

UNIVERSITÉ NOUVEAUX HORIZONS



**Étude de la minimisation d'erreur dans
l'apprentissage supervisé en utilisant la
technologie ANPR**

Auteur :

TSHELEKA KAJILA Hassan

Directeur :

Prof. MASAKUNA Jordan

*Mémoire présenté à la Faculté des Sciences Informatiques en vue de
l'obtention du grade de Licencié en informatique.*

en

Calcul Scientifique

24 avril 2022

RÉSUMÉ

Au cours de la dernière décennie, la taille des données a augmenté plus rapidement que la vitesse des processeurs. Dans ce contexte, faire un traitement de reconnaissance des formes dans des images et vidéos, les ensembles de données d'entraînement pour les problèmes de détection d'objets sont généralement très volumineux et les capacités des méthodes d'apprentissage automatique statistique sont limitées par le temps de calcul plutôt que par la taille de l'échantillon.

Le cas des problèmes d'apprentissage à grande échelle implique la complexité de calcul de l'algorithme d'optimisation sous-jacent de manière non triviale. Des algorithmes d'optimisation improbables tels que la **descente de gradient stochastique** (en anglais : **Stochastic Gradient Descent** ou SGD) montre des performances étonnantes pour les problèmes à grande échelle, lorsque l'ensemble d'apprentissage est volumineux.

En particulier, les variants du SGD n'utilisent qu'un seul nouvel échantillon d'apprentissage à chaque itération, sont asymptotiquement efficaces après un seul passage sur l'ensemble d'apprentissage.

Ce travail vise à proposer une méthode intelligente, basée sur l'intelligence artificielle, qui permet aux ordinateurs et aux systèmes informatiques de dériver des informations significatives à partir d'images numériques, de vidéos et d'autres entrées visuelles, avec un coût plus bas que possible. Dans notre contexte la reconnaissance des plaques d'immatriculation des véhicules à l'aide d'un classificateur de la famille de descente de gradient stochastique. Pour minimiser la **fonction coût** du classificateur, la SGD adopte un modèle d'optimisation convexe. De plus, pour augmenter la vitesse de convergence du classificateur, la descente de gradient stochastique, à chaque étape, elle tire un échantillon aléatoire de l'ensemble des fonctions (f_i), de la fonction objectif, constituant la somme.

Mots clés : Apprentissage supervisé, vision par ordinateur, Descente de gradient stochastique, Adaline, ANPR, ALPR.

ABSTRACT

Over the past decade, data size has grown faster than processor speeds. In this context, doing pattern recognition processing in real-time videos, training datasets for object detection problems are usually very large, and the capabilities of statistical machine learning methods are limited by computation time rather than sample size.

The case of large scale learning problems involves the computational complexity of the underlying optimization algorithm in a nontrivial way.

Improbable optimization algorithms such as **Stochastic Gradient Descent** (SGD) show amazing performance for large scale problems, when the training set is bulky.

In particular, SGD variants use only one new training sample at each iteration, are asymptotically efficient after a single pass over the training set.

This work aims to provide an intelligent method, based on artificial intelligence, that allows computers and computer systems to derive meaningful information from digital images, videos and other visual inputs, with a lower cost. as possible. In our context the recognition of vehicle license plates using a classifier of the family of stochastic gradient descent. To minimize the **cost function** of the classifier, the SGD adopts a convex optimization model. Moreover, to increase the speed of convergence of the classifier, the stochastic gradient descent, at each step, it draws a random sample from the set of functions (f_i), of the objective function, constituting the sum.

Key words : Supervised learning, computer vision, Stochastic gradient descent, Adaline, ANPR, ALPR.

TABLE DES MATIÈRES

Résumé	ii
o Introduction	2
0.1 Choix et intérêt du sujet	2
0.2 Problématique	3
0.3 Hypothèse	4
0.4 Objectifs et division du travail	4
 Annexes et Bibliographies	
 Bibliographie	 7

LISTE DES ACRONYMES

ML Machine Learning
CV Computer Vision
OCR Optical character recognition
ANPR Automatic number-plate recognition
ALPR Automatic license plate recognition
API Application Programming Interface
UML Unified Modeling Language



