

التصنيف الآلي للسور القرآنية

عبد الباقي محمد شرف
قسم الحاسوب – كلية الهندسة
جامعة ليدز
ليدز – LS2 9JT
المملكة المتحدة
{a.m.sharaf08 , e.s.atwell}@leeds.ac.uk

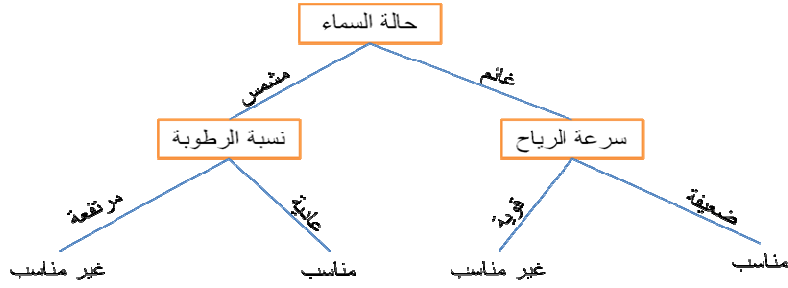
الخلاصة. التلقين الآلي هو مجال علمي يتيح للحاسوب اتخاذ قرار واكتشاف أنماط من قائمة طويلة جداً من البيانات بحيث يصعب التتقيب عن مثلها بالتحري البشري اليدوي. وفي هذه الدراسة أثبتنا كيفية الاستفادة من هذا العلم في دراسة خصائص السور وإجراء تصنيفات وتجميعات آلية لسور القرآن بناءً على خصائص حيوية تم تحديدها من أبحاث قرآنية. وبعد تحديد الخواص المعينة للتصنيف أجرينا التجارب باستخدام برنامج ويكا مفتوح المصدر وأوضحنا وشرحنا النتائج وناقشناها واستخلصنا فوائد وأهمية مثل هذه التجارب.

الكلمات الجوهرية: التلقين الآلي، التصنيف الآلي، معالجة اللغات الطبيعية، التتقيب النصي، الذكاء الاصطناعي.

١ مقدمة

تهدف خوارزميات التلقين الآلي إلى اكتشاف أنماط وتصانيف وتجميعات مثيرة للإهتمام في مجموعة بيانات مدخلة. وتتلخص طريقة عمل خوارزميات التصنيف الآلي في تلقينها أولاً بأمثلة مصنفة يدوياً تصنيفاً بشرياً، ومن ثم إختبار فعاليتها ومدى "تعلمها" من التلقين بطلب تصنيف أمثلة أخرى غير مصنفة.

وفيما يلي نوضح مفهوم التلقين الآلي من خلال المثال التالي. لنفترض أننا نريد تصميم برنامج للتلقين الآلي يمكنه الإدلاء برأيه إزاء مناسبة ممارسة رياضة ما في الخارج في يوم يتم تحديد بعض خواصه المناخية من الحرارة الخارجية ونسبة الرطوبة و سرعة الرياح وفيما إذا كانت السماء غائمة أو مشمسة. وحتى تتمكن خوارزميات التلقين الآلي من اتخاذ القرار الصائب يجب تدريبها أولاً من خلال أمثلة كثيرة لأيام ذات طقس مناخية مختلفة مع إعطاء القرار الصائب فيما إذا كان ممارسة الرياضة في ذلك اليوم مناسبة أو غير مناسبة. وتقوم خوارزميات التلقين الآلي بـ "التعلم" من هذه الأمثلة بطرق شتى ومنها بناء "شجرة قرارات" من هذه الأمثلة المحلولة ومثل هذه الشجرات تمكنها من إعطاء قرار لأي مثال مستقبلي. الشكل رقم (١) يعطي مثال مبسط لشجرة قرار عن ممارسة الرياضة في الخارج.



الشكل ١. شجرة قرار من قائمة بيانات الرياضة الخارجية وحالة الطقس

وكما هو ملاحظ فإنه يمكن فرز أي مثال عن طريق الهبوط من أعلى الشجرة لأدناها والتفرع حسب الخصائص المناخية في كل فرع.

