pre-layer and output them into the next layer. The feed–forward network with Back Propagation (BP) is widely used for credit scoring. ANN possess a strong learning ability and does not make assumptions on the relationship between input variables. ANNs' major disadvantage is their black box nature hence they are difficult to understand and also it is difficult to design the network for experimental purposes because of its complexity and also long training time and demand for large training samples.

SVM is easy to implement on small samples and does not limit the data distribution. Many researches that did work around the implementation of SVM in credit scoring concluded that it was a superior approach to ANN in the aspect of classification accuracy, however, SVM is also black box in nature hence it is complicated to implement on large data. GA simulate natural selection of Darwinian biological evolution theory in order to search for an optimal solution. It is self-adaptive, globally optimal and robust, because it is evolutionary in nature, it does not need to understand the inherent nature of the problem, be it linear or non-linear, continuous or discrete.

KNN is a clustering method that learns by analogy. It searches the pattern space for training samples that are closest to the unknown sample. Hsieh *et al.* [8] built a new classifier called Clustering Launched Classification (CLC) and concluded that it was more effective than the SVM approach. Case Based reasoning is another approach that can be used in credit scoring and it works in such a way that it compares each case to see if there is an identical one, if not, the search continues until an identical case is found.

5 Hybrid/Ensemble Approaches

Hybrid models have become very common in credit scoring. A hybrid is a combination of different approaches in order to come up with one model. Simple hybrid models are made from performing three steps namely, feature selection, determination of model parameters and then classification. The hybrid models can also be built from a class wise classifier. In this approach samples of data are clustered in order to decide the number of labels before they are classified. Ensemble classification is another approach that can be used in credit scoring. It works by first producing several classifiers to obtain classification results trained on different samples.

The difference between Ensemble classification and hybrid approaches is that the Hybrid approach uses one classifier for learning while Ensemble learning produces various classifiers with different parameters. There is a difference between credit scoring and behavioural scoring, the difference being that, the latter is a dynamic approach to credit scoring. Zhong and Li [7] recommended that it is important to incorporate economic conditions to credit scoring since they have an effect on the evolution criterion of credit institutions. They suggested that, when the economic conditions are in a depression, evolution criteria should be lenient to increase the revenue for credit institutions.

6 General Approaches to Handling Drift

Credit rating models that are able to adapt to changes in population samples due to unexpected changes in the economic conditions are the most favourable ones. Drift can be classified into two categories namely abrupt and gradual drifts. When the drift is abrupt, distribution changes at different time points normally referred to as change points. When moving from one change point to the other, the distribution is static. When the drift is gradual, the distribution changes at each time step. A general approach to handle population drift would be to re-estimate the parameters of the classifier at different intervals using a current section of the available data. Unfortunately this approach is not so accurate because population drifts occur at unpredictable times [2].

Another approach that is common is to rebuild the classifier using current data after observing a major degradation in the prediction accuracy. Rebuilding can be a good approach if the population drift does not occur soon after rebuilding, if this occurs, it means all the predictions done between the drift period and the next rebuilding will be inaccurate. Another challenge to this approach is the time factor; it is time consuming to start gathering new data for new clients in order to use it for the rebuilding process.

Forgarty [9] investigated how genetic algorithms can be used in credit scoring. Genetic algorithms work in a way that is similar to the way biological genes are passed from one generation to the next. He noted that one of the key challenges faced in maintaining scoring systems include population drift which happens to be a process by which scores deteriorate over a certain given period. As data grows in this bid data era, most organizations do not have enough resources to curb the population drift problem hence do not commit enough resources to mitigate this situation. The common approach that has been used to deal with population drift is the use of sophisticated semi-automated prediction tools, which operate in the background to ensure that the models meet minimal standards of efficiency. These tools include genetic algorithms. The concept behind genetic algorithms is beneficial because it enables the production of robust and effective model offspring from a given set of surviving genes thereby curbing the population drift challenge.

In order for Forgarty [9] to prove that genetic algorithms can be used to facilitate scoring system maintenance functions, he created a proprietary genetic algorithm that could facilitate the credit scoring function. The genetic algorithm that is implemented in scoring models is based on Darwin's survival of the fittest principle hence in this case, better scoring models must live longer as compared to poor performers. The genes represent the data attributes or independent variable, chromosomes are composed of genes and represent a model. The evolutionary process starts by breeding the first population of randomly selected models and evaluating the worthiness of each model, then new generation models are bred by cloning, mating and mutating with the best evolved model being the final solution.

