تسخير شبكة العصبونات في التعرف الآلي على الخط اليدوي العربي

الأستادة : سهام موساوي

تخصص : لسانيات تطبيقية

العنوان : ص.ب: 212 إمامة تلمسان 1300 الجزائر

الهاتف:

00213797250692 /0021372506595

الفاكس:

0021343277431

البريد الإليكتروني:

sihem _1380@ hotmail.com

ملخص

تعد شبكة العصبونات الاصطناعية فرع من فروع الذكاء الاصطناعية لتمييز المواد اللغوية وتوليدها ،و نقصد التعرف الآلي على الخط اليدوي العربي هي عملية فهم و تمييز العربية المكتوبة بخط اليد من قبل الحاسوب لإخراجها حروف مطبوعة بعد مرورها بعدة عمليات رقمية و إلكترونية مطواعة للحاسوب ،بواسطة التدريب في شبكة العصبونات الاصطناعية ذلك نظرا لأهمية المخطوطات العربية باعتبارها المخزون الثقافي الذي يحفظ لنا واقعا حضاريا و مرجعا علميا لا يمكن الاستغناء عنه يستلزم المحافظة عليه في ذاكرة الحاسوب و التعرف عليه آليا و عليه يعتمد مستقبل اللغة العلمي و الاقتصادي خصوصا بعد أن توطدت العلاقة بين اللغة العربية و التكنولوجيا.

لذا ، تتمثل إشكالية المقال فيما إذا كان بإمكاننا إيجاد قواعد لسانية وقوالب تضبط شكل الحرف وتسهل عملية فهم الحاسوب لهذه الرموز الخطية البسيطة والمعقدة الشاقة والمكلفة مما يجعل البحث فيها في منأى عن الوقوع في الأخطاء الطبيعية والرسمية (نحوية، صرفية ...) خلال التعامل مع المخطوطات اليدوية.

من بين التقنيات الآلية المساعدة في عملية التعرف الآلي على المواد اللغوية هي شبكة العصبونات الاصطناعية ،« ويتلخص مفهوم هذه الأخيرة بشكل عام ، في استعمال ميكنيزمات المخ البشري لبناء أنظمة حساب أكثر قدرة على حل نوع ما من المشاكل التي يصعب على الإنسان حلها بشكل دقيق و سريع »[1]والاختلاف الأساسي و الجوهري بين شبكة العصبونات

و الدماغ البشري، يكمن في نمط الأنظمة الداخلية و ذلك « قصد تحقيق وظيفة التعرف على معلومات معينة سبق و قد قدمت للدماغ على شكل معطيات معينة لتسهيل هذه العملية ، فخلايا النورون تستقبل المعطيات المرسلة إليها عن طريق المدخل ، وتتم عملية التدريب في العلبة السوداء المتواجدة فيها خلايا النورون ، ثم نصل إلى النتيجة الفعلية ، فأما أن تكن مطابقة للقاعد النظرية ، أو أن تكون عكس ذلك و في هذه الحالة تقوم بعملية اختبار لمعرفة مدى استيعاب خلايا النورون لتلك القواعد المخزونة لديه في قاعدة بيانات محددة .

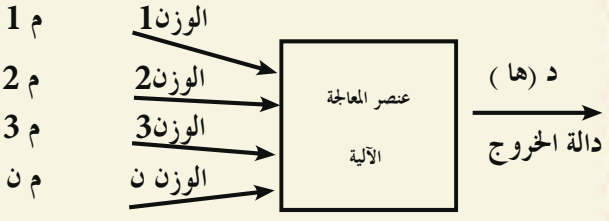
و يرتكز خلايا العصبونات في تغير و ترتيب للأوزان على قاعدة المقارنة بين الخروج الفعلي و الخروج النظري للمعطيات اللغوية المتمثل في دالة حسابية ، إلى غاية أن يصبح التباين بينهما أقل قيمة »[2].