INTRODUCTION

0.1 CHOIX ET INTÉRÊT DU SUJET

L'intelligence désigne communément le potentiel des capacités mentales et cognitives d'un individu, animal ou humain, lui permettant de résoudre un problème ou de s'adapter à son environnement. L'intelligence nous fait ressentir ce besoin d'apprendre pour arriver à nos fins, extresinquement l'intelligence c'est l'apprentissage. Pour que nous puissions dire qu'une machine est intelligente, premièrement elle doit passer par une phase d'apprentissage. Apprendre à résoudre des problèmes ou à réaliser des tâches par lui-même d'une façon autonome. Dans le IA nous parlons de l'apprentissage automatique (en anglais : machine Learning, ML), nous utilisons plusieurs paradigmes d'apprentissage automatique : apprentissage supervisé, apprentissage non supervisé, apprentissage par renforcement, apprentissage en profondeur.

L'apprentissage supervisé représente une grande partie de l'activité de recherche en apprentissage automatique et de nombreuses techniques de ce paradigme ont trouvé une application dans le traitement de contenu multimédia [8]. La caractéristique qui définit ce type d'apprentissage est la disponibilité de données d'apprentissage annotées.

Les algorithmes d'apprentissage supervisé font l'expérience d'un ensemble de données contenant des caractéristiques, et chaque exemple est également associé à une étiquette ou à une cible [13].

L'application de cette étude dans l'apprentissage supervisé est orientée vers la reconnaissance automatique des plaques d'immatriculation (en anglais : automatic number plate recognition, ANPR) dans les images. Une des applications intéressantes parmi tant d'autres dans l'intelligence artificielle. Nous présentons une étude approfondie sur les algorithmes de minimisation de la fonction coût (en anglais : loss function) d'un modèle d'apprentissage appliqué à l'ANPR.

Lorsque nous voulons faire une application dans le traitement de reconnaissance des formes dans des vidéos, les ensembles de données d'entraînement pour les problèmes de détection d'objets sont généralement très volumineux et les capacités des méthodes d'apprentissage automatique statistique sont limitées par le temps de calcul plutôt que par la taille de l'échantillon [4]. Par exemple, pour entraîner une machine à reconnaître des plaques d'immatriculation de voiture, elle doit recevoir de grandes quantités d'images de plaques d'immatriculation et d'éléments liés aux plaques pour apprendre les différences et reconnaître une plaque, en particulier la voiture qui porte une plaque sans défaut. Plus nous avons des données, plus nous gagnons en précision et plus la complexité en

temps augmente.

Des contraintes d'exploitation découlent des observations citées ci-dessus, parmi lesquelles nous citerons celles qui sont liées à la reconnaissance des objets dans les vidéos et images. Par exemple, de nos jours, un très grand nombre de caméras est déployé exclusivement pour la surveillance vidéo [1]. Souvent, le contenu de ces vidéos est interprété par des opérateurs humains qui engendrent des coûts exorbitants pour le suivi et l'analyse du contenu, sans mentionner les erreurs qui peuvent être induites par la fatigue et l'inattention humaine.

0.2 PROBLÉMATIQUE

La complexité de calcul de l'algorithme d'apprentissage devient le facteur limitant critique lorsque l'on envisage de très grands ensembles de données. C'est à ce point critique qu'entre en jeu cette étude, la minimisation des erreurs sans alourdir la complexité en temps et espace de l'algorithme d'apprentissage.

Les ensembles de données d'entraînement pour les problèmes de détection d'objets dans des images sont généralement très volumineux. Minimiser les erreurs dans ces modèles d'apprentissage est une tâche très importante pour renforcer la fiabilité de notre *modèle entraîné* [14].

Établir un algorithme d'apprentissage qui s'adapte au mieux à notre modèle, selon la nature du problème métier traité, il existe différentes approches qui varient selon le type et le volume des données.

L'un des piliers de l'apprentissage automatique est l'optimisation mathématique [6], qui, dans ce contexte, implique le calcul numérique de minimisation des paramètres d'un système conçu pour prendre des décisions en fonction des données disponibles. Ces paramètres sont choisis pour être optimaux par rapport à notre problème d'apprentissage.

Dans l'ensemble, ce document tente d'apporter des réponses aux questions suivantes.