The proprietary genetic algorithm model was applied to the datasets on 25 models developed in a traditional manner using logistic regression using data of a large finance company. The genetic algorithm outperformed the traditionally built models thus indicating that they are suitable for model redevelopment maintenance functions. In

72% of the trials, the genetic algorithms performed better than the traditional techniques, 16% of the trials, genetic algorithms and the traditional ones performed the same then 12% of the trials failed due to data over fitting. There was never a situation when the traditional ones outperformed the genetic algorithms. The genetic algorithms are best used for maintaining credit score systems in order to curb population drift but are not very popular because they require high processing power and also it is difficult to come up with a genetic model for all class of problems and data in credit scoring scenarios. Forgaty [9] suggested the use of neural networks and support vector machines as other methods that could solve problems in credit scoring.

7 Drift Effects and Proposed Solutions

Hand and Adams [10] explained the problems that arise when inferential statements are made about a population based on a non-random population sample. One example that they concentrated on was the issue of reject reference, a situation that arises when a score card is built only from previously accepted customers. In credit scoring, it is of paramount importance to keep evaluating the existing scorecards to maintain the prediction accuracy. Usually performance degrades due to changes in applicant population, economic climate and also the competitive environment. If there is a bias in the selection, the resulting situation will be favouring a new scorecard as compared to the existing one. When a new scorecard is favoured, unnecessary costs might be incurred in an effort to phase out the old score card. More costs like the cost of new software and retraining of staff are incurred.

Kelly *et al.* [11] performed a study on a UK dataset of 92958 unsecured personal loans with a 24 month term from 1993 to 1997. They used two classifications to categorize the clients. A client with at least three months in arrears was classified as bad and good was for opposite scenario. They observed that there was need for a waiting period of about 2 years after the granting of the loan for a client to be classified in their accurate class. This misclassification is evidence enough to show that there is some shifting of classification conditions that occurs to a model overtime. In most cases the change in the posterior distribution of the class membership is the one that affects classification accuracy.

Kadwe and Suryawanshi [12] in their review on concept drift did an analysis on how concept drift occurs and how it affects the performance and accuracy of the model. The general current trend is the handling of large streams of data and this data can change over a certain period of time and this is what is being referred to as concept drift. Concept drift occurs when the concept about which data is being collected changes from time to time. There is that need to incorporate a model that can detect the changes so that when there is a change in concept it will be easily noticed and dealt with. There is no way one can detect that a change has occurred unless the change has been detected and also the level of the change has been noted. It is against this background that there exist different kinds of concept drift. The relation between the input data and target variable makes concept change take different forms. Concept drift between time point *t0* and time point *t1* can be defined using Eq. (1) as:

$$\exists X\colon pt0(X, y) \neq pt1(X, y) \tag{1}$$

Where pt0 denotes the joint distribution at time t0 between the set of input variables X and the target variable y. There are three ways in which concept drift may occur and these are prior probabilities of classes, p(y) may change over time, class-conditional probability distributions, p(X, y) might change and also posterior probabilities p(y|X) might change hence concept drift can then be grouped in terms of reason of change and also the rate at which the change has taken place. When the class labels are different from point to point, that is referred to as a real drift but when the concept does not change but there is a change in data distribution, it is called a virtual drift. Drifts can be sudden, incremental and also gradual. When it is a sudden drift, there is a sharp change between the old data and the current data. When it is incremental, the change will be happening sequentially over a given time and when it is gradual, there could be sudden change in some cases but the concept might go back to the previous trend then have a sharp change again. A lot of real world applications are subject to drift issues and these range from monitoring systems, personal assistance systems, Decision support systems and also artificial Intelligence Systems.

Knotek and Pereira [13] did a survey on concept drift. When instances are no longer from the same distribution, it is an indication that concept drift would have occurred. As indicted by other authors, Knotek and Pereira [13] mentioned that concept drift can occur in different ways and these are sudden, gradual, incremental and reoccurring and suggested some of the main approaches to handling it. They suggested the following approaches: instance selection, instance weighing, ensemble learning and statistical methods.

Instance selection mainly deals with the identification of instances that are applicable to the current concept. In general there will be a window that moves over recently arrived instances, as new examples arrive they are placed into the beginning of the window, the same number of examples are removed from the end of the window, and the learner is reapplied. The learnt concepts are used for prediction only in the immediate future. In some cases the window size can be changed and in some cases it can be fixed.

In Instance weighting, the main concept lies behind forgetting. The new training examples are regarded as of higher priority than the older ones and their priority is reduced as time lapses. A gradual forgetting function or a kernel function can be used to calculate the weights. SVM and neural networks have been used for this purpose.

Ensemble learning keeps a set of concept descriptions and the predictions are combined using voting or the most relevant description is selected. In most cases the weight is a function of the previous performance and shows the future capability of the base learner. When accuracy of the model takes precedence as compared to the time taken to run and update the ensemble, an ensemble is a very good solution to solving concept drift. Statistical methods usually calculate a statistic that picks the similarity between two example sets of a multivariate data. The value of the statistic is then compared to the expected value under the null hypothesis that both sets are sampled from the same distribution. The resulting value can be seen as a measure of the extent to which concept drift would have occurred. There is no concrete approach to solving concept drift problems. Any given solution depends on the problem domain and also nature of the data that is being used, if the data is artificial, the model may behave differently when exposed to real world problems.