المفهوم الإلكتروني للعصبون :

يتمثل نموذج العصبون الإلكتروني عن طريق وحدات المعالجة المسماة بالعصبونات الصورية حيث تتصل كل وحدة معالجة بأخرى و تكون متشابهة مع بعضها ، فلدينا خمس عناصر

لعمل العصبون الإلكتروني و هي :

- المداخل التي تمثل طبيعة المعطيات
- دالة التنشيط
- معالجة العناصر المكونة لقاعدة معطيات
- المخارج التي تمثل طبيعة خروج المعطيات كالشكل الآتي:



الشكل الأول: عمل العصبون الإصطناعي

فبعد تزويد الشبكة بالمعطيات اللازمة و تدريب العصبون على القواعد نقوم بعملية اختبار الكفاءة بإدخال معطيات من نفس الصنف ، فإذا حصلنا على الدالة مثلا

تا (س) = 1 فإن العملية ناجحة ، و إذا كانت النتيجة تا (س) =0 فإن العصبون لا يزال بحاجة إلى تدريب »[3]

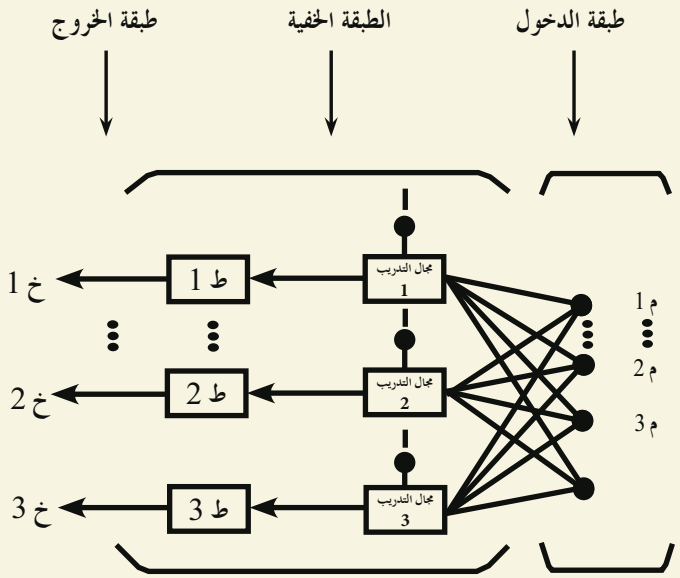
أنواع الشبكات العصبية: « تتكون الشبكات العصبية من وحدات معالجة بسيطة ، هذه الوحدات ماهي إلا عناصر حسابية تسمى عصبونات و التي لها خاصية عصبية، من حيث أنها تقوم بتخزين المعرفة العملية و المعلومات التجريبية لتجعلها متاحة للمستخدم، وذلك عن طريق ضبط الأوزان

و تتكون شبكة أحادية الطبقة من ثلاثة مكونات رئيسية:

1- الشبكية: تسمح بتقديم معطيات (دخول خروج) على شكل ثنائي

2- طبقة الخلايا المقررة : تستقبل المعطيات من الخلايا

3- الشبكة المتعددة الطبقات: هي شبكة ذات الطبقة الواحدة لكن بإدماج عدة طبقات وسيطة بينها و تسمى الطبقة الخفية و في هذه الحالة تكون عصبونات الطبقة الواحدة في معزل عن بعضها البعض في حين تكون مع اتصال بعصبونات الطبقة المجاورة (السابقة و اللاحقة) و يتعلق الأمر دائما بمحاولة التقليل من قيمة الخطأ و تكون على الشكل التالي»[4] :



الشكل الثاني: مخطط الشبكة المتعددة الطبقات

م = عدد المداخل، ط = الطبقة الخفية ، خ = شعاع الخروج

تطبيق الشبكة العصبية للتعرف الآلي على الوحدة الخطية

نقصد بعملية التعرف الآلي على الخط اليدوي العربي و هي« عملية فهم و تمييز الحروف العربية المكتوبة بخط اليد، بعد محاولة فك رموز الرسالة المراد تبليغها، وتهتم هذه الطريقة بالنصوص