1. Comment les problèmes de minimisation surviennent-ils dans les applications d'apprentissage automatique ?
2. Quelles ont été les méthodes de minimisation les plus efficaces pour les ensembles des données d'apprentissage supervisé à grande échelle et pourquoi ?
3. Comment des algorithmes d'apprentissage supervisé arrivent-t-ils résoudre le problème de la reconnaissance automatique d'objet ?
4. Quelles avancées récentes ont été réalisées dans la conception d'algorithmes de minimisation des erreurs dans l'apprentissage et quelles sont les questions ouvertes dans ce domaine de recherche ?

0.3 HYPOTHÈSE

Le cas des problèmes d'apprentissage à grande ou à petite échelle implique la complexité de calcul de l'algorithme d'optimisation sous-jacent de manière non triviale.

En effet, dans ce travail, nous discutons des algorithmes de descente de gradient stochastique parce qu'ils montrent des performances d'optimisation incroyables pour les problèmes à grande échelle [4].

Le travail de Léon Bottou et al (e. g., dans [4] [20] [5]), présente *la descente de gradient stochastique comme un algorithme d'apprentissage fondamental*.

Une analyse plus précise révèle des compromis qualitativement différents pour le cas des problèmes d'apprentissage à grande échelle [6]. Des algorithmes d'optimisation improbables tels que la **descente de gradient stochastique** (en anglais : **Stochastic Gradient Descent** ou SGD) montre des performances étonnantes pour les problèmes à grande échelle, lorsque l'ensemble d'apprentissage est volumineux. En particulier, le gradient stochastique du second ordre et le gradient stochastique moyennée sont asymptotiquement efficaces après un seul passage sur l' ensemble d'entraînement [4]. Les optimiseurs SGD n'utilisent qu'un seul nouvel échantillon d'apprentissage à chaque itération.

0.4 OBJECTIFS ET DIVISION DU TRAVAIL

Nous nous proposons dans ce mémoire d'aborder sur l'utilisation des algorithmes d'optimisation numérique, précisément de minimisation. Appliquée à l'apprentissage automatique qui permet aux ordinateurs et aux systèmes informatiques de dériver des informations significatives à partir d'images numériques et/ou d'autres entrées visuelles, avec un coût plus bas que possible.

En fait, nous faisons la reconnaissance des plaques d'immatriculation des véhicules à l'aide d'un classificateur de la famille de descente de gradient stochastique (SGD). Pour minimiser la fonction de coût du classificateur, la SGD adopte un modèle d'optimisation convexe [10]. De plus, pour augmenter la vitesse de convergence du classificateur, la descente de gradient stochastique, à chaque étape, tire un échantillon aléatoire de l'ensemble des fonctions (f_i), qui est notre fonction objectif, constituant une somme.

Pour chaque algorithme, nous examinons l'efficacité et comparons le score pour différents cas.

En dehors de cette introduction, la partie conclusive et l'annexe, ce mémoire est organisé en quatre chapitres comme suit.

Chapitre 1 est consacré à quelques rappels des matières sur lesquels je me base pour constituer l'ensemble de ce travail. Nous traitons des considérations de méthodes numériques impliquées dans la résolution de problèmes de minimisation des erreurs d'apprentissage. Certaines discussions

sur les modèles de régression linéaire convexe et de classification dans d'apprentissage supervisé. Nous discutons également du réseau neuronal convolutif le plus adapté pour analyser l'imagerie visuelle.

Chapitre 2 explore une méthodologie parmi tant d'autres, pour entraîner les modèles d'apprentissage automatique de façon optimale, qui nous permettra par la suite de faire une classification d'images pour reconnaissance automatique de plaque d'immatriculation.

Pour la minimisation de la fonction coût nous utilisons des algorithmes comme ASGD, ADAM, ADADELTA, NAG. Puis faire une étude comparative de leurs performances.

Chapitre 3, Ici nous construirons des modèles à partir d'une base de données annotée pour l'apprentissage et pour les tests de reconnaissance d'objets. Les résultats concluants de cette étude pourront conduire à un déploiement de notre système dans les domaines comme celui de la surveillance vidéo de voitures dans une entrée de parking. Des métriques connues pour mesurer les erreurs et en déduire le score du classificateur seront utilisées pour évaluer la qualité de la reconnaissance automatique des plaques d'immatriculation (ANPR) par notre approche.