Sun *et al.* [14] did a concept drift adaptation framework by exploiting historical knowledge. They came up with a new model that could adapt to changes in the concept. They named the ensemble Diversity and Transfer based Ensemble Learning (DTEL). They did experiments with 15 synthetic data streams and also with 4 real life situations and observed that the results given by DTEL were better than the existing approaches to adapt to concept drift changes. Assuming that some new data has just arrived, DTEL uses each preserved historical model as the starting model and further trains it with the new data through transfer learning. In general, there are two research questions that one has to answer during the design of an ensemble method for incremental learning with concept drift. These two questions are, which previously trained models are we going to maintain for use in future and also which approach is going to be adopted to exploit the maintained models to enable future learning with concept drift?

DTEL makes use of a decision tree as the base learner and a diversity-based strategy for maintaining historical models. When there is the introduction of new data, the preserved models are exploited as the stating points for training the new models, then the recently acquired models are joined to form the new ensemble. This approach becomes very effective in curbing population drift challenges and hence improves on the prediction accuracy.

Zliobaite *et al.* [15] did an analysis of the current approaches that are being used in adaptive learning with an effort to make it more applicable in real life situations. They appreciated that there is a growing amount of data that is continuously being exchanged and requires real time processing and mining so that it becomes meaningful. The fact that data evolves over time is reason enough to make a predicting model flexible and to adjust accordingly so that it does not give false predictions as the data evolves. Zliobaite *et al.* [15] identified six areas which need to be addressed when one wants to build a meaningful adaptive model and these are, making adaptive systems scalable, dealing with realistic data, improving usability and trust, integrating expert knowledge, taking into account various application needs, and moving from adaptive algorithms towards adaptive tools. Making adaptive systems scalable can be achieved by making sure that data is processed as it arrives or it can be handled at the hardware side where its main application areas are in cloud computing, grid computing and mobile applications. Scaling up can also be achieved by dedicating the training process to different processors so that the task is not only managed on one machine. The building of new incremental algorithms and the conversion of current learning algorithms so that they will be able to work in incremental online mode, building techniques that allow learning algorithms to work in changing hardware environments, coming up with approaches that would be able to run data mining algorithms in resource aware devices and also developing algorithms that are able to return an estimate of the correct solution, depending on the amount of processing they were able to perform would be a good way to achieve scalability.

There is a difference between dealing with a synthetic data stream and dealing with a real data set. Usually the general assumption is that data comes in an orderly way and it would have already been pre-processed such that feedback is instantaneously available after performing each prediction and before the arrival of new data. Real data usually comes from complex environments and is usually noisy, redundant and sometimes has missing values hence it is very important to dedicate much time to pre-processing when dealing with real data in order to achieve high levels of prediction accuracy. Improving usability and trust can be done by making sure that the process of setting parameters that determine that the prediction is made easy for the user or better still to make the parameters self-adjusting to suit the change in environment. When one is building an adaptive system, there is that need for them to merge their proposed model with the knowledge that has already been unearthed by the experts in this field. A lot of research work has been conducted but some of the work is not being used to solve real life problems. Adaptive learning models can be used in different application areas and it is of paramount importance to note that there is no software tool that has been developed so far to solve adaptive learning problems in different application areas using one solution. Coming up with such a software solution is quite a challenge because it requires the synchronization of different data sources, data formats, size and also arrival frequency so that they become easy to process. Zliobaite *et al.* [15] appreciated that adaptive learning has come to help solve real life problems but a lot still needs to be done to increase its use.

Krempl and Hofer [16] did a study on population drift. They highlighted that it is of paramount importance to study the different categories of drift because many adaptive classification methods make assumptions concerning the drift type. They highlighted that drift can affect the Posterior distribution P(Y|X), the feature distribution P(X) and the class Prior distribution P(Y) where X is the explanatory variables or features and Y is the binary response. They also defined verification latency as the time interval between classification and verification of the prediction. Drift mining needs to be done on prediction systems in order to come up with ways to handle the drift to improve on accuracy. There is that need to identify and track sub-populations in un-labelled data over time in order to observe the drift pattern. Krempl and Hofer's [16] work helps in understanding how the population drift comes into play to degrade the prediction accuracy.

