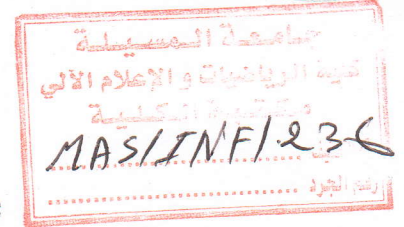




UNIVERSITE MOHAMED BOUDIAF - M'SILA
FACULTE DES MATHÉMATIQUES ET
DE L'INFORMATIQUE



DEPARTEMENT D'INFORMATIQUE



MEMOIRE de fin d'étude
Présenté pour l'obtention du diplôme de MASTER
Domaine : Mathématiques et Informatique

Filière : Informatique

Spécialité : Systèmes d'Informations Avancés

Par: Alahoum Oussama

SUJET

Traitement des chèques postaux (le cas de reconnaissance du montant littéral Arabe)

Soutenu publiquement le : 31/ 05/ 2016 devant le jury composé de :

A.Therafi	Université de M'sila	Président
Dr. Assas Ouarda	Université de M'sila	Rapporteur
A.Guemmougui	Université de M'sila	Examineur

Promotion : 2015 /2016

Sommaire :

Listes des figures.....	viii
Listes des tableaux :.....	ix
Listes des Abréviations :	x
Introduction générales :.....	2
CHAPITRE 1 L'ECRITURE ARABE.....	4
1. Introduction.....	5
2. Ecriture Arabe :.....	5
2.1. L'écriture arabe :.....	5
2.2. Alphabet arabe :	5
2.3. Le type de l'écriture :	6
2.4. Caractéristiques de l'écriture Arabe :.....	7
2.4.1. Signes diacritiques :.....	9
2.4.2. Une ou plusieurs composantes connexes par mot.....	12
3. Les systèmes de reconnaissance de l'écriture :	13
3.1. Systèmes de reconnaissance en ligne.....	14
3.2. Systèmes de reconnaissance hors ligne.....	15
3.3. Quelques exemples d'applications.....	15
3.4. Etapes du processus de reconnaissance hors-ligne automatique de l'écriture manuscrite.....	16
3.4.1. Etape d'acquisition.....	16
3.4.2. Etape de prétraitement.....	17
3.4.3. Etape d'extraction de caractéristiques ou de primitives :.....	20
4. Conclusion.....	21
CHAPITRE 2 RECONNAISSANCE DES MONTANTS DES CHEQUES.....	22
1. Introduction :.....	23
2. Les chèques :.....	23
2.1. Historique des chèques	23
2.2. Les avantages du chèque.....	23
2.3. Les inconvénients du chèque	23
2.4. Quelques exemples des chèques en Algérie.....	24
2.5. Le chèque postal algérienne :.....	25
3. Architecture et description du système de reconnaissance utilisé :	25

4. Les Méthodes :	31
4.1. Méthodes neuro-mimétiques:	31
4.2. Méthodes markoviennes :	31
4.3. Classificateur euclidien.....	32
4.4. La méthode du plus proche voisin :	32
5. Quelques systèmes de traitement automatique des montants littéraux Arabes :.....	32
6. Conclusion :	37

CHAPITRE 3 GENERALITES SUR LES RESEAUX DE NEUROUNE ET MACHINE A VECTEURS DE SUPPORT

1. Introduction	39
2. Les réseaux de neurone	39
2.1. Historique	39
2.2. Base biologique	39
2.3. Les Différents Types Des RNA :	42
2.3.1. Les réseaux de propagation (feed-forward).....	42
2.3.2. Les réseaux de retro-propagation (feedback) :	43
2.4. L'apprentissage :	44
2.4.1. Les Types D'apprentissage :	44
2.4.2. Le Perceptron Multicouches :	45
2.5. Les avantages des réseaux de neurones :	46
3. Les Séparateurs à Vaste Marge (SVM).....	46
3.1. Historique	47
3.2. Définition	47
3.3. Domaines d'application	48
3.4. Principe des SVMs :Notions de base: Hyperplan, Marge, Vecteurs de support.....	48
3.5. Fondements mathématiques	49
3.6. SVM multi-classes	52
4. Avantages et inconvénients.....	52
5. Conclusion.....	53