وقد دخلت تطبيقات التلقين الآلي بنجاح في العديد من حلول الذكاء الإصطناعي مثل "الرؤية الحاسوبية" و معالجة اللغات الطبيعية ومحركات البحث و التعرف الآلي على الأنماط ومجالات علم الأحياء المعلوماتية. وفي جميع هذه التطبيقات تحتفظ حلول التلقين الآلي بنفس الهيكل العام من حيث وجود مجموعة بيانات تحتوي على أمثلة محلولة للتدريب مبنية فيها الخصائص المهمة للوصول للقرار مع تحديد القرار الصائب عن طريق خبير بشري، ومن ثم تدريب الخوارزميات من خلال هذه البيانات لتمكن من اتخاذ قرار لأمثلة مستقبلية.

والذي يخصنا في هذا البحث هو تطبيقات التلقين الآلي في مجال معالجة اللغات الطبيعية وفي هذا البحث حاولنا الإستفادة من هذا الهيكل العام لتطبيقات "التلقين الآلي" وتطويعه لحل مشكلة التصنيف الآلي للسور القرآنية إلي سور مكية ومدنية. وقد استفدنا من وجود أدوات متداولة لتطبيقات التعلم الآلي مثل أداة Weka¹ والتي تتيح تطبيق العديد من خوارزميات التلقين الآلي وعرض النتائج بعد إدخال مجموعة البيانات المعطاه.

وبسير نهج هذه الورقة العلمية على النحو التالي. في الفصل الثاني نعطي صورة إجمالية عن بعض التجارب السابقة في مجال تطبيقات المعالجة الحاسوبية للنص القرآني. أما الفصل الثالث فسنفصل فيه كيفية معالجة تصنيف السور القرآنية آلياً من قبل خوارزميات التلقين الآلي ونبين فيه الخصائص اللغوية والأسلوبية التي اعتمدنا عليها في تصنيف السور القرآنية إلى مكية ومدنية. وفي الفصل الرابع سرد لخطوات إجراء التجربة باستخدام أداة Weka المفتوحة المصدر. وفي الفصل الخامس نناقش نتائج التجربة وطرق تحسين الأداء وتطبيقات مستقبلية لحل مسائل أخرى على النص القرآني. نختم هذه الورقة بخاتمة نتطلع فيها لأهم الدروس المستفادة من هذه التجربة والآمال

¹ موقع الأداة هو <http://www.cs.waikato.ac.nz/ml/weka/>

التي يمكن أن نعدها على تقنيات الحاسوب لمعالجة اللغة العربية الفصحى بشكل عام والنص القرآني بشكل خاص.

٢ الأعمال المتصلة

اكتسبت ابحاث التصنيف النصي المزيد من الاهتمام في الآونة الأخيرة نظراً لتزايد المحتوى النصي الرقمي عبر الإنترنت وغيرها. وفيما كانت غالبية الحلول في ثمانينيات القرن الماضي للتصنيف الآلي تعتمد على ايجاد مجموعة كبيرة من قوانين يدوية تساعد في التصنيف، فإن الإتجاه تغير في العقد الأخير من القرن الماضي وبدأ الإقبال على حلول التلقين الآلي. وهذا الحل الأخير وفر الكثير من المجهود البشري خصوصاً مع تزايد كمية النص المراد تصنيفه. وفي الآونة الأخيرة توسع نطاق التصنيف الآلي ليشمل "التنقيب النصي" والذي يعنى باستخلاص معلومات وأنماط مهمة تلقائياً من النص. ويمكن معرفة تفاصيل عمل دوال التصنيف النصي بالرجوع إلى [سيبستاني ٢٠٠٢].