A new Dynamic Credit Scoring Model Based on Clustering Ensemble was developed to solve the problem that could not predict customer credit dynamically as well as population drifts in customer credit scoring [17]. Firstly, the training set samples were clustered into multiple subareas, then the entire observation period was fractionized into several fractional periods. Finally, customer credit scoring sub classifiers were established using cost-sensitive support vector machine. The empirical results showed that the dynamic model that they proposed not only had lower misclassification rate than static model, but also could predict the bad customers as early as possible.

Adams *et al.* [18] suggested a new streaming technology might be adapted to handle drift without an explicit drift model. A data stream comprises data sets that arrive at high frequency generated by a process that is subject to unknown changes, these changes are generally referred to as drift. Credit card transaction data is a very good example of a data stream. The nature of streaming data requires algorithms that are efficient and also adaptive in order for them to handle the high frequency and the drift also. A number of different approaches for streaming classifications have been proposed and these are data selection approaches with standard classifications and ensemble methods.

Adams *et al.* [18] mainly concentrated on modifying standard classifiers to incorporate forgetting factors, which are parameters that control the contribution of old data to parameter estimation. They used the main concept from the adaptive filter theory, so that the forgetting factor is automatically turned on. They concluded that adaptive forgetting methods have some advantage of reducing performance degradation between classifier rebuilds.

Nikolaidis *et al.* [19] concentrated on how population drift affects behavioural scoring. They gave examples of different data sources that can be used to start behavioural scoring. The examples that they gave are, delinquency history, usage history, static information such as age and demographic data, payment history, collections activity, type of credit and bureau data. The data set that they used was from Greece. The country's economic situation had been affected since 2009. Data used in behavioural scoring is dynamic in nature because it evolves over time yet most scoring models are static in nature hence there is that need to study the relationship between the two. The measure that they used in their experiment is the probability of default (PD) where their default definition was either good or bad using a performance window of 12 to 24 months and an outcome window of 12 months.

An entity that was being awarded the good status is the one that had 0–1 months in arrears, 2–3 months in arrears was indeterminate and 4 or more months in arrears was classified as bad. They concluded that behavioural scoring is a little bit affected by population drift and in situations where the population is good, it becomes better but when the population is bad, it becomes worse.

Whittaker *et al.* [20] used the Kalman filtering algorithm to come up with a technique for monitoring the performance of a client credit scorecard over a period of time. In their new approach, they allowed systematic updates on the scorecard after comparing the new applicant information to the previous ones. The Kalman filtering algorithm in this instance was used to cater for scorecard degradations, which can occur due to different factors which include but not limited to population drifts and changes in the economic conditions.

They adapted the Kalman filter so that it can be used as a diagnostic tool to monitor credit scorecards. Whenever there is a new applicant, the filter adaptively estimates a value so that current observations are given higher weight. The updated scorecard can be observed over time and suggestions can be made if there is need to rebuild the scorecard. The data that they used was that of a commercial company's mortgage portfolio which contained records on about 180 000 applicants and used about 30 variables from the application form. After the implementation of the dynamic scorecard, there was clear evidence that it properly and accurately scored new clients as compared to its baseline counterparts.

Pavlidis *et al.* [2] developed an adaptive and sequential approach for logistic regression, which caters for any type of population drift storing any previously captured data. Their main idea was to incorporate new data upon arrival into the system.

They defined a weighted likelihood function that regularly removes the effects of past data on current parameter estimates. The previously used information is forgotten and this produces an online algorithm that can cater for the changes in the population. The proposed method was evaluated on artificial data sets that show gradual and abrupt drift changes. The proposed method was also done on static artificial data sets. The performance of the proposed method, Adaptive Online Logistic Regression (AOLR) proved that it works better than the general logistic classifiers. The results proved that, updating the classifier as data becomes available gives an improvement as opposed to just re-estimating the classifier parameters.

There is an aspect related to population drift that Pavlidis *et al.* [2] did not consider, which is the delay between the observation of the predictor variables and the corresponding class label. This aspect can be generally solved by a survival analysis model, which changes over time as more data becomes available. A combination of the adaptive online logistic regression and the survival analysis model would make a great enhancement in solving the population drift problem.

Romanyuk [21] developed a credit scoring model based on contour subspaces. In his approach, he suggested the separation of a client's personal data and credit terms so that there is the creation of a contour subspace for credit scoring. Banks must take cognisant of the fact that the creditworthiness of a borrowing entity can change and must allow a platform to adjust the credit terms so that credit risk is effectively managed. Proper adjustments to the changes that occur can improve on the number of applicants that are more credit worthy. In his proposed model, credit terms form a contour subspace for each creditworthiness value to such an extent that when this model is being used, it must give assistance in making a decision on whether to grant a loan or not.

