Applying Ensemble-based Online Learning Techniques on Crime Forecasting

Anderson José de Souza, André Pinz Borges, Heitor Murilo Gomes, Jean Paul Barddal and Fabrício Enembreck
Programa de Pós-Graduação em Informática, Pontifícia Universidade Católica do Paraná, Av. Imaculada Conceição, 1155, Curitiba, Paraná, Brazil

Keywords: Data Stream Classification, Crime Forecasting, Public Safety, Concept Drift.

Abstract: Traditional prediction algorithms assume that the underlying concept is stationary, i.e., no changes are expected to happen during the deployment of an algorithm that would render it obsolete. Although, for many real world scenarios changes in the data distribution, namely concept drifts, are expected to occur due to variations in the hidden context, e.g., new government regulations, climatic changes, or adversary adaptation. In this paper, we analyze the problem of predicting the most susceptible types of victims of crimes occurred in a large city of Brazil. It is expected that criminals change their victims' types to counter police methods and vice-versa. Therefore, the challenge is to obtain a model capable of adapting rapidly to the current preferred criminal victims, such that police resources can be allocated accordingly. In this type of problem the most appropriate learning models are provided by data stream mining, since the learning algorithms from this domain assume that concept drifts may occur over time, and are ready to adapt to them. In this paper we apply ensemble-based data stream methods, since they provide good accuracy and the ability to adapt to concept drifts. Results show that the application of these ensemble-based algorithms (Leveraging Bagging, SFNClassifier, ADWIN Bagging and Online Bagging) reach feasible accuracy for this task.

1 INTRODUCTION

Data Stream Mining is an active research area that aims to extract knowledge from large amounts of continuously generated data, namely data streams. In addition to the conventional problems that impact conventional machine learning, Data Stream Mining algorithms must be able to perform both fast and incremental processing of arriving instances, addressing time and memory limitations without jeopardizing predictions' accuracy. One of the major characteristics of data streams is the assumption that data distribution changes over time, phenomenon known as concept drifts (Schlimmer and Granger, 1986; Widmer and Kubate, 1996). In order to deal with concept drifts, ensemble-based methods are widely used as they are able to achieve high accuracy, while adapting to both gradual and abrupt concept drifts (Barddal et al., 2014; Bifet et al., 2009). In these methods, a set of classifiers is trained, usually on different chunks of instances, and their predictions are combined in order to obtain a global prediction. Training classifiers in different chunks of instances cause their models to be different, i.e., a diverse set of classifiers is induced. Intuitively, a diverse set is better than a homogeneous set, since their combination may achieve higher accuracy than any of the individual classifiers. Even though this claim has not been theoretically proven (Kuncheva et al., 2003), many works back this argument through empirical tests (Bifet et al., 2010b; Gomes and Enembreck, 2013, 2014; Barddal et al., 2014). Also, for data stream mining ensemble classifiers provide another benefit, which is gradual adaptations to the model by resetting individual classifiers in response to changes in data, instead of the whole model, which is often the case when a single classifier is used. This trait causes the ensemble to be less susceptible to noise (false concept drifts) and to adapt naturally to gradual concept drifts.

Data Stream Mining has been successfully applied in many real world problems, such as ATM and credit card operations fraud detection (Domingos and Hulten, 2000), mass flow (Pechenizkiy et al., 2010), wind power prediction for smart grids (Bessa et al., 2009), crime prediction by region (Gama et al., 2014) and fraud crime detection (Zliobaite, 2010).

de Souza A., Pinz Borges A., Gomes H., Barddal J. and Enembreck F..
Applying Ensemble-based Online Learning Techniques on Crime Forecasting.
DOI: 10.5220/0005335700170024
In *Proceedings of the 17th International Conference on Enterprise Information Systems* (ICEIS-2015), pages 17-24
ISBN: 978-989-758-096-3