ANNEXES ET BIBLIOGRAPHIES

BIBLIOGRAPHIE

- [1] Yaovi AHADJITSE. "Reconnaissance d'objets en mouvement dans la vidéo par description géométrique et apprentissage supervisé". Thèse de doct. Université du Québec en Outaouais, 2013.
- [2] Vincent Barra ANTOINE CORNUÉJOLS Laurent Michet. *Apprentissage automatique : Deep learning, concepts et algorithmes*. 3rd. Eyrolles, 2018, p. 239-263.
- [3] Christopher M. BISHOP. *Pattern Recognition and Machine Learning*. First. Springer-Verlag New York, 2006, p. 179-195.
- [4] Léon BOTTOU. "Large-scale machine learning with stochastic gradient descent". In : *Proceedings of COMPSTAT'2010*. Springer, 2010, p. 177-186.
- [5] Léon BOTTOU. "Stochastic gradient descent tricks". In : *Neural networks : Tricks of the trade*. Springer, 2012, p. 421-436.
- [6] Léon BOTTOU, Frank E CURTIS et Jorge NOCEDAL. "Optimization methods for large-scale machine learning". In : *Siam Review* 60.2 (2018), p. 223-311.
- [7] F. COULOMBEAU, G. DEBEAUMARCHÉ, B. DAVID, F. DORRA, S. DUPONT et M. HOCHART. *Mathématiques MPSI-PCSI : Programme 2013 avec algorithmique en Scilab*. Cap Prépa. Pearson, 2013. ISBN : 9782744076527. URL : <https://books.google.cd/books?id=e4vfnQEACAAJ>.
- [8] Pádraig CUNNINGHAM, Matthieu CORD et Sarah Jane DELANY. "Supervised learning". In : *Machine learning techniques for multimedia*. Springer, 2008, p. 21-49.
- [9] R.B. DARLINGTON et A.F. HAYES. *Regression Analysis and Linear Models : Concepts, Applications, and Implementation*. Methodology in the Social Sciences. Guilford Publications, 2016. ISBN : 9781462521135. URL : <https://books.google.cd/books?id=YDgoDAAAQBAJ>.
- [10] Natarajan DEEPA, B PRABADEVI, Praveen Kumar MADDIKUNTA, Thippa Reddy GADEKALLU, Thar BAKER, M Ajmal KHAN et Usman TARIQ. "An AI-based intelligent system for healthcare analysis using Ridge-Adaline Stochastic Gradient Descent Classifier". In : *The Journal of Supercomputing* 77 (2021), p. 1998-2017.
- [11] Kary FRÄMLING. "Scaled Gradient Descent Learning Rate". In : *Reinforcement Learning With Light-Seeking Robot, Proceedings of ICINCO* (2004), p. 1-8.
- [12] Yoav FREUND et Robert E SCHAPIRE. "Large margin classification using the perceptron algorithm". In : *Machine learning* 37.3 (1999), p. 277-296.
- [13] I. GOODFELLOW, Y. BENGIO et A. COURVILLE. *Deep Learning*. Adaptive Computation and Machine Learning series. MIT Press, 2016. ISBN : 9780262035613. URL : <https://books.google.cd/books?id=Np9SDQAAQBAJ>.

- [14] Daniel Kirsch JUDITH HURWITZ. *Machine Learning For Dummies*. IBM Limited Edition. John Wiley et Sons, Inc., 2018.
- [15] N. MATLOFF. *Statistical Regression and Classification : From Linear Models to Machine Learning*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press, 2017. ISBN : 9781351645898. URL : <https://books.google.cd/books?id=IHs2DwAAQBAJ>.
- [16] Praneeth NETRAPALLI. "Stochastic gradient descent and its variants in machine learning". In : *Journal of the Indian Institute of Science* 99.2 (2019), p. 201-213.
- [17] Jorge NOCEDAL et Stephen J WRIGHT. *Numerical optimization*. Springer, 1999.
- [18] Carl-Erik SÄRNDAL, Bengt SWENSSON et Jan WRETMAN. *Model assisted survey sampling*. Springer Science & Business Media, 2003.
- [19] Vahid Mirjalili SEBASTIEN RASCHKA. *Python Machine Learning and Deep Learning, with sckit-learn and Tensorflow*. 2nd. Packt, 2017, p. 17-139.
- [20] Rob GJ WIJNHOFEN et PHN de WITH. "Fast training of object detection using stochastic gradient descent". In : *2010 20th International Conference on Pattern Recognition*. IEEE. 2010, p. 424-427.