The decision can be reached through the visualization of a contour subspace of credit terms for an applicant in relation with an individual creditworthiness, giving choices on credit terms from this contour subspace and manage credit terms online in association with the changes in creditworthiness valuation. In general credit scoring, the applicants' personal data is combined with credit attributes in order to decide whether to grant a loan or not. Romanyuk's [21] model distinguishes the applicant's creditworthiness from the riskiness of the credit terms through the use of contour subspaces. This approach makes future decisions become less hectic because the bank does not need to go through the complete recalculation process when the same applicant applies for a different credit type. Whenever a client's credit worthiness changes, there should be a change also in the loan rate. Assuming that there is an increase in creditworthiness, the bank must respond by decreasing the loan rate, if creditworthiness remains constant then the bank must maintain the same loan rate, if creditworthiness decreases then the bank should ideally increase the loan rate. The introduction of contour spaces enables the bank to manage the dynamism in credit worthiness.

Babu and Satish [22] came up with a model based on the application of K-Nearest Neighbor (K-NN). They used the K-NN method to perform an approximating of good or bad risk likelihoods for an applicant. In order for one to be classified good and bad, it is based on the k most similar points in the training samples. The similarity of points accessed by the appropriate distance metric. They gave the advantages of using K-NN credit scoring as the automatic updating of the design set, providing a reason for refusal of credit by exploiting the information about class separation in the data provided by the regression weights and also it is less prone to declining credit on the basis of one characteristic regardless of all other attributes as compared to its linear or logistic regression counterparts.

Barakat [23] proposed the inclusion of context awareness in an effort to adapt classification models to different changes that might occur. Most changes occur to the data due to changes in the environment. There is need to study, understand and analyse the contextual issues in order to unearth the causes of the drift. Information regarding contextual issues helps in deciding on the relevant data to train and also to detect the changes that would have occurred in the model.

Identifying the appropriate training data size is one approach that can be used to manage drifts but this can bring a dilemma between stability and plasticity. A smaller window will easily capture the changes in the data hence detecting the drift easily but might be too small to bring stability in the concept description of the observed data. Barakat [23] proposed the incorporation of a context learning model that makes use of historical data to pick on the variables that have an effect on the concept of interest. The proposal suggested the use of drift detection mechanisms. It is more advantageous to implement drift detection algorithms since there will be no need to revisit the model or data that existed before the detection. The learner is designed in such a way that it has multiple levels and each model will be representing a different concept. Once the same concept resurfaces, the previously learned concept is the one that is run hence there is no need to continue giving reference to the previously trained data. Understanding the concept drift and handling it helps in maintaining the prediction model accuracy by adapting to the changes that would have occurred.

Bifet *et al.* [24] presented on handling concept drift and highlighted that, the fact that data continues to evolve over time is the reason why models suffer changes in data distribution normally referred to as concept drift. When a learning model is built, there is need to make sure that the model adapts to the changing environment in order to maintain model prediction accuracy. The adaptation of the model has to occur in a timeous and accurate manner. For a system to be able to handle concept drift, it has to be possessing some desirable properties like prompt detection and adaptation to the drift, be able to differentiate between drift and noise because in machine learning, noise can also affect the prediction accuracy. The adaptation must occur in such a way that it does not have much effect on the system resources like time and memory. Different adaptive learning strategies can be used in order to manage concept drift. The strategies can be grouped as single classifiers or ensembles. Single classifiers can make use of detectors, which can detect the changes they can implement the forgetting approach in which the old data is forgotten and retraining occurs using new data. In the ensemble approach, the strategy can be contextual based on or it can be in the form of a dynamic ensemble. In the contextual approach, many models are switched according to the observed incoming data. In dynamic ensemble, many models are built and the models are dynamically combined.

The best approach to use in different circumstances depends on the nature of change that would have occurred to the data. Changes can be categorized in terms of speed that it can be a gradual change or a sudden change. The other category is in terms of occurrence, whether it is always a new change or a reoccurrence. Sudden changes are best solved using single classifiers either by detection of by forgetting in which fixed windows are used with instance weighting. Gradual drift can be solved using dynamic ensembles, which have the implementation of adaptive fission rules. Reoccurring drift is best solved using contextual ensembles.

Drifts can occur due to different circumstances hence the sources of change can vary from adversarial in which there is an input to the learning model that is sent intentionally to cause an error in the model. Changes can also be caused by a change in interest, change in population or even a change in model complexity. These changes can be predictable, unpredictable or identifiable hence there is that need to keep the model up to date, detect the change whenever it happens and explain the change in order to deal with it in the case of reoccurrence.