In this paper we present a real application of data stream ensemble-based algorithms for crime forecasting. We analyze a dataset obtained from the emergency call center of Brazilian Military Police (PMSC), namely CRE190, which acts 24/7/365. CRE190 receives approximately 2,000 calls a day, of which 300 are converted to police occurrences. Police occurrences are classified into 9 subgroups: Community Aid, Crimes and Contraventions, Diverse Occurrences, Emergencies, Environmental Crimes, Maintenance Activities and Transit Occurrences. Therefore, in order to provide more intelligent support for PMSC, we study online learning techniques to determine which type of target (private individual entities, commercial stores or residencies) is more susceptible to crimes and contraventions in a short-term future. We aim at pointing out the most appropriate ensemble-learning algorithm for crime forecasting to support PMSC to prevent crimes by applying statistical hypothesis testing to corroborate results obtained. This paper is divided as follows. Section 2 presents the problem definition and briefly described the dataset used while Section 3 presents ensemble-based methods applied in this work. Section 4 presents the experimental protocol and discusses about empirical results obtained. Finally, Section 5 concludes this paper and presents future work.

2 THE BRAZILIAN CRIME FORECASTING PROBLEM

There are few academic works relating the usage of knowledge extraction from Brazilian Military Police public safety databases. We believe that the main reason for this fact is the lack of data mining experts inside these entities. Also, we must emphasize the difficulties on retrieving this kind of data, which access' is usually restricted to inside personnel, therefore, not much academic research is allowed. Most part of the existing studies are limited to Brasil's Civil Police from many different parts of the country, which do not abjure Military Police reality of the same regions. This occurs since there is no integrated system which merges data obtained from Federal, Civil and Military polices. It happens that when a victim makes an emergency call to CRE190 and opens a new occurrence, this data is not shared to Civil and Federal polices, and vice-versa.

In Azevedo et al. (2011), authors emphasize the lack of data acquisition and storage by Military Polices, which do not allow planning using statistics or more sophisticated techniques, such as forecasting. Also, intelligent planning is accounted by authors in Machado (2009), which enlightened the need of a historical database of criminal actions to perform strategical planning of police actions to prevent crimes.

Some geographic-based approaches are also relevant to determine the most susceptible areas for crimes in cities (Ferreira, 2012; Nath, 2006). Approaches such as Formal Concept Analysis (FCA) have also been applied in Siklóssy and Ayel (1997) to map and structure specific criminal organizations. These organizations are defined by its constituents, *modus operandi*, resources and types of crime (e.g. robbery, drug trade, homicide, art and luxury auto theft). Also, FCA and rule-based classification presented interesting results on domestic violence in Amsterdam, Netherlands (Poelmans et al., 2011).

Time series models were also applied for crime forecasting in Wang et al. (2013) in order to determine their order and whether they would be performed by the same individuals.

Focusing on Brazilian's Military Police, most part of its operations are planned accordingly to empirical knowledge which resides on public safety personnel. Also, an interesting fact occurs: it is common that a very low occurrence rate for a certain type of victim might occur during a period of time, followed by a sudden increase in these rates and vice-versa. The cause of these sudden increases and decreases cannot be determined due to the limited aspect of the data acquired and stored by the military police. Nevertheless, it is feasible to assume that these crime rates change accordingly to the hidden context (Tsymbal, 2004), i.e., these changes may be due to attributes not being monitored. Instead of focusing on finding why these changes happens, it is easier to adapt the forecasting model to it. Concept drifts might occur due to a variety of factors such as: period of the month, holiday eves, payment days and economic inflations which might characterize a higher cash flow in certain time span. In this sense, we hypothesize that concept drift learning techniques are needed to forecast crimes in online fashion, thus, presenting up to date predictions for the most susceptible types of crimes in the near future.

2.1 The Santa Catarina's Military Police Dataset

The dataset extracted from PMSC's Crime and Contraventions Group contains three types of police occurrences bulletins regarding robbery and assaults against private individual entities, residencies and commercial stores. These three types were chosen due to its numerous appearances all over the country; thus, diminishing these crimes are highly related to public safety goals. All extracted data is originally generated at the emergency call center, namely CRE190, where the call attendee fills a form regarding the caller's emergency, which is later used to notify policemen. The next step occurs when policemen insert extra data about the emergency, such as people involved and procedure used. Until now, no analytical process has ever been made over this data in order to provide PMSC extra information for decision-making.