CHAPITRE 4 IMPLIMENTATION DE SYSTEME PROPOSE ET RESULTATS

1. Introduction	55
2. L'environnement de Développement :	55
3. Architecture du système de reconnaissance de montant littéral de chèque postal :	56
3.1. L'acquisition :	57
3.1.1. La base de données utilisée :	58

3.2.	Prétraitement :	59
3.3.	Détection de zones d'intérêt :	61
3.4.	Segmentation :	62
3.4.1.	La segmentation du texte en lignes :	62
3.4.2.	La segmentation des lignes en parties connexes	63
3.5.	Normalisation	63
3.6.	L'extraction des primitives	64
3.7.	La classification et reconnaissance :	65
3.7.1.	Classifieur basé a Svm :	65
3.7.2.	Classifieur basé a Knn :	67
3.7.3.	Classifieur basé RNA	68
4.	Comparaison des résultats de reconnaissance entre les différents classifieurs :	69
5.	Mesure de performance de base des chèques :	70
6.	Interface utilisateur	72
7.	Conclusion	74
	CONCLUSION GENERALE	75
	Bibliographie :	77
	Figure 2.1 Différent modèle des cheque	24
	Figure 2.2 cheque postal algérien	25
	Figure 2.3 Architecture du système de reconnaissance de montant littéral	26
	Figure 2.4 Image d'un cheque de la base d'images CNPARI	33
	Figure 2.5 le montant littéral	34
	Figure 2.6 le processus du système de reconnaissance de montant littéral	35
	Chapitre 3	
	Figure 3.1 Schéma d'un Neuron Biologique	40
	Figure 3.2 Schéma d'un modèle formel	41
	Figure 3.3 Fonctions d'activation	41
	Figure 3.4 Architecture d'un Perceptron Multicouche à une seule couche cachée	46
	Figure 3.5 Hyperplan optimal, marge maximale et vecteurs de support	48
	Figure 3.6 Hyperplan séparateur dans le cas d'un linéairement séparable	49
	Chapitre 4	
	Figure 4.1 L'interface de Matlab	56
	Figure 4.2 Schéma général de notre système de reconnaissance	57
	Figure 4.3 Exemple de 15 exemple de partie connexe ($L=3$) de notre base	58
	Figure 4.4 Quelques exemples des chèques postaux de notre base	59
	Figure 4.5 Un cheque postal avant et après la binarisation	60
	Figure 4.6 l'image d'un partie connexe de base et sa binarisation	60
	Figure 4.7 Un cheque postal avant et après la Dilatation	61
	Figure 4.8 détection de zone d'intérêt de cheque postal (le montant littéral)	61
	Figure 4.9 la segmentation du de montant littéral	62

Introduction générales :

La reconnaissance de l'écriture manuscrite est le vieux rêve de tous ceux qui ont eu besoin d'entrer des données dans un ordinateur. Il remonte à plus d'une trentaine d'années. Aujourd'hui, il existe plusieurs domaines dans lesquels la reconnaissance de l'écriture manuscrite est attendue avec impatience, par exemple dans le tri automatique du courrier, le traitement automatique de dossiers administratifs, des formulaires d'enquêtes, ou encore l'enregistrement des chèques bancaires.

La reconnaissance des textes cursifs reste toujours un problème ouvert aussi bien dans sa forme imprimée que manuscrite. Ceci à cause des difficultés auxquelles sont confrontés les chercheurs et les développeurs, telles que la variabilité de la forme, du style, et de l'inclinaison de l'écriture. L'écriture manuscrite arabe est naturellement cursive, difficile à traiter, et présente une grande variabilité.