المكتوبة ، حيث تهدف إلى تشخيص الحروف الفردية حرفا ،حرفا ثم تحديد الكلمات و الجمل و تحتاج هذه العملية لتطبيقها إلى التدقيق في أشكال الوحدات الخطية من حيث الطول والعرض و الحجم و تجاورها مع غيرها ، فتحتاج هذه العملية الآلية إلى معارف علمية دقيقة و معلومات لسانية، فيقارن الحاسوب بين الوحدة الخطية التي يستقبلها و الأطياف المخزنة لديه مسبقا بحيث، يقوم النظام بتخزين أنماط خطية معينة على هيئة طيف من الموجات الكهرومغناطيسية »[5]، ولكي يتمكن النظام الآلي من التعرف على الرمز الخطي الذي تحمله هذه الإشارة يقوم بتحليلها بطرق هندسية و رياضية لنستخلص منها الخصائص الأساسية للحروف كسعتها و درجة شدتها و معدلات انحدارها و صعودها و مركز ثقلها ثم نقارن هذه المعطيات بتلك المخزنة في النظام ، وتحتوي هذه الطريقة على مناهج مختلفة في تمييز الخط باختلاف الهدف والوسيلة و تشمل هذه الطريقة على سلسلة مكونة من أربع خورزمات هي :

1- تحليل الطيف الخطي :

و تستعمل فيها طرق التحليل الترددي و الترابط الذاتي و تقنيات التقاطع مع الصفر للإشارة الخطية ، وكلها تقنيات تسهل عملية ترشيح الخط و إزالة الشوائب المحيطة به

2- تحديد الخصائص العامة للوحدة الخطية:

تستعمل هذه الخورزميات لاستخراج السمات في العينة الخطية المختلفة على وتيرة زمنية مناسبة على طول الوحدة الخطية و حجمها ، ثم تخزن في مصفوفة معيارية و من أمثلة هذه السمات الارتفاع ، و العرض و، المساحة ....و غيرها

3- تسوية شكل الوحدة الخطية:

إن السرعة و طريقة الكتابة تختلف من شخص لآخر أو حتى من زمن لآخر لذا لابد من توحيد هذه الطريقة

و السرعة في الكتابة ، ضمن الحاسوب ، وذلك بضبط جميع اختلافات الخط عن طريق أخذ العينات المختلفة

4- إنشاء قاعد معطيات:

« هي الخورزميات الأخيرة في عملية التعرف ، وتتضمن تقنيات مختلفة انطلاقا من سلسلة من المعارف والمرور بسلسلة من المستويات اللغوية خصوصا، وأن الخط اليدوي العربي متعدد الأحجام و الأشكال في مجموعة منتهية تحتوي على 29 حرفا ذوي أربع مواضيع هي : بداية الحرف : وسط الحرف، وآخر الحرف بالإضافة إلى حالة تطرف الحرف و حركات الفتح و الكسر والضم و السكون التي لها حالات التنوين بالفتح، و التنويت بالكسر و حركات المد (ا ، و ، ي) كل هذه الأشكال المختلفة و المتعددة التي تحمل 116 شكلا للحروف (29 X 4 = 116)و 12 شكلا للحركات (4 X 3 = 12)و 3 حركات المدى»6 كما نجد مستويات لغوية تنطلق منها الإشارة الخطية لإنجاح عملية التعرف منها :

ا المستوى الصرفي :

تدرس هيئة الوحدة الخطية ، وتشمل قواعد الجمع و التصريف و الإبدال لشكل الوحدة الخطية ..و غيرها من التبديلات

ا المستوى التركيبي :

يدرس تركيب الوحدة الخطية و مطابقتها لقواعد النحو

ا المستوى الدلالي :