The data extracted contains 210 days (instances) which regards occurrences from January 1^{st} 2010 to July 30^{th} 2010 and are divided in the following attributes: day of the month, holiday eve, holiday, dawn, morning, afternoon, night, female amount, male amount, part of the week and the most susceptible type of occurrence, i.e. private individual entities, commercial stores or residencies (class). This dataset accounts for an average of 2.36 occurrences per day. Table 1 presents these attributes' descriptions and domains.

In this paper we focus on the task of predicting the most susceptible type of victim: private individual entities, commercial stores or residences. Crimes against private individual entities usually happen when people are arriving home, specially victimizing women while waiting to front gates open. Additionally, this type of crime is also very common against women that visit banks, since they usually do not react to crimes, therefore being easy targets.

Crimes against commercial stores, in majority, involve markets, bakeries, pharmacies and clothing stores. In all of these situations, most part of cashiers are women, therefore, are also very susceptible to crimes. Thus, a possibility would be to exchange cashiers during specific times of the day to reduce its own chances to be a target.

Finally, crimes against residences occur during the morning period or early evening. This type of crime occur when a residence is left alone or as a continuation of crimes against private individuals that are arriving home as presented early.

Based on the predictions obtained, we expect to allow policemen to determine efficient routes accordingly to the most susceptible types of victims, therefore patrolling residential, commercial or banking areas more efficiently.

3 ENSEMBLE-BASED ALGORITHMS

Most of the existing work on ensemble classifiers relies on developing algorithms to improve overall classification accuracy that copes with concept drift explicitly (Bifet et al., 2009) or implicitly (Kolter and Maloof, 2007, 2005; Widmer and Kubate, 1996). An ensemble classifier can surpass an individual classifier's accuracy if its component classifiers are diverse and achieve a classification error below 50% (Kuncheva, 2004). Thus, an ensemble is said diverse if its members misclassify different instances. Another important trait of an ensemble refers to how it combines individual decisions. If the combination strategy fails to highlight correct and obfuscate incorrect decisions then the method menaced. In addition, when dealing with data streams it is important to consider concept drifts and hence include some kind of adaptation strategy into the ensemble. Therefore, a successful ensemble classifier for data streams must induce a diverse set of component classifiers, including a combination strategy that amplifies correct predictions and adapts to concept drifts.

In the next sections we survey the utilized ensemble-based algorithms for this comparison.

3.1 Additive Expert Ensemble

The Additive Expert Ensemble (AddExp) (Kolter and Maloof, 2005) is an extension to the DWM algorithm (Kolter and Maloof, 2007), thus AddExp associates each classifier a weight ω which is decreased by an user-given factor β every time it misclassifies an instance. The ensemble prediction corresponds to the class value with the greatest weight after combining classifiers' weights. If the overall prediction is incorrect a new classifier is added to the ensemble with weight τ equal to the total weight of the ensemble times some user-given constant γ. Since this strategy can yield a large number of experts, authors consider oldest first and weakest first pruning methods to address efficiency issues.

3.2 Online Bagging

The Online Bagging algorithm is a variation for data streams of the traditional ensemble classifier Bagging (Oza and Russell, 2001). Originally, a bagging ensemble is composed of k classifiers, which are trained with subsets (bootstraps) D_j of the whole training set D. However, sampling usually is not feasible in a data stream configuration, since that would require storing all instances before creating subsets. Also, the probability of each instance to be selected for a given subset is governed by a Binomial distribution and therefore can be approximate by evaluating every instance one at a time and randomly choosing if the instance will be included in a given subset (Oza and Russell,

Table 1: Attributes extracted from the PMSC's database.

Attribute	Type	Description
Day of the Month	Numeric	Day of the month represented by the instance
Holiday Eve	Binary	Determines whether this instance represents a holiday eve
Holiday	Binary	Determines whether this instance represents a holiday
Dawn	Numeric	Amount of occurrences during the 00:01 until 05:59 period
Morning	Numeric	Amount of occurrences during the 06:00 until 11:59 period
Afternoon	Numeric	Amount of occurrences during the 12:00 until 17:59 period
Night	Numeric	Amount of occurrences during the 18:00 until 00:00 period
Female Amount	Numeric	Amount of female victims
Male Amount	Numeric	Amount of male victims
Part of the Week	{Beginning, Middle, End}	Monday and Tuesday stand for Beginning, Wednesday and Thursday stand for Middle and Friday, Saturday and Sunday stand for End
Most Susceptible Type of Occurrence (class)	{Private Individual Entity, Commercial Store, Residence}	Determines who is the most susceptible target for crimes