Les chèques bancaires et postaux font partie des documents les plus utilisés, et l'automatisation de son traitement devient un domaine de recherche important du point de vue économique et scientifique. Un chèque contient beaucoup d'informations qui sont souvent manuscrites, on cite parmi autres un montant numérique et un montant littéral. Ce dernier est le seul acceptable d'un point de vue légal. Cependant l'information redondante entre les deux montants (littéral et numérique) peut augmenter le taux de reconnaissance et par conséquent l'amélioration des performances du système.

Notre travail s'insère dans le cadre général de la reconnaissance automatique hors-ligne(off-line) de l'écriture arabe manuscrite à vocabulaire limité, avec application sur la reconnaissance des montants littéraux arabe des chèques postaux.

Notre mémoire se subdivise donc comme suit :

- ❖ Dans le premier chapitre, nous présentons un aperçu sur l'écriture arabe et leur caractéristiques Puis un vu générale sur les étapes suivies pour la réalisation d'un système de reconnaissance en général.
- ❖ Le deuxième chapitre est consacré aux quelques caractéristique des chèques postaux ainsi l'état de l'art dans le domaine de la reconnaissance automatique des chèques où nous décrivons les étapes nécessaires au développement d'un système de reconnaissance des chèques.

- ❖ Le troisième chapitre se répartit en deux sections, la première un vu générale sur les réseaux de neurone avec ces types et la seconde présente une idée globale sur les machine a vecteur de support.
- ❖ Le dernier chapitre décrit les étapes de la reconnaissance appliquées dans notre système et les expériences effectuées dans le cadre de ce travail et l'analyse des résultats obtenus avec quelques exemples des interfaces de notre application.

Nous finirons ce travail par une conclusion et des perspectives sur les travaux futurs dans ce domaine de recherche.

Conclusion et Perspectives

Malgré les efforts et les travaux réalisés dans le domaine de la reconnaissance des chèques arabes, le système idéal n'existe pas. Et un bon système nécessite de faibles taux de rejets et des taux d'erreurs très faibles car le coût des erreurs est extrêmement élevé. Cela dépend d'une grande part des techniques utilisées au cours du processus de développement.

Dans ce travail nous avons présenté en détail les différentes étapes d'un système de reconnaissance hors-ligne de l'écriture, dédié à une application spécifique à vocabulaire limité : la lecture automatique des montants littéraux arabe des chèques postaux.

Ce travail repose sur la reconnaissance globale du partie connexe (PAW). Le choix de cette approche se justifie par une raison simple, lorsqu'on examine de plus près les caractéristiques de l'écriture arabe, on arrive et sans aucune difficulté, pour réduire le degré de complexité de conception et pour ne pas emprunter des voies pleines d'obstacles, au choix de la méthode globale comme stratégie de reconnaissance.

Notre contribution s'articule autours du quatre points suivants:

- a. réalisation d'une base de donnée pour les parties connexes rencontrés dans les montants littéraux arabe des chèques.
- b. on appliqués quelques prétraitements (binarissations et dilatation).
- c. La détection de la zone d'intérêt l'étape le plus importante dans le domaine de la reconnaissance de chèques qui contient le montant littéral arabe du chèque.
- d. La segmentation de montant littéral en ligne puis en partie connexe(PAW).
- e. L'utilisation des moments de hu pour l'extraction des caractéristiques des parties connexes.
- f. La reconnaissance et le classification du montant avec trois différent classifieur (Rna ,Svm ,Knn).

Grâce à l'introduction de ces propositions, notre système donne des résultats très encourageants, dans ce qui concerne les performances et la complexité.

Les perspectives pour ce travail sont nombreuses. Les plus importants sont :

- Le prétraitement : tel que l'adaptation des lignes de base pour chaque élément d'un mot composé, la correction des caractères penchés,...etc.
- D'après le premier jeu de tests effectués, la majorité des erreurs de reconnaissance peuvent être attribuée à la phase d'extraction de caractéristiques où les méthodes employées donnent, dans certains cas, des fausses alertes. Donc, si on pourra dans les prochaines années édifier un module d'extraction robuste.
- Une perspective majeure consiste à établir la phase de post-traitement en combinant plusieurs classifieurs(HMM,RBF....) .