يختبر الوحدات الخطية الصحيحة إملائيا و الخاطئة دلاليا»7

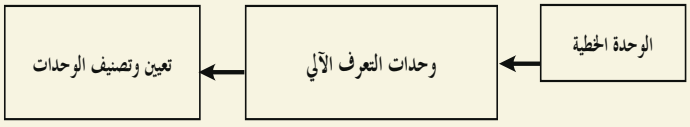
5-التطبيق :

تعد الشبكة العصبية للتعرف الآلي على الوحدة الخطية العربية فرع من فروع الذكاء الاصطناعي و يعني :

أولا: تمييز الخط أي التعرف الآلي على الخط اليدوي العربي

ثانيا : إنتاج الخط اليدوي العربي، أي توليد حروف مطبعية تحاكي حروف الخط العربي كما يوضح الشكل التالي :

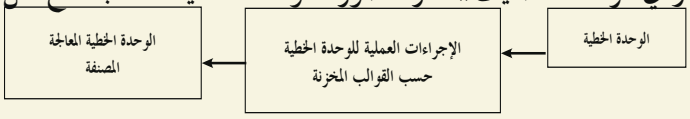
وتتم هذه العملية بعد المرور بمراحل تعليم الشبكة التي تقوم بتخزين



المشكل الثالث: عملية تمييز وإنتاج الخط اليدوي العربي

مجموعة من الأطياف الخطية و التي يعتبر القالب الذي يكتب عليه خط الوحدة الخطية مثل الشكل التالي :

و في مرحلة التصنيف« تقارن صورة الوحدة الخطية المستقبلة مع كل

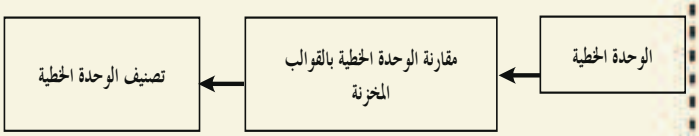


المشكل الرابع: االعملية التصنيفية للوحدة الخطية العربية

صنف من القوالب ، فإن كانت نتيجة مقارنتها مع الصنف س مثلا أكبر من نتيجة مقارنتها مع الصنف ص فإنها تصنف ضمن الصنف س و هكذا تصنف ضمن مصفوفة معينة لتحديد معايير المقارنة و تصحيح صورة الوحدة الخطية إذا كانت مشوهة مثل أصناف رسم الهمزة العربي التي تمثل كالآتي :

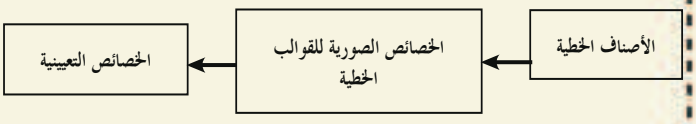
الصنف الأول :ء (ءِ ،ءَ ، ءُ ، ء)

الصنف الثاني :أ (ـأ ، ـإ ، آ ،لإ ،لأ ...)
الصنف الثالث :ؤ (ـؤ)
الصنف الرابع :ئ (ـئ ،ـئـ ، ئـ)»8
كما يوضح الشكل التالي :
ثم تـأتي مرحلة التوصيف أو التعيين، بحيث يوصف و يعين قالب الوحدة الخطية بواسطة مجموعة من الخصائص التعيينية، و التي من



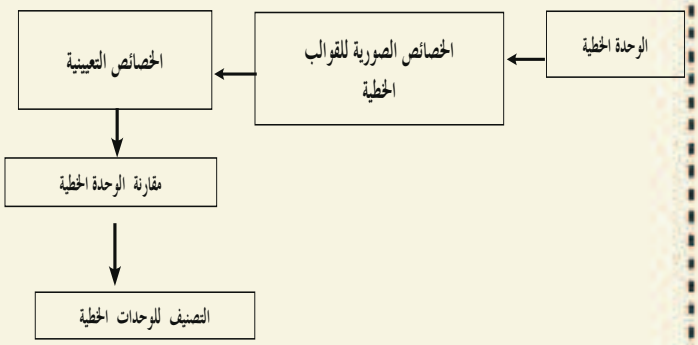
المشكل الخامس: المقارنة بين الوحدة الخطية والقوالب من أجل تصنيفها