2001). Nevertheless, this solution is not applicable as it is necessary to know beforehand the size N of the training set for an accurate approximation of the original sampling process. However, if we assume that $N \to \infty$, i.e. an infinite stream, then the binomial distribution tends to a Poisson distribution with $\lambda = 1$, thus it is feasible to approximate the sampling process (Oza and Russell, 2001).

3.3 Adaptive Windowing Bagging

Adaptive Windowing (ADWIN) Bagging extends Online Bagging, presenting ADWIN as a change detector (Bifet et al., 2009). ADWIN keeps a variable-length window of recently seen instances, with the property that the window has the maximal length statistically consistent with the hypothesis "there has been no change in the average output value inside the window". The ADWIN change detector is both parameter and assumption-free in sense that is automatically detects and adapts to the current rate of change. Its only parameter is a confidence threshold δ that indicates the confidence level we want to be in the algorithm's output since base classifiers outputs are inherently random. Therefore, ADWIN Bagging is the Online Bagging method presented in Oza and Russell (2001) with the addition of the ADWIN algorithm as a drift detector. When a drift is detected, the worst classifier of the ensemble of classifiers is removed and a new classifier is added trained with instances chosen by the Online Bagging method.

3.4 Leveraging Bagging

Leveraging Bagging uses the ADWIN algorithm to detect drifts, but enhances the training method by proposing two improvements (Bifet et al., 2010b). First, the bagging performance is leveraged by increasing the resampling and using output detection codes. The resampling process is done by using the Online Bagging method using a Poisson distribution. In this sense, the weights of the resamples are increased by increasing the variable λ. Second, a randomization is added at the output of the ensemble using output codes. To each class possible to predict, a binary string of length n is assigned and then an ensemble of n classifiers is created. Each of the classifiers learns one bit for each position in this binary string. When a new instance arrives, it is assigned x to the class whose binary code is closest. Random output codes are used instead of deterministic codes. In standard ensemble methods, all classifiers try to predict the same function. Nevertheless, by using random output codes, each classifier will predict a different function, therefore increasing the diversity of the ensemble.

3.5 Online Updated Accuracy Ensemble

The Online Accuracy Updated Ensemble (OAUE) maintains a weighted set of k component classifiers, such that the weighting is given by an adaptation to incremental learning of the weight function presented in Brzezinski and Stefanowski (2013). Every w instances, the least accurate classifier is replaced by a candidate classifier, which has been trained only in the last w instances. Since there is no drift detection algorithm, OAUE relies on these periodic reconstructs of the ensemble to adapt to concept drifts.

3.6 Social Adaptive Ensemble 2

SAE2 (Gomes and Enembreck, 2014) is an improved version of the original Social Adaptive Ensemble algorithm (Gomes and Enembreck, 2013). Both algorithms arrange the ensemble members in a weighted graph structure, where nodes are classifiers and connections between them are weighted according to their “similarity”. Similarity is measured as a function of their recent individual predictions over the same instances. This graph structure is used to combine similar classifiers into groups (Maximal Cliques). These groups are used during vote, such that individual predictions are combined into each group decision, which are then combined to obtain the overall prediction. The reason to combine predictions this way is to prevent a large set of similar classifiers from dominating the overall prediction. SAE2 uses a fixed window w to control how often updates are made to the ensemble structure, such as classifiers removals and additions accordingly to maximum and minimum similarity parameters s_{max} and s_{min}; the maximum ensemble size k_{max} and the minimum accuracy for each classifier e_{min}. Also, similarly to OAUE, SAE2 includes a background classifier trained during the last window, which allow the ensemble to quickly recover from concept drifts.