Žliobaitė *et al.* [1] in their overview of concept drift applications did an analysis of the major applications that are usually affected by concept drift. They defined concept drift as the change that occurs in personal interests, population or adversary activities. The change is given the term concept drift in machine learning, data mining and predictive analysis but it has a different names in other area of study. When the change occurs in pattern recognition, it is referred to as covariate shift or data set shift, when it occurs in signal processing, it is referred to as non-stationarity. There is a difference in data source between the training data and the application data. Usually there is a wrong assumption when it comes to model building, test data source is different from train data source and this causes a deviation in the model accuracy for as long as data is distributed in streams rather than in static form evolution will always occur. There is no single solution that can solve the concept drift problem because it occurs in different scenarios.

Žliobaitė *et al.* [1] identified the main categories that are in use which suffer concept drift. The major application areas were identified as monitoring and control applications which mainly concentrate on detection and taking corrective action, information management which mainly deals with recommender systems [35] and personalized learning and also analytics and diagnostics, which deals with predictive analytics like evaluation of credit worthiness.

Garcia *et al.* [25] suggested an approach to handling concept drift in situations where there is both abrupt and slow gradual changes. They named their proposed solution as the Early Drift Detection Method (EDDM). The solution was based on taking into consideration the distance between two erroneous classifications instead of just concentrating on the number of errors as suggested by other researchers that designed the general detection methods. Their experimental results showed that it was also worth considering the distance between the errors in order to detect drift occurrence earlier, however, their work is not enough to use to model a dynamic credit scoring model because of the existing technological change which introduces a lot of dynamism.

Klinkenberg and Joachims [26] proposed a method of handling concept drift in Support Vector Machines (SVM). The proposal was based on keeping a window on the training data. Once a change has occurred, the window size must automatically adjust in order to reduce the estimated generalization error. The major strength of this proposal is on giving the correct window size per given scenario. SVM is based on structural risk minimization principle hence it basically learns linear decision rules explained by weight and a vector. In many common machine learning problems, researchers use a fixed window size but this approach's major drawback is that it makes a lot of assumptions on the rate of concept drift. A fixed window would work perfectly fine if concept drift will never occur or occurs after a very long period of time.

Klinkenberg and Joachims [26] proved that there is a direct relationship between the training window size and the rate at which a model responds to concept drift hence they suggested an algorithm that automatically adjusts the training window to suit the change. The choice of a training window must be done with caution because it must not be too big or too small. A training window must not be too big, otherwise it will keep including some old data hence giving a false prediction. The window must also not be too small in order to train the model widely to improve accuracy. In their experimental set up, they used four data management approaches which are full memory, no memory, fixed size window and adaptive window size. In the full memory approach, there is no forgetting, all previously used examples are considered, the no memory approach bases its data on the most recently used batch, there is no reference to the older versions. Fixed window size uses an unchanging window size for three batches and the adaptive window size implements the proposed window adjustment algorithm.

For as long as there was no concept drift, the full memory and adaptive window size performed almost the same. When concept drift occurred, the other three non-adaptive methods underperformed hence the adaptive method is the only one that produced accurate results. For as long as there is the occurrence of concept drift, it is more advantageous to make the window adjustable.

Hofer and Krempl [27] supported the fact that scoring data suffers changes over time hence there is that need to predict and monitor these changes in order to achieve the highest level of prediction accuracy. The main goal is to establish when the change would have occurred and act accordingly. Change can be revealed through change detection algorithms or change diagnosis which can be based on spatio-temporal density estimation, velocity density estimation or post analysis of changes in the distribution. There is need for an appreciation of the general distribution pattern so that when changes occur, it will be easily detected. Hofer and Krempl [27] suggested the use of un-labelled data to detect class priori changes. They also suggested the use of temporal goodness of fit and also the implementation of a controlled adaptation process. All these approaches help in monitoring and controlling drift in order to maintain prediction accuracy.

Tsymbal [28] explained the subject of concept drift in detail. When there is a change in the hidden context, it can cause a big change in the target concept. Models need to be set correctly in order to be able to distinguish between noise and concept drift. If an algorithm over reacts to noise, it might regard it as concept drift and also making it less reactive, it might react too slowly to concept drift when it occurs and it will not be efficient. Many concept drift handling approaches are based on updating the model when new data arrives but this approach is very costly because data can arrive at very fast unmanageable rates. In some systems, the data needs to be labelled based on user feedback and this can be very time consuming. The best approach to handling drift is to detect and adapt to the change. Tsymbal [28] also highlighted the major approaches to concept drift handling as instance selection, instance weighting and ensemble learning.

Ang *et al.* [29] analysed the existence of drift in distributed computing environment in which peer machines learn from each other in order to classify different concepts. When a change has occurred in one peer, the other peer must be able to detect the change and react accordingly in order to learn the correct model and produce correct classification results. Ang *et al.* [29] developed an ensemble approach that they referred to as Predictive and Parameter Insensitive Ensemble (PINE). PINE is an approach which enables drift to be handled in two ways namely the reactive and the proactive approach. In the reactive approach, drift would have occurred and it is detected and corrective action is taken. In the proactive, upcoming events are assessed and warning signs and adaptation approaches are communicated across the network to other peers.