الفهرسة الآلية هي من أوائل تطبيقات التصنيف الآلي وهي متداولة في أبحاث استرجاع المعلومات [زيراس و هارتمان ١٩٩٣]. واستخدم [لاركي ١٩٩٩] التصنيف الآلي لفرز الوثائق إلى مجموعة مواضيع محددة مسبقاً. ومن تطبيقات التصنيف الآلي ترشيح النص حيث يرشح النص بغية التقاط المواد المرغوبة فقط وعزل المواد غير المرغوبة ومن أمثلة ذلك ترشيح البريد الإلكتروني وعزل الرسائل الدعائية [أندروتسوبولوس ومن معه ٢٠٠٠]. وفي الآونة الأخيرة ومع تزايد الإقبال على محتوى الشبكة العنكبوتية تركزت أبحاث عديدة في التصنيف الآلي لصفحة عنكبوتية إلى موضوعات متدرجة محددة [يانغ ومن معه ٢٠٠٢]. ومن التطبيقات المشهورة للتصنيف الآلي فك اللبس عن كلمة غامضة فمثلاً كلمة "أمة" تأتي بمعان عديدة في القرآن فهي تعني الجماعة من الناس في الآية "كنتم خير أمة أخرجت للناس" ولكنها تعني القدوة والقائد في قوله تعالى "إن ابراهيم كان أمة قانتاً لله حنيفاً". أما في قوله تعالى "ولإن أخرجنا عنهم العذاب إلى أمة معدودة" فهي تعنى البرهه من الزمن. ويمكن صياغة هذه المشكلة في قالب التلقين الآلي وبالتحديد التصنيف الآلي ومن ثم معالجتها [اسكوديرو ومن معه ٢٠٠٠].

فيما يخص التطبيقات السابقة للتلقين الآلي على النص القرآني فهي نادرة ونورد هنا بحثان توصلنا إليهما بعد بحث طويل وهما دراسة نجلاء ثابت [ثابت ٢٠٠٥] وهيرمان مويسل [مويسل ٢٠٠٩] وفيما يلي شيء من التفصيل عن هاذين الباحثين.

عمدت ثابت إلى إنشاء مصفوفة ضخمة مكونة من ١١٤ صف تمثل سور القرآن، و٣٦٧٢ عمود تمثل مفردات القرآن الكريم. وهذا العدد من المفردات تم اختياره بعد فصل الكلمات الغير دلالية من حروف وأدوات مختلفة وإزالة التكرار وكذلك إزالة كل كلمة وردت مرة واحدة في النص القرآني. وكذلك استخدمت أداة لتجذير الكلمات واستخراج مشتقات الكلمة الواحدة. وتوصلت بعد ذلك لمصفوفة توضح عدد مرات تكرار كل كلمة في سور القرآن الكريم. ولكنها واجهت مشكلة خلو السور القصيرة من غالبية هذه الكلمات فلجأت إلى تقليص المصفوفة لتضم ٢١ أطول سورة من القرآن فقط والتي تحوى على أكثر من ألف كلمة. وانتهت بعد ذلك إلى إصدار آلي لتجميعات

تدرجية توضح مدى تماثل سور القرآن والمسافة بينها قرباً أو بعداً. أما موصل فلجأ إلي تحديد تسع كلمات قرآنية اعتبرها كلمات ممثلة تعطي معياراً لدراسة سور القرآن ومن ثم طبّق مفاهيم معقدة من علم الإحصاء التوزيعي لتكوين تجميعات تدرجية تضم هذه السور. وهذه الكلمات التسع هي: الله، لا، رب، قال، كان، يوم، ناس، يومئذ، شر. وليس واضحاً المغزي من اختيار هذه بالذات.

٣ تصنيف السور القرآنية

حرصنا في هذه التجارب على دراسة تصنيف السور المكية والمدنية لسببين رئيسيين: أولاً لأنها حظيت باهتمام من قبل العلماء وافرد هذا الباب بالتصنيف قديماً وحديثاً مثل كتاب "المكي والمدني" لمكي بن أبي طالب القيسي المتوفي عام ٤٣٧ هـ وغيره. وقد ذكروا فوائد عديدة لمعرفة المكي والمدني منها معرفة النسخ والمنسوخ وفهم الآيات بعمق أكثر بمعرفة ظروف النزول وتأمل سيرة الرسول في تطور دعوته واختلاف الخطاب الديني بين المرحلتين، وكذلك معرفة الحكمة في تدرج كثير من أحكام الشريعة [رباني ٢٠٠٤]. والسبب الآخر هو قابلية هذا الموضوع ومناسبته لأن يصاغ على شكل مشكلة تحل عن طريق التصنيف الآلي وذلك أنه لتطبيق التصنيف الآلي لابد من توفر مجموعة من الخواص ونتيجة نهائية للتصنيف وقائمة بيانات لتدريب خوارزميات التلقين الآلي، وهذه الإعدادات يمكن إيجادها لهذه المسألة التي بأيدينا.