3.7 Scale-free Network Classifier

The Scale-free Network Classifier (SFNC) (Barddal et al., 2014) weights classifiers predictions based on an adaptation of the scale-free network construction model (Albert and Barabási, 2002). In SFNC, classifiers are arranged in a graph structure, similarly to SAE and SAE2, such that classifiers with higher accuracy are more likely to receive connections. During prediction, classifier weighting is based on a centrality metric α, such as degree or eigenvector, instead of solely the individual accuracy. Since, high accuracy classifiers tend to receive more connections, depending on the centrality measure used, there is an indirect relation between accuracy and the centrality measure. A new classifier is added every w instances if the ensemble overall accuracy is below a given user threshold θ and the ensemble size is below a parameter k_{max}, and it may receive connections from other classifiers according to their individual accuracy. In addition to that, the classifier with the lowest accuracy is removed every w instances, and that triggers a rewiring process to maintain the graph connected.

Table 2: Algorithms evaluated and its parameters.

Algorithm	Parameter	Value
AddExp	β	0.5
	γ	0.1
	τ	0.05
	Pruning Method	Weakest First
Online Bagging	k	10
ADWIN Bagging	k	10
	δ	10%
Leveraging Bagging	k	10
	δ	10%
	k	6
OUAE	k	10
	w	$\{3,7,10\}$
SAE2	k_{max}	10
	w	$\{3,7,10\}$
	e_{min}	66%
	s_{min}	89%
	s_{max}	99%
SFNC	k_{max}	3
	w	{3,7,10}
	α	Eigenvector
	θ	95%

4 EMPIRICAL EVALUATION

There is a great monitoring necessity in regions that try to maintain order and public safety. Nowadays, reports are manually produced by PMSC every 6 hours in order to provide both feedback for managers and to inform policemen which are about to start patrolling. In this section we evaluate a variety of ensemble-based algorithms in order to determine which present feasible accuracy to perform real-time crime forecasting. Firstly we present the experimental protocol and later discuss the results obtained.

4.1 Experimental Protocol

Since the dataset represents daily instances, we evaluated all algorithms using 3, 7 and 10-day window sizes, although a 6 hour period would be essential. The latter window size of 3 days allows us to analyze police occurrences that happened during the weekend, which are currently treated by the next Monday. The window size of 7 days allows us to analyze whether policemen work was effective when compared to an early week. Finally, a 10-day window size is a special case which also helps managers to verify the effectiveness of police operations.

In order to compare algorithms, we used all of the algorithms presented in Section 3 with the default values presented at original papers, with the exception of internal window sizes w (when existing), which were set accordingly to the evaluation ones (3, 7 and 10). In addition, we compare the results for a single Updatable Naïve Bayes classifier and an incremental ZeroR classifier to determine the lowest bound. All algorithms were implemented under the Massive Online Analysis framework (Bifet et al., 2010a). Also, all ensembles use an Updatable Naïve Bayes base learner (John and Langley, 1995). Table 2 ensemble-based algorithms' parameter values used in this evaluation.

To obtain accuracy for algorithms, we adopted the Prequential test-then-train procedure (Gama and Rodrigues, 2009) due the monitoring of the evolution of performance of models over time although it may be pessimistic in comparison to the holdout estimative. Nevertheless, authors in Gama and Rodrigues (2009) observe that the Prequential error converges to an periodic holdout estimative (Bifet et al., 2010a) when estimated over a sliding window.

The Prequential error of a classifier is computed, at a timestamp i, over a sliding window of size w' accordingly to Equation 1 where $L(\cdot,\cdot)$ is a loss function for the obtained class value y_k and the expected $\hat{y}_k$.

$$P_{w'}(i) = \frac{1}{w'} \sum_{k=i-w'+1}^{i} L(y_k, \hat{y}_k) \quad (1)$$

Finally, the accuracy of each classifier on a timestamp i is computed as $1 - P_{w'}(i)$.

In this comparison we determined evaluations over windows of 3, 7 and 10 days (instances) for our experiments since these values are used to develop structural police operations as discussed earlier.

4.2 Results Obtained

In Table 3 we present the accuracy results obtained, where one can see that Leveraging Bagging outperforms all other algorithms in all different evaluation window sizes. This occurs mainly due the leveraged resampling process, which is responsible for indirectly boosting accuracy (Bifet et al., 2009). Additionally, results show that both OUAE and SAE2 do not perform well when submitted to the crime-forecasting problem while algorithms ADWIN Bagging, Leveraging Bagging and SFNC presented the best overall average results.