Ang *et al.* [29] did a comparison of the performance of the Ensemble of Proactive Models (PEM), the Ensemble of Reactive Models (REM) and PINE. They concluded that PINE being a combination of the proactive and the reactive approaches performed better in detecting and managing drift in situations where the drift was occurring at different times for different peers. This study was done for distributed computing but it is very relevant in detecting drift in data streams scenarios such as credit scoring hence the ensemble approach can be used to detect and manage drifts in such cases.

Zliobaite *et al.* [30] produced a framework that can be used for active learning on changing data streams. They produced this framework as further work to the initial work on Massive Online (MOA) system. Bifet *et al.* [32] under normal circumstance active learning strategies concentrate on queuing the most unlikely instances, which are usually found on the decision boundary. This implies that whenever a change occurs away from the decision boundary it is most likely unidentified hence there is no adaptation. The major goal in coming up with adaptive learning strategies is to maximize prediction accuracy as time lapses but at the same time managing resources such as labelling costs and time. The software framework that they produced is suitable for use in active learning classifications on data streams. The platform can be used by researchers to do experiments on data stream learning benchmarks for active learning. Credit scoring [31, 33, 34] is ever challenging research area with a number of researchers providing solutions.

8 Credit Scoring Models Application Areas

When used in Financial Institutions credit scoring models are used in loan applications to make credit decisions for loan applications, set credit limits, manage existing accounts and forecast the profitability of consumers and customers.

When used in the insurance industry, they are used to decide on the applications of new insurance policies and the renewal of existing polices. They can also be used in Real Estate where landlords can make use of credit scores to determine whether potential tenants are likely to pay their rent on time. In Human Resources, some employers make use of credit history and credit scores to decide whether to hire a potential employee, especially for posts where employees need to handle huge sums of money.

9 Conclusion

The design of a dynamic credit scoring model that caters for the drift problem is of paramount importance in improving the prediction and the rating accuracy. Most of the models that have been implemented do not consider the drift problem. Some models do consider drift and adapt to it by revisiting the model and making adjustments to it but this process has proven to be costly in terms of time and other resources hence designing a model that can detect a change and automatically adapt to it is recommended.

References

1. Žliobaitė, I., Pechenizkiy, M., Gama, J.: An overview of concept drift applications. In: Japkowicz, N., Stefanowski, J. (eds.) Big Data Analysis: New Algorithms for a New Society. SBD, vol. 16, pp. 91–114. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-26989-4_4
2. Pavlidis, N.G., Tasoulis, D.K., Adams, N.M., Hand, D.J.: Adaptive consumer credit classification. J. Oper. Res. Soc. **63**(12), 1645–1654 (2012)
3. Huang, S., Day, M.: A comparative study of data mining techniques for credit scoring in banking. In: IEEE IRI, San Francisco, California, USA, 14–16 August 2013 (2013)
4. Huang, Z., et al.: Credit rating analysis with support vector machines and neural networks: a market comparative study. Decis. Support Syst. **37**(4), 543–558 (2004)
5. Keramati, A., Yousefi, N.: A proposed classification of data mining techniques in credit scoring. In: Proceedings of the International Conference on Industrial Engineering and Operations Management (2011)
6. Paleologo, G., Elisseeff, A., Antonini, G.: Subagging for credit scoring models. J. Oper. Res. **201**(2), 490–499 (2010)
7. Li, X.L., Zhong, Y.: An overview of personal credit scoring: techniques and future work. Int. J. Intell. Sci. **2**(04), 181 (2012)
8. Luo, S.T., Cheng, B.W., Hsieh, C.H.: Prediction model building with clustering-launched classification and support vector machines in credit scoring. Expert Syst. Appl. **36**(4), 7562–7566 (2009)
9. Fogarty, D.J.: Using genetic algorithms for credit scoring system maintenance functions. Int. J. Artif. Intell. Appl. **3**(6), 1 (2012)
10. Hand, D.J., Adams, N.M.: Selection bias in credit scorecard evaluation. J. Oper. Res. Soc. **65** (3), 408–415 (2014)
11. Kelly, M.G., Hand, D.J., Adams, N.M.: The impact of changing populations on classifier performance. In: Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 367–371. ACM, August 1999
12. Kadwe, Y., Suryawanshi, V.: A review on concept drift. IOSR J. Comput. Eng. **17**, 20–26 (2015)
13. Knotek, J., Pereira, W.: Survey on Concept Drift. Faculty of Economics, University of Porto, Portugal. https://is.muni.cz/el/1433/podzim2011/PA164/um/drift_detection_methods.pdf
14. Sun, Y., Tang, K., Zhu, Z., Yao, X.: Concept drift adaptation by exploiting historical knowledge. IEEE Trans. Neural Netw. Learn. Syst. **29**, 4822–4832 (2018)
15. Zliobaite, I., et al.: Next challenges for adaptive learning systems. ACM SIGKDD Explorations Newsl **14**(1), 48–55 (2012)
16. Krempl, G., Hofer, V.: Classification in presence of drift and latency. In: 2011 IEEE 11th International Conference on Data Mining Workshops (ICDMW), pp. 596–603. IEEE, December 2011