وعلى ذلك يجب أولاً إيجاد مجموعة من الخصائص التي تميز الآيات المكية والمدنية، وإعداد قائمة بيانات تكون بمثابة أمثلة لتدريب خوارزميات التلقين الآلي. وبما أن هدف هذا البحث هو تطبيق تقنيات التلقين الآلي فإننا لا ندخل في حوض غمار تفاصيل اختلاف علماء السور المكية والمدنية وإنما نعتد في هذه الدراسة على الخواص التي ذكرت في [رباني ٢٠٠٤] وعدّلنا هذه الخواص بما يتناسب مع امكانيات تعدادها بشكل آلي. والجدول المبين في الشكل رقم (٢) في الصفحة التالية يلخص تصنيف سور القرآن مع بيان السور المختلف عليها.

٣.١ خصائص السور المكية والمدنية

هناك خصائص وضوابط عديدة ذكرها العلماء للتمييز بين السور المكية والمدنية، وقد اخترنا منها بعض تلك التي نستطيع احصائها بطرق شبه آلية والتي تتناسب مع طبيعة تجارب التلقين الآلي. وفيما يلي سرد موجز لهذه الخصائص:

١. كل سورة فيها سجدة فهي مكية، ويستثنى من ذلك سورتا الرعد والحج.
٢. كل سورة فيها كلمة "كلا" فهي مكية.
٣. كل سورة فيها عبارة "يا أيها الناس" فهي مكية، إلا سورة الحج.
٤. كل سورة فيها عبارة "يا أيها اللذين آمنوا" فهي مدنية.
٥. كل سورة في أولها حروف الهجاء ك "الم" ، و "الر" ، ونحو ذلك فهي مكية سوى البقرة وآل عمران والرعد.

٦. كل سورة فيها قصة آدم وإبليس فهي مكية سوى البقرة.
٧. كل سورة فيها قصص الأنبياء والأمم السابقة فهي مكية سوى البقرة.
٨. تتميز السور المكية بكثرة استخدام أدوات التأكيد والتقريب والردع والزجر وغيرها.
٩. تتميز السور المكية بقصر آياتها.
١٠. تتميز السور المكية بالإكثار من ذكر الآخرة وأهوال يوم القيامة والجنة والنار.
١١. تتميز السور المدنية بذكر الجهاد في سبيل الله.
١٢. تتميز السور المدنية بكثرة سرد أحكام المعاملات من نكاح وبيع وطلاق ورضاع وغيرها.
١٣. تتميز السور المدنية بخطاب أهل الكتاب ومناقشتهم.
١٤. تتميز السور المدنية بذكر أركان الإسلام العملية من صلاة وزكاة وصيام وحج.

No. Category	No. Category	No. Category	No. Category	No. Category	No. Category	No. Category	No. Category
1	K*	21	K	41	K	61	D*
2	D	22	D*	42	K	62	D
3	D	23	K	43	K	63	D
4	D	24	D	44	K	64	D*
5	D	25	K	45	K	65	D
6	K	26	K	46	K	66	D
7	K	27	K	47	D*	67	K
8	D	28	K	48	D	68	K
9	D	29	K*	49	D	69	K
10	K	30	K	50	K	70	K
11	K	31	K	51	K	71	K
12	K	32	K	52	K	72	K
13	D*	33	D	53	K	73	K
14	K	34	K	54	K	74	K
15	K	35	K	55	D*	75	K
16	K*	36	K	56	K	76	D*
17	K	37	K	57	D	77	K
18	K	38	K	58	D	78	K
19	K	39	K	59	D	79	K
20	K	40	K	60	D	80	K*
						81	K
						82	K
						83	K*
						84	K
						85	K
						86	K
						87	K
						88	K
						89	K*
						90	K
						91	K
						92	K*
						93	K
						94	K
						95	K
						96	K
						97	K*
						98	D*
						99	D*
						100	K
						101	K
						102	K
						103	K
						104	K
						105	K
						106	K
						107	K
						108	K
						109	K
						110	D*
						111	K
						112	K*
						113	K*
						114	K*

الشكل رقم (٢): تصنيف سور القرآن الكريم بأرقامها إلى مكية (K) ومدنية (D) مع الإشارة إلى المختلف فيها بعلامة (*)

وبعد تحديد الخصائص المستخدمة يجب علينا إحصاء وتعداد هذه الخصائص في القرآن الكريم، وهنا لا بد من الاستعانة بأدوات حاسوبية وذخائر معنونة للقرآن لإحصاء هذه الخصائص حيث أنه يتبين لنا أن أدوات البحث العادية لا تفي وحدها بالغرض وإنما يستلزم وجود ذخائر معنونة تعنى بتوسيم النص القرآني بأجزاء الكلام، وهنا تم الإستعانة بـ "الذخيرة القرآنية العربية" (Quranic Arabic Corpus) وهي ذخيرة متوفرة مجاناً [ديوكس وحيش ٢٠١٠] ونختصر هذا المصدر في المرات القادمة بكلمة "ذخيرة". كما تم الإستعانة لبحث مفردات القرآن بكتاب "المعجم المفهرس لألفاظ القرآن الكريم" [عبد الباقي ١٩٤٣] والمتوفر أيضاً بالموقع الإلكتروني لمجمع الملك فهد لطباعة القرآن الكريم ونختصره بكلمة "المعجم" في المرات القادمة.

٣.٢ إحصاء الخصائص من النص القرآني

بالاستعانة بالذخائر والأدوات آفة الذكر نقوم بإحصاء مرات تكرار هذه الخصائص في كل سورة لإنشاء قائمة بيانات قبل اعتمادها في تجارب التلقين الآلي. وفيما يلي نعيد ذكر كل خاصية من الخصائص الأربعة عشر الواردة في الفقرة الماضية ولكن نركز على بيان طريقة تمثيلنا واختزالنا لهذه الخاصية حتى تكون قابلة للتعداد. كما نورد ذكر بعض الصعوبات والتحديات عند تعداد كل خاصية.

١. آيات السجود: تجاوزنا في تعدادنا عن آيات السجود من تعداد سجديات القرآن المعروفة إلى تعداد جذور "سجد" في القرآن والبالغة ٩٢ مرة وذلك باستخدام الذخيرة ولكن هذا يعني دخول بعض الآيات المدنية مثل كلمة "المسجد الحرام" الواردة مرات عدة في سورة البقرة المدنية.
٢. كلمة الردع "كلا": وردت ٣٣ مرة في القرآن وكلها مكية.
٣. عبارة "ياأيها الناس": بالبحث عن هذه العبارة وجدناها تتكرر ٢٠ مرة نصفها في السور المدنية، وهذه الحقيقة يجعلها خاصية غير مفيدة للتصنيف وهذا جعلنا نستبعدنا من قائمة الخصائص النهائية.
٤. عبارة "ياأيها اللذين آمنوا": وهذا العبارة تتكرر ٩٠ مرة كلها مدنية ولذلك اعتمدناها ضمن الخصائص بخلاف العبارة الماضية.
٥. الحروف المقطعة: وردت ٣٠ مرة في القرآن ويمكن تتبع ذلك بوسم (INL) في الذخيرة. كما ذكرنا فإن سور البقرة، آل عمران والرعد تمثل إستثناءات لهذه الخاصية.
٦. قصة آدم وحواء وإبليس: وجدنا هذه القصة تتكرر خمس مرات في القرآن كلها مكية ما عدا سورة البقرة. وقد اعتمدنا في التعداد على البحث المتقدم للموقع الإلكتروني للمجمع والذي يتيح البحث بالمسافة بين الكلمتين (وهما "آدم" و"إبليس" في هذا المثال).
٧. قصص الأنبياء: عثرنا على ٥٨١ مرة تكرار ذكر الأنبياء في القرآن، واستعنا بالبحث في الذخيرة عن أسماء الأنبياء بعزل الزوائد. وتم البحث عن ٢١ من الأنبياء كما يلي: إبراهيم، اسماعيل، يعقوب، إسحاق، موسى، عيسى، داوود، نوح، زكريا، يحيى، يونس، هارون، سليمان، يوسف، إلياس، اليسع، لوط، صالح، هود، آدم، أيوب، وإدريس.
٨. حروف الردع والزجر والتأكيد: وهنا استعنا في البحث كذلك على توسيم الحروف في الذخيرة. واستخدمنا وسم (AVR) لحروف الردع مثل "كلا"، ووسم (CERT) لأدوات التأكيد مثل "قد"، ووسم (SUP) لحروف الفجاءة مثل "إذ"، ووسم (EXH) لأدوات التمني مثل "لولا". وعثرنا بهذه الطريقة على عدد ١٤٧٨ من هذه الحروف والغالبية من السور كانت بالفعل مكية.
٩. معدل طول الآيات: لمعرفة معدل حجم الآيات قمنا بتنزيل نص القرآن من موقع "تنزيل" ثم صياغة برنامج لعد عدد الكلمات لكل سورة ومن ثم تقسيم عدد الكلمات على عدد الآيات لمعرفة معدل طول الآيات لكل سورة. وبهذه الطريقة وجدنا معدل الطول العام لآيات القرآن هو عشر كلمات، وتقل لثمان كلمات للسور المكية وتتضاعف لـ ١٦ كلمة للسور المدنية.

١٠. مواضيع الآخرة: كمؤشر لهذا الموضوع اخترنا سبع كلمات دالة مع ملحقاتها وهي: جهنم، الجنة، النار، سعير، قيامة، عذاب، و آخرة. وعثرنا على ٨٢٢ موضع في الذخيرة. ولكن هناك مواضع في السور المدنية أيضاً تشمل بعض هذه الكلمات.
١١. آيات الجهاد: وهذه ميزة للسور المدنية. واستعنا بالذخيرة لهذا الغرض وبحثنا عن كلمتي "جهاد" و "قتل" بجذورهما، وعثرنا على عدد ٢١١ مرة. ووجدنا حالات استثنائية تعكر على المغزي من هذا البحث، مثل ورود جذر الجهاد بمعنى غير القتال كما في سورة الأنعام المكية "وأقسموا بالله جهد أيمانهم" أو الجهاد الفكري كما في سورة الفرقان "وجاهدكم به [اي بالقرآن] جهاداً كبيراً" ، وكما في أمثلة قتل المشركين لأولادهم في سورة الأنعام أيضاً، أو قتل فرعون لأولاد بني إسرائيل في مواضع من سور مكية.
١٢. آيات النكاح والطلاق: وهي ميزة للسور المدنية، واخترنا لتمثيلها بجذور كلمتي "نكح" و "طلق"، ولم ندخل في البحث جذور "زوج" لدخول آيات كثيرة غير دالة على المعنى المراد هنا. وبهذه الطريقة عثرنا على ٣٧ موضع في الذخيرة كلها في صميم البحث المراد ما عدا ثلاث مواضع تكرر لكلمة "انطلقوا" من سورة المرسلات المكية.
١٣. التحدث عن "أهل الكتاب": واخترنا خمس كلمات دالة على هذه الخاصية وهي: أهل الكتاب، التوراة، الإنجيل، اليهود، النصارى. وبهذه الطريقة عثرنا على عدد ٦٣ موضع في الذخيرة كلها مدنية ما عدا موضعين في سورتي العنكبوت والأعراف.
١٤. ذكر أركان الإسلام العملية: واخترنا الكلمات الدالة مع مشتقاتها كما يلي: الصلاة، الزكاة، الصوم، والحج. ووجدنا أن هذه الكلمات ليست حكرأ على سور مدنية وإن كانت الغالبية. فمثلاً نجد سورة الأنعام المكية ذكرت "الصلاة" في موضعين. وكذلك ذكرت سورة الأعراف الزكاة في معرض الحديث عن قصة موسى.

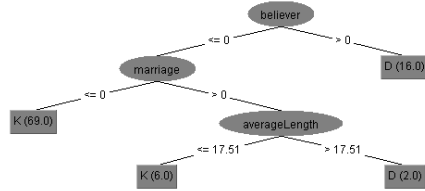
٤ إجراء تجارب التصنيف الآلي باستخدام weka

يتيح برنامج ويكا المفتوح المصدر للباحثين في مجال التلقين الآلي بإجراء تجاربهم على قائمة البيانات المدخلة بعد تعيين خصائص التصنيف [ويتن وفرانك ٢٠٠٥]. وهكذا يريح ويكا المستخدم من تفاصيل معقدة بتوفير شاشة واجهة مبسطة. وكل ما على المستخدم هو توفير ملف البيانات كمدخل لويكا. وبعد إدخال هذا الملف لويكا يمكن عرض توزيع هذه الخصائص بطرق شتى. وتوفر ويكا وظائف أخرى عديدة مثل مقارنة خاصيتين حسب عدد التكرار والتوزيع إلى الفئتين المكي والمدني وغير ذلك.

٤.١ التصنيف الآلي

يجب أولاً إنشاء ملفين ويكا من نوع arff أحدهما يمثل البيانات المحلولة لتدريب وتلقين خوارزميات التصنيف الآلي، والملف الثاني مجموعة من بيانات الإختبار لمعرفة كفاءة الخوارزميات لتصنيف الأمثلة غير المحلولة. وفي تجربتنا اعتمدنا في تحديد ملف التدريب على ٩٣ مثال للسور المتفق عليها بين العلماء في كونها مكية أو مدنية. أما ملف الإختبار فجعلناها ٢١

مثال من أمثلة السور المختلف عليها كما يتضح ذلك في الجدول المدرج في الشكل رقم (٢). وبعد ذلك تم تلقين خوارزم للتلقين الآلي يسمى J48 والذي قام بإنشاء شجرة قرار من هذه الأمثلة جاءت كما هو مبين في الشكل رقم (٣) أدناه.



الشكل رقم (٣): شجرة قرار منتجة من قبل خوارزم J48

وتم بعد ذلك اختبار كفاءة التصنيف الآلي لتمرير الأمثلة الـ ٢١ لسور مختلف عليها بين العلماء وجاءت النتيجة كما هو موضح في جدول الشكل رقم (٤).

نجد مثلاً في هذا الجدول أن التصنيف الآلي اعتبر السورة رقم ١٣ وهي سورة الرعد مكية ولهذا الاختيار دافع قوي كونها تستفتح بحروف مقطعة وتحوي آية سجدة وإن كانت طول حجم آياتها تجذبها لفئة السور المدنية.

sura no.	kalla	prostrate	believer	initials	prophets	adam-iblis	emphasis	length	eschatology	jihad	marriage	other-religion	pillars	category	prediction
1	0	0	0	NO	0	NO	0	4.1	0	0	0	0	0	K	K
13	0	1	0	YES	0	NO	16	19.84	9	0	0	0	1	D	K
16	0	2	0	NO	4	NO	41	14.41	25	2	0	0	0	K	K
22	0	5	1	NO	6	NO	21	16.4	17	4	0	1	7	D	D
29	0	0	0	YES	13	NO	46	14.17	17	5	0	1	2	K	K
47	0	0	2	NO	0	NO	7	14.26	4	3	0	0	0	D	D
55	0	1	0	NO	0	NO	0	4.51	6	0	0	0	0	D	K
61	0	0	3	NO	3	NO	1	16.14	1	2	0	1	0	D	D
64	0	0	1	NO	1	NO	2	13.44	2	0	0	0	0	D	D
76	0	1	0	NO	0	NO	0	7.84	3	0	0	0	0	D	K
80	2	0	0	NO	0	NO	2	3.17	0	1	0	0	0	K	K
83	4	0	0	NO	0	NO	10	4.69	0	0	0	0	0	K	K
89	2	0	0	NO	0	NO	3	4.63	4	0	0	0	0	K	K
92	0	0	0	NO	0	NO	4	3.38	2	0	0	0	0	K	K
97	0	0	0	NO	0	NO	0	6	0	0	0	0	0	K	K
98	0	0	0	NO	0	NO	2	11.75	2	0	0	2	2	D	K
99	0	0	0	NO	0	NO	0	4.5	0	0	0	0	0	D	K
110	0	0	0	NO	0	NO	0	6.33	0	0	0	0	0	D	K
112	0	0	0	NO	0	NO	0	3.75	0	0	0	0	0	K	K
113	0	0	0	NO	0	NO	0	4.6	0	0	0	0	0	K	K
114	0	0	0	NO	0	NO	0	3.33	0	0	0	0	0	K	K

الشكل رقم (٤): مقارنة نتائج التصنيف الآلي لـ ٢١ سورة وحالات الاختلاف الستة مظلمة

٤.٢ التجميع الآلي

يتيح ويكا تكوين تجميعات آلية ذات خصائص مشتركة بناء على البيانات المدخلة وتعتمد على خوارزميات للتجميع تستخدم المسافة الأوكليدية بين بؤر التجميع [ديزا وديزا ٢٠٠٩]. والشكل رقم (٥) يوضح هذه التجميعات الثمانية المنتجة آلياً عن طريق ويكا.

1	2	3	4	5	6	7	8
2 D	6 K	15 K	74 K	8 D	1 K	31 K	3 D
7 K	10 K	19 K	83 K	24 D	50 K	35 K	4 D
	11 K	21 K	96 K	33 D	53 K	36 K	5 D
	12 K	23 K			55 D	44 K	9 D
	13 D	26 K			56 K	45 K	
	14 K	30 K			69 K	46 K	
	16 K	32 K			72 K	47 D	
	17 K	34 K			76 D	49 D	
	18 K	37 K			77 K	51 K	
	20 K	43 K			79 K	52 K	
	22 D	54 K			84 K	57 D	
	25 K	68 K			86 K	58 D	
	27 K				88 K	59 D	
	28 K				90 K	60 D	
	29 K				91 K	61 D	
	38 K				92 K	62 D	
	39 K				93 K	63 D	
	40 K				94 K	64 D	
	41 K				95 K	65 D	
	42 K				97 K	66 D	
	48 D				99 D	67 K	
					100 K	70 K	
					101 K	71 K	
					103 K	73 K	
					105 K	75 K	
					106 K	78 K	
					108 K	80 K	
					109 K	81 K	
					110 D	82 K	
					111 K	85 K	
					112 K	87 K	
					113 K	89 K	
					114 K	98 D	
					102 K		

الشكل رقم (٥): تجميعات آية لسور القرآن الكريم

نلاحظ أن غالبية التجميعات تنحاز لفئة معينة من مكية ومدنية و التجميعات أرقام ٣ و ٤ و ٥ و ٨ تحتوي على سور متجانسة بينما نجد في التجميع رقم ٧ خليط من الفئتين. ونستطيع تفسير التجميع رقم (١) وكونها تحوي سورتي البقرة والأعراف حيث أنهما تشتركان في خصائص عديدة مثل تفاصيل قصة موسى وبنو إسرائيل وكونهما من السور الطوال المستفتح بحروف متقطعة وغير ذلك.

٥ تقييم النتائج

ندرك أن هناك مجموعة من العوامل قد ساهمت في تحديد كفاءة النتائج أعلاه. منها أن تصنيفنا كان على مستوى السور وليس على مستوى الآيات وهذا هو مصدر اختلاف العلماء في تصنيف بعض السور مثل سورة الرعد