Between the best ranked algorithms, one can see that Leveraging Bagging presents the best results for all window sizes (3, 7 and 10). In order to determine whether there is significant statistical difference between Leveraging Bagging and other algorithms, we used a combination of Friedman and Bonferroni-Dunn's non-parametric tests. Firstly, Friedman test pointed out that there is statistical difference between algorithms using $\alpha = 0.05$. Therefore, we used Bonferroni-Dunn's $1 \times N$ test pivoting Leveraging Bagging and determined that it is significantly superior to other algorithms with the exception of ADWIN Bagging, Online Bagging and SFNC when compared with a 95% confidence level.

In Figure 1 one can see the accuracy of algorithms during the whole stream, where Leveraging Bagging is able to surpass all others in a 3-day window. Also, one can see that algorithms do not perform well during the first 50 instances. This occurs mainly due the lack of instances seen by the classifiers and the fact that the analyzed data refers to a coastal region, therefore, from January until March, most part of the population is absent and victims realize that the crime occurred only after this period.

We also emphasize the good adaptation of both SFNC and Online Bagging algorithms, outperforming others during the stream. In Figure 1 we present results for a single Naïve Bayes and an Incremental ZeroR classifier, in order to determine the lower bound for accuracy, therefore, confirming that the Bagging variations and SFNC yield remarking results.

Although most part of algorithms presents a 60% accuracy during the stream, we believe these results are interesting for a generic and naïve approach based on a limited amount of features and instances.

4.3 Discussion

Generally, the algorithms used in this benchmark presented a reasonable behavior when submitted to the crime forecasting problem. All the analytics here presented was performed aiming at reproducing the Military Police procedures for strategical planning. Online learning algorithms with the capability of detecting and adapting to concept drifts are essential for the crime forecasting task since it empowers decision-making in order to prevent crimes with diminished resource usage and best manpower allocation. Results here presented contribute positively for Public Safety, which so far did not explicitly used online learning techniques to prevent crimes in specific city regions. We must emphasize that online learning algorithms can be used by any Public Safety related agency that pursuits intelligent automatic support for decision making.

Another important trait is the lack of professionals with adequate knowledge to apply the techniques here presented in Military Police. In order to take advantage of the techniques here presented and other

Table 3: Accuracy Obtained for the Tested Algorithms.

	Average Accuracy (%)								
Window Size	Naïve Bayes	Incremental ZeroR	AddExp	Online Bagging	ADWIN Bagging	Leveraging Bagging	OUAE	SAE2	SFNC
3	50.86	39.81	61.08	64.60	64.60	**67.21**	57.36	54.82	65.12
7	50.33	39.54	60.49	63.85	61.10	**64.29**	62.93	54.25	63.85
10	49.77	38.89	60.26	63.52	59.95	**63.83**	61.28	43.02	63.72
Average Ranking	*7.67*	*9.00*	*5.67*	*3.50*	*3.17*	***1.00***	*5.33*	*7.33*	*2.33*