الممكن أن نعبر عنها بقيم حقيقية، في مرحلة التعلم فيقدم كل نمط كمتجه من الخصائص كما يوضح الشكل التالي :
أما في مرحلة التعرف، أو التمييز: و التصنيف فهذه عادة تتم عن طريق تقسيم مساحة الوحدات الخطية إلى مناطق مجزأة، كل منطقة تقارن بالخصائص الموجودة فيها مع خصائص الصنف المخزنة من حيث



المشكل السادس: االعملية التعيينية للوحدة الخطية

الشكل، و الحجم ،والمحيط، و المساحة و تتم هذه المرحلة على الشكل التالي :



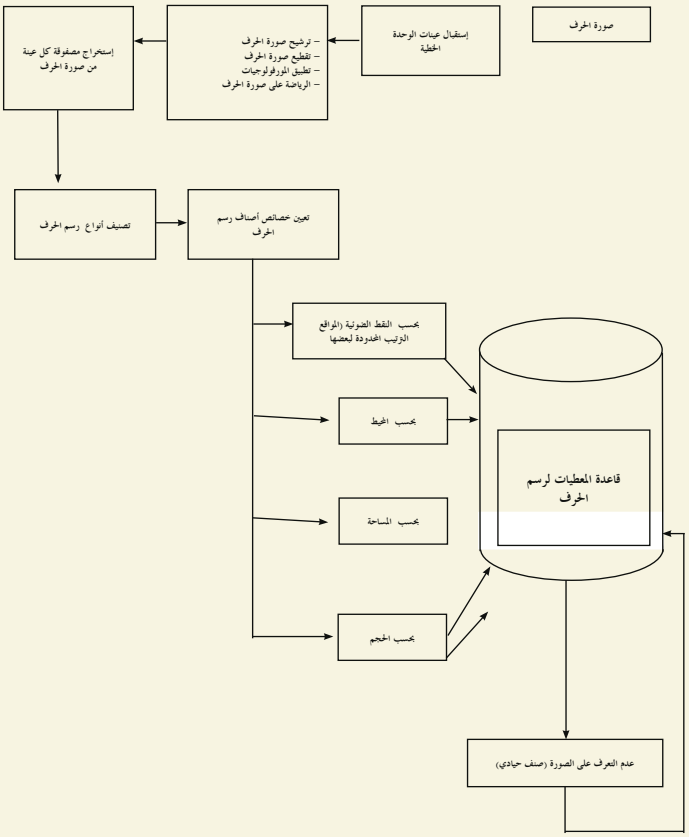
الشكل السابع: العملية التصنيفية والتعيينية للوحدة الخطية

– النظام الرقمي لمعالجة صورة الهمزة العربية :

يتكون النظام الرقمي لمعالجة الصورة بصفة عامة من الشكل التالي:

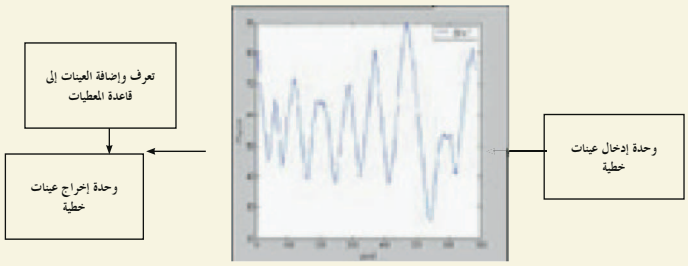
ا «نظام استقبال: يتم استقبال صورة الحرف من قبل الحاسوب عبر جهاز المسح الضوئي الذي يقل الصورة إلى الحاسوب من خلال الأشعة الضوئية التي تسهل عملية المعاينة الرقمية لمصفوفة الصورة، و تحديد حجمها
ا ذاكرة حفظ الصورة الرقمية: تخزن صورة الحرف بشكلها و مصفوفتها داخل ذاكرة الحاسوب لنقلها إلى مرحلة ما قبل المعالجة و تهيئتها لتعيين مكوناتها العامة
ا نظام التجزيء أو التقطيع: يسهل هذا النظام عملية تجزيئ صورة الحرف إلى عدة مناطق من أجل استخراج أهم الخصائص المكونة لرسم الوحدة الخطية من محيط و حجم ومساحة
ا نظام المورفلوجيات الرياضية: يسهل هذا النظام إعادة جمع أجزاء صورة الهمزة باستخراج مصفوفة أصلية لرسم الحرف نظام التعرف الآلي:

يتطلب هذا النظام قاعدة إدخال و إخراج الوحدات الخطية لإنتاج و توليد الصورة الخطية» 9تتم عملية النظام على الشكل التالي:



الشكل الثامن: مخطط التعرف الآلي على رسم الوحدة الخطية العربية

فبعد الحصول على عينات مختلفة لرسم الحرف ، أجرينا عملية المسح الضوئي على بعض العينات من أنواع رسم الوحدة الخطية ثم تخزينها في ذاكرة الحاسوب على الشكل الآتي:



وأخيرا إن هذه الدراسة تهدف إلى تطوير برنامج للتعامل مع الحروف العربية ، وهي تقنية رائدة تساعد في التعرف إلكترونيا على حروف اللغة العربية المكتوبة بخط اليد بهذا يمكن لهذا المحرك التعرف على النصوص المكتوبة بخط اليد وتحويلها إلى صيغة نصية مطبوعة.
ولتسهيل هذه العملية يجب إعداد برنامج يمثل المرجعية الخطية في مداخل هذه الخطوط بحيث يكون في كل اختبار للنصوص توسيع لمرجعية الخط تتمثل في قاعدة بيانات تحمل الضوابط اللغوية للوحدة الخطية. وإدخال أكبر عدد ممكن من الأشكال المختلفة التي تمكننا من تنويع الصفات التمييزية المعتمدة في التعرف الآلي.

البيبليوغرافيا

1. حركات مصطفى الكتابة والقراءة قضايا الخط العربي دار الآفاق، الأبيار، الجزائر.–
2. علي تعونيات، صعوبات تعلّم اللغة العربية المكتوبة، ديوان المجموعات الجامعية، د.ط
3. مسعودي فضيلة « التكرارية الصوتية في لبقرآآت القرآنية –قراءة نافع أنموذج – رسالة ماجستير قسم اللغة و الأدب العربي كلية العلوم الاجتماعية و الإنسانية 2002–2001
4– هاشم محمد الخطاط : قواعد الخط العربي (مجموعة خطية لأنواع الخط العربي)ط المزيدة 1400هـ 1980 م بغداد دار القلم مكتبة النهضة و ينظر كتاب – علي تعونيات، صعوبات تعلّم اللغة العربية المكتوبة، ديوان المجموعات الجامعية، د.ط

– 5 نوام تشو مسكي ترجمة حمزة بن قبلان المزيني اللغة ومشكلة المعرفة الرياض المريخ للنشر 1409ه/13/8 1989م/25/3 ص 45

–المراجع باللغة الأجنبية :

1–Hervé abdr « les reseaux de neurons» press univ .de
greonfle Normal net work –
––2 Hervé abdr « les reseaux de neurons» press univ .de greonfle Normal net work P 129

–3Hassane Chelayh :» analyse phonographématique de l'Arabe en vue d'application informatique». Thèse de doctorat présentée à'université de Paris VII Inria Sophia Antropolis – 123 ص

–4–james broesch Comprendre le traitement numérique de signal. Publitonic fgfg. Paris ISBN 28661109–8 OCT.008079