17. Wei, G., Mingshu, C.: A new dynamic credit scoring model based on clustering ensemble. In: 3rd International Conference on Computer Science and Network Technology (2013)
18. Adams, N.M., Tasoulis, D.K., Anagnostopoulos, C., Hand, D.J.: Temporally-adaptive linear classification for handling population drift in credit scoring. In: Lechevallier, Y., Saporta, G. (eds.) Proceedings of COMPSTAT'2010, pp. 167–176. Physica-Verlag HD, Heidelberg (2010). https://doi.org/10.1007/978-3-7908-2604-3_15
19. Nikolaidis, D., Doumpos, M., Zopounidis, C.: Exploring population drift on consumer credit behavioral scoring. In: Grigoroudis, E., Doumpos, M. (eds.) Operational Research in Business and Economics. SPBE, pp. 145–165. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-33003-7_7
20. Whittaker, J., Whitehead, C., Somers, M.: A dynamic scorecard for monitoring baseline performance with application to tracking a mortgage portfolio. J. Oper. Res. Soc. **58**(7), 911–921 (2007)
21. Romanyuk, K.: A dynamic credit scoring model based on contour subspaces. In: Proceedings of SAI Intelligent Systems Conference (IntelliSys). IntelliSys (2016)
22. Babu, R., Satish, A.R.: Improved of k-nearest neighbor techniques in credit scoring. Int. J. Dev. Comput. Sci. Technol. **1**(2), 1–4 (2013)
23. Barakat, L., Pavlidis, N., Crone, S.: A context-aware approach for handling concept drift in classification (Doctoral dissertation, Lancaster University) (2018)
24. Bifet, A., Gama, J., Pechenizkiy, M., Liobait, I.: Handling concept drift: importance challenges & solutions. Tutorial. In: Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining (2011)
25. Baena-Garcia, M., del Campo-Ávila, J., Fidalgo, R., Bifet, A., Gavalda, R., Morales-Bueno, R.: Early drift detection method. In: Fourth International Workshop on Knowledge Discovery from Data Streams, vol. 6, pp. 77–86, September 2006
26. Klinkenberg, R., Joachims, T.: Detecting concept drift with support vector machines. In: ICML, pp. 487–494, June 2000
27. Hofer, V., Krempl, G.: Predicting and Monitoring Changes in Scoring Data (2015)
28. Tsymbal, A.: The problem of concept drift: definitions and related work, vol. 106, no. 2. Computer Science Department, Trinity College Dublin (2004)
29. Ang, H.H., Gopalkrishnan, V., Zliobaite, I., Pechenizkiy, M., Hoi, S.C.: Predictive handling of asynchronous concept drifts in distributed environments. IEEE Trans. Knowl. Data Eng. **25**(10), 2343–2355 (2013)
30. Zliobaite, I., Bifet, A., Holmes, G., Pfahringer, B.: MOA concept drift active learning strategies for streaming data. In: Proceedings of the Second Workshop on Applications of Pattern Analysis, pp. 48–55, October 2011
31. Bhatia, S., Sharma, P., Burman, R., Hazari, S., Hande, R.: Credit scoring using machine learning techniques. Int. J. Comput. Appl. **161**(11), 1 (2017)
32. Bifet, A., Holmes, G., Kirkby, R., Pfahringer, B.: MOA: massive online analysis. J. Mach. Learn. Res. **11**, 1601–1604 (2010)
33. Kogeda, O.P., Vumane, N.N.: A model augmenting credit risk management in the banking industry. Int. J. Technol. Diffus. (IJTD) **8**(4), 47–66 (2017)
34. Chikoore, R., Kogeda, O.P.: A credit rating model for Zimbabwe. In: Proceedings of the Southern Africa Telecommunication Networks and Applications Conference (SATNAC), Fancourt, George, Western Cape, South Africa, 4–7 September 2016, pp. 88–89 (2016)
35. Bhebe, W., Kogeda, O.P.: Shilling attack detection in collaborative recommender systems using a meta learning strategy. In: Proceedings of the IEEE International Conference on Emerging Trends in Networks and Computer Communications, ETNCC 2015, Windhoek Country Club Resort, Namibia, 17–20 May 2015, pp. 56–61 (2015)