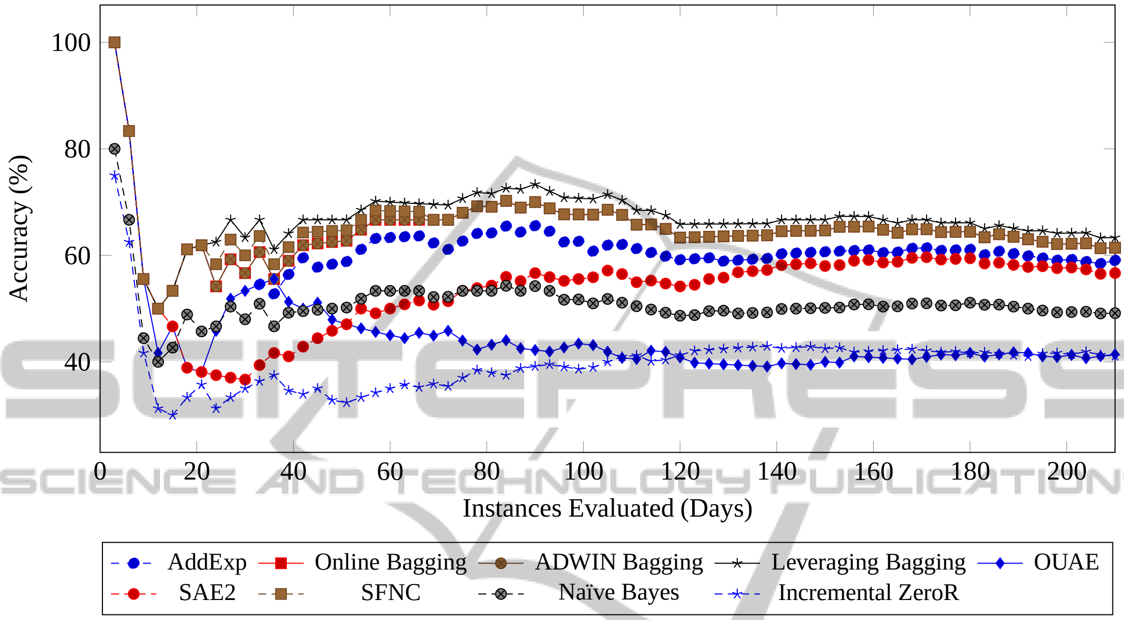

Figure 1: Accuracy obtained using a 3-day window evaluation.

data mining algorithms, a great effort must be put on to break bureaucratic paradigms so that Public Safety endorses and applies machine learning in day to day occasions, therefore enabling future research and the development or optimization of algorithms for this specific task.

5 CONCLUSION

In this paper we analyze and compare seven ensemble-based algorithms for the crime-forecasting problem. This approach was chosen due the possibility to perform forecasting incrementally accordingly to the arrival of occurrences, i.e. CRE190. Specifically at CRE190, estimates show that 93,000 police occurrences annually at 24/7/365.

Assuming the great need of Public Safety organizations regarding tools for crime forecasting and prevention, we believe this kind of technique is helpful and helps these organizations on establishing specific plans to act against crime. We also assume that an initial 60+% accuracy is feasible and tends to improve on its effective implementation, where police occurrences (instances) would arrive intermittently, therefore, allowing intelligent actions by crime prevention teams. So far we were unable to apply our approach in a real-time application, therefore we were incapable of measuring the effect size on crime prevention.

One of the major limitations of this work is the dataset size. Therefore, in future work, we look forward to continue this research by analyzing larger datasets with an extended set of features, other crime types and criminal data from other regions.

REFERENCES

Albert, R. and Barabási, A. L. (2002). Statistical mechanics of complex networks. In *Reviews of Modern Physics*, pages 139–148. The American Physical Society.

Azevedo, A. L. V. d., Riccio, V., and Ruediger, M. A. (2011). A utilização das estatísticas criminais no planejamento da ação policial: Cultura e contexto organizacional como elementos centrais. 40:9 – 21.

Barddal, J. P., Gomes, H. M., and Enembreck, F. (2014). Sfnclassifier: A scale-free social network method to handle concept drift. In *Proceedings of the 29th Annual ACM Symposium on Applied Computing (SAC)*, SAC 2014. ACM.

Bessa, R., Miranda, V., and Gama, J. (2009). Entropy and correntropy against minimum square error in offline and online three-day ahead wind power forecasting. *Power Systems, IEEE Transactions on*, 24(4):1657–1666.

Bifet, A., Holmes, G., Kirkby, R., and Pfahringer, B. (2010a). Moa: Massive online analysis. *The Journal of Machine Learning Research*, 11:1601–1604.

Bifet, A., Holmes, G., and Pfahringer, B. (2010b). Leveraging bagging for evolving data streams. In *Machine Learning and Knowledge Discovery in Databases*, pages 135–150. ECML PKDD.

Bifet, A., Holmes, G., Pfahringer, B., Kirkby, R., and Gavaldà, R. (2009). New ensemble methods for evolving data streams. In *Proc. of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 139–148. ACM SIGKDD.

Brzezinski, D. and Stefanowski, J. (2013). Reacting to different types of concept drift: The accuracy updated ensemble algorithm. *IEEE Transactions on Neural Networks and Learning Systems*, 25(1):81–94.