Bibliographie :

- [1] Abdelhak Boukharouba • Abdelhak Bennia « Recognition of Handwritten Arabic Literal Amounts Using a Hybrid Approach » Published online: 24 December 2010
- [2] Abdi H, " Intelligence Artificielle et Intelligence Naturelle", Presse Universitaires de France in J. P Rossi J. F Le Ny J. F (Eds), 1993.
- [3] A.Belaid, «Reconnaissance automatique de l'écriture et du document» , Pour la science, disponible sur le lien web :<http://webloria.loria.fr/~abelaid/Publications.html>, 2001.
- [4] Achour K, Zenati N, Djekoune O, "Contribution d'image par un modèle de réseaux de neurones", 1998.
- [5] Al badr B, mahmoud S. A. " survey and bibliography of arabic optical text recognition" signale processing, vol 41, pp:49-77,1995.
- [6] AYAT.N.E, CHERIET.M, SUEN.C.Y. "Automatic model selection for the optimization of SVM kernels". ELSEVIER, the journal of pattern recognition society. Pattern recognition 38 (2005) 1733-1745.
- [7] Bèdda. Mouldi, Ramdani. Messaoud, Doghmane Nourreddine, classification des caractères manuscrits arabes par un perceptron multicouches, revue des sciences et technologies synthèse n°1, université Annaba, pp. 6-11, juin, 1996.
- [8] BenJ.A.krose P.patrick van der smagt, " An introduction to neural networks" universityof amsterdam faculty of mathematics and computer science: pages 33- 42, Fifth edition january 1993.
- [9] BOTTOU.L."Comparison of Classifier Methods : A Case Study in Hand written Digit Recognition". Pattern Recognition,1994. Vol. 2 -Conference B: Computer Vision & Image Processing. Proceedings of the 12th IAPR International. Conference on: oct1994. Jerusalem.
- [10] Boukhari.w, "Identification Biométrique des Individus par leurs Empreintes Palmaires (Palprints) ", Mémoire de Magister, USTOran, Octobre 2007.
- [11] C. Chatelain, «Extraction de séquences numériques dans des documents manuscrits quelconques», Thèse de doctorat ; Université de Rouen, 2006.
- [12] C.C. Tappert, C.Y. Suen, T. Wakahara, «The state of the art in online handwriting recognition», IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 12, n°. 8,pp.787-808, 1990.

- [13] Cohen David, 1993, « Arabe : langue arabe " Enc. Universalis, vol 2, pp 707 –714. Encyclopédie de l'Islam, 1975-, collectif, « Arabiyya » vol 1, pp579-601.
- [14] CORTES.C, VAPNIK.V. "support-vector networks, Machine learn". 20 (3) (1995) 273-297.
- [15] Dayan P, Abbott L. F, " Theoretical Neuroscience", Book, 2000
- [16] Doermann and S. Jaeger. "Arabic and Chinese Handwriting Recognition", Springer, Germany, 2006.
- [17] Dreyfus G, " Les Réseaux de Neurones, Mécanique Industrielle et Matériaux ", No.51, septembre 1998.
- [18] F. Menasri, "Contributions à la reconnaissance de l'écriture arabe manuscrite," in thèse de l'Université Paris Descartes, 2008.
- [19] Freeman J. A, Shapriua D.M, "Neural Networks: algorithms, Application and Programming Techniques", Book, Addison Wesley Publishing Company, INC, 1991.
- [20] H. Miled, «Stratégies de reconnaissance de l'écriture semi cursive : application aux mots manuscrits arabes», Thèse de doctorat, Université de Rouen, 1998.
- [21] H. Mouchère, "Étude des mécanismes d'adaptation et de rejet pour l'optimisation de classifieurs: Application à la reconnaissance de l'écriture manuscrite en-ligne," in thèse de l'Institut National des Sciences Appliquées de Rennes, 2007.
- [22] <http://svms.org/history.html>. Date d'accès 01-05-2016.
- [23] <https://fr.wikipedia.org/wiki/> consulter le 26 avril 2016.
- [24] J.P. Crettez, G. Lorette, «Reconnaissance de l'écriture manuscrite », Techniques de l'Ingénieur, Dossier : H1358, Date de parution : 02/1998
- [25] KECMAN.V, "Learning and Soft Computing Support Vector Machines, Neural Networks, and fuzzyLogicModels", The MIT Press 2001.
- [26] K. S. Nathan, H. S. Beigi, J. Subrahmonia, G. J. Clary, et H. Maruyama, "Real-time on-line unconstrained handwriting recognition using statistical methods," in International Conference on Acoustics, Speech, and Signal Processing, p. 2619–2622, 1995.
- [27] L. Souici-Meslati, M. Sellami, «A hybrid approach for arabic literal amounts recognition», AJSE, the Arabian Journal for Science and Engineering, Volume 29, Number 2B, pp. 177-194, October 2004.
- [28] Maddouri S S, Amiri H. Combination of local and global vi-sion modelling for Arabic handwritten words recognition. In Proc. the 8th International Workshop on Frontiers in Hand-writing Recognition, Aug. 2002, pp.128-135.

- [29] N. Mezghani, A. Mitiche, et M. Cheriet, "A new representation of shape and its use for high performance in online Arabic character recognition by an associative memory," *International Journal on Document Analysis and Recognition*, vol. 7, n°. 4, p. 201–210, 2005.
- [30] N.E. Ayat , «Selection de modele automatique des machines a vecteurs de support: application a la reconnaissance d'images de chiffres manuscrits» , Thèse de doctorat , 2004.
- [31] N. Farah, L. Souici, M. Sellami, «Classifiers combination and syntax analysis for Arabic literal amount recognition», *Engineering Applications of Artificial Intelligence*, vol. 19, pp. 29- 39, 2006.
- [32] N Nemmour, Application des machines à vecteurs de support pour la reconnaissance des chiffres manuscrits. Thèse de doctorat, USTHB Alger, Algérie, 2010.
- [33] O. Hachour, «Reconnaissance hybride des caractères Arabes imprimés», *JEP-TALN*, Fès, Avril 2004.
- [34] OSUNA.E, FREUND.R, GIROSI.F. "Training support vector machines: an application to face detection". In: *CVPR 97: proceeding of the 1997 conference on computer vision and pattern recognition*. IEEE computer society, 1997, pp.130-136.
- [35] R. Plamondon, S. Srihari, «On-line and Off-line Handwriting Recognition: A Comprehensive Survey», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, n°. 1, pp. 63-84, 2000.
- [36] S. Alma'adeed, C. Higgins, D. Elliman, «Off-line recognition of handwritten Arabic words using multiple hidden Markov models», *Knowledge-Based Systems*, vol. 17, pp.75–79, 2004.
- [37] Sander Bohte, " Spiking neural networks ", Thesis University Leiden.2003,
- [38] S. Carbonnel, E. Anquetil, «Apprentissage automatique d'une distance d'édition dédiée à la reconnaissance de l'écriture manuscrite», *Conférence Internationale Francophone sur l'Écrit et le Document CIFED'04*, 2004.
- [39] S. Chevalier, M. Lemaître, E. Geoffrois, «Étude de primitives spectrales pour la reconnaissance de caractères manuscrits dans le cadre d'une approche markovienne 2D» , Actes 15ème Congrès Francophone AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle (RFIA'2006), Tours, France (CDROM), 2006.
- [40] SCHMIDT.M. "Identifying speakers with Support Vector networks". In: *Proceedings of the 28th symposium on the interface*, 1996, Sydney.
- [41] SCHOLKOPF.B, SOMLA.A.J, "Learning with kernels, Support vector

machineregularization, optimization, and beyond”, the MIT press 2002.

[42] Shih-chi huang and yih-fang huang, "Learning algorithms for perceptrons using back- propagation with selective update" , paper, 1990.

[43] Simon haykin, " Neural networks a comprehensive foundation" ,SBN0-13-9083855 prentice hall international, Inc pages: 34-78, second edition, 1999

[44] S.Kermi " Classifieur neuronal base connaissances, application a la reconnaissance des caractères arabes isoles manuscrits ". Thèse de magister,université Badji Mokhtar, Annaba, Algérie 1999.

[45] Suici-Meslati, " Reconnaissance de formes et écriture arabe manuscrite" ,support de cours 2008.

[46] Vapnik, V., The Nature of Statistical Learning Theory. Springer Verlag, New York; 1995.

[47] W, Boukhari., "Identification Biométrique des Individus par leurs Empreintes Palmaires (Palmprints) ", Mémoire de Magister, USTOran, Octobre 2007.

[48] www.memoireonline.com/05/10/3498/m_les-differents-modes-de-paiement-0.html consulter le 26 avril 2016.

[49] Xueying Zhang & al. "Evaluation of a set of new ORF kernelfunctions of SVM for speech recognition", engappai, Elsevier ltd, 2013.

[50] Y. Al-Ohali, M. Cheriet, C. Suen, «Databases for recognition of handwritten Arabic cheques», Pattern Recognition vol. 36 , pp. 111-121, 2003

ملخص

في إطار العمل الذي قمنا به ، و المتمثل في وضع نظام للتعرف الآلي على المبالغ الحرفية المخطوطة بحروف عربية و تطبيق هذا النظام على الشيكات البريدية الجزائرية و ذلك باستعمال عدة مصنفات هي الشبكات العصبونية و شعاع الدعم الآلي و ن-الأقرب الجار، وكذلك إنشاء قاعدة بيانات جديدة. من أجل ذلك، في بادئ الأمر قمنا في المرحلة الأولى والتي تتمثل في معالجة صورة الصك التي تم إدخالها إلى الجهاز عن طريق المساح الضوئي ثم القيام بتحديد مكان كتابة المبلغ الحرفي الذي تم تقسيمه . المرحلة التالية تتمثل في استخلاص الخصائص المميزة لكل كلمة و ذلك من أجل استخدامها فيما بعد بواسطة المصنف وبعد ذلك تحويل المبلغ الحرفي إلى مبلغ رقمي.

كلمات مفتاح: التعرف، الصك البريدي، شبكات العصبونية، الكتابة العربية، المبلغ الحرفي، شعاع الدعم الآلي، ن-الأقرب جار.

Abstract

This work is interested more particularly in the realization of a system of recognition of Arab handwritten in the automatic reading of the Arabic literal amounts of postal checks based on different classifiers which are: RNN, SVM and KNN. Our aim is to propose a methodology for the implementation of this system. With this intention, The initial phase consists in the pretreatment and detection of the literal amount that is segmented and that allows the preparation of the data. This data is then exploited by the module of extraction that introduced the most relevant characteristics for the recognition. The results of this stage are transformed into sequences of information in the form of vectors. The latter are used as entered for the module of recognition then, convert Arabic literal amounts to numeric amount.

Key words: recognition, postal check, neural network, KNN, SVM, Arabic handwriting, literal amounts.

Résumé

Dans le cadre de notre travail, qui consiste à développer un système de reconnaissance de l'écriture manuscrite appliquée aux montants littéraux Arabe des chèques postaux algériens, basé sur des différentes classifieurs qui sont :RNA ,SVM ,KPPv Notre travail consiste également à la conception d'une nouvelle base de données. L'objectif visé est de proposer une méthodologie pour la mise en œuvre de ce système. Pour ce faire, nous proposons dans un premier temps une phase de prétraitement, puis la détection de la zone du montant littéral arabe qui est segmenté et qui permet de préparer les données. Ces données sont ensuite exploitées par le module d'extraction des caractéristiques les plus pertinentes pour la reconnaissance. Les résultats de cette étape sont transformés en séquences d'informations sous forme de vecteurs. Ces derniers se servent des entrées au module de reconnaissance pour convertir le montant littéral arabe vers un montant numérique.

Mots clés : reconnaissance, chèque postal, réseaux de nouerons, SVM, kppv écriture arabe, montant littéral.