Domingos, P. and Hulten, G. (2000). Mining high-speed data streams. In *Proc. of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 71–80. ACM SIGKDD.

Ferreira, J., J. P. M. J. (2012). Gis for crime analysis: Geography for predictive models. 15:36–49.

Gama, J. and Rodrigues, P. (2009). Issues in evaluation of stream learning algorithms. In *Proc. of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 329–338. ACM SIGKDD.

Gama, J. a., Žliobaitė, I., Bifet, A., Pechenizkiy, M., and Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Comput. Surv.*, 46(4):44:1–44:37.

Gomes, H. M. and Enembreck, F. (2013). Sae: Social adaptive ensemble classifier for data streams. In *Computational Intelligence and Data Mining (CIDM), 2013 IEEE Symposium on*, pages 199–206.

Gomes, H. M. and Enembreck, F. (2014). Sae2: Advances on the social adaptive ensemble classifier for data streams. In *Proceedings of the 29th Annual ACM Symposium on Applied Computing (SAC)*, SAC 2014. ACM.

John, G. H. and Langley, P. (1995). Estimating continuous distributions in bayesian classifiers. In *Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 338–345, San Mateo. Morgan Kaufmann.

Kolter, J. Z. and Maloof, M. A. (2005). Using additive expert ensembles to cope with concept drift. In *ICML '05 Proc. of the 22nd international conference on Machine learning*, pages 449–456. ACM.

Kolter, J. Z. and Maloof, M. A. (2007). Dynamic weighted majority: An ensemble method for drifting concepts. In *The Journal of Machine Learning Research*, pages 123–130. JMLR.

Kuncheva, L. I. (2004). *Combining Pattern Classifiers: Methods and Algorithms*. John Wiley and Sons, New Jersey.

Kuncheva, L. I., Whitaker, C. J., Shipp, C. A., and Duin, R. P. (2003). Limits on the majority vote accuracy in classifier fusion. *Pattern Analysis & Applications*, 6(1):22–31.

Machado, D. M. S. (2009). O uso da informação na gestão inteligente da segurança pública. In *A força policial: órgão de informação e doutrina da instituição policial militar*, pages 77–85.

Nath, S. V. (2006). Crime pattern detection using data mining. In *Web Intelligence and Intelligent Agent Technology Workshops, 2006. WI-IAT 2006 Workshops. 2006 IEEE/WIC/ACM International Conference on*, pages 41–44.

Oza, N. C. and Russell, S. (2001). Online bagging and boosting. In *Artificial Intelligence and Statistics*, pages 105–112. Society for Artificial Intelligence and Statistics.

Pechenizkiy, M., Bakker, J., Žliobaitė, I., Ivannikov, A., and Kärkkäinen, T. (2010). Online mass flow prediction in cfb boilers with explicit detection of sudden concept drift. *SIGKDD Explor. Newsl.*, 11(2):109–116.

Poelmans, J., Elzinga, P., Viaene, S., and Dedene, G. (2011). Formally analysing the concepts of domestic violence. *Expert Systems with Applications*, 38(4):3116 – 3130.

Schlimmer, J. C. and Granger, R. H. (1986). Incremental learning from noisy data. *Machine Learning*, 1(3):317–354.

Siklóssy, L. and Ayel, M. (1997). Datum discovery. In *Advances in Intelligent Data Analysis, Reasoning about Data, Second International Symposium, IDA-97, London, UK, August 4-6, 1997, Proceedings*, pages 459–463.

Tsymbal, A. (2004). The problem of concept drift: definitions and related work. Technical Report TCD-CS-2004-15, The University of Dublin, Trinity College, Department of Computer Science, Dublin, Ireland.

Wang, T., Rudin, C., Wagner, D., and Sevieri, R. (2013). Learning to detect patterns of crime. In *Machine Learning and Knowledge Discovery in Databases*, volume 8190 of *Lecture Notes in Computer Science*, pages 515–530. Springer Berlin Heidelberg.

Widmer, G. and Kubate, M. (1996). Learning in the presence of concept drift and hidden contexts. *Machine Learning*, 23(1):69–101.

Zliobaite, I. (2010). Learning under concept drift: an overview. *CoRR*, abs/1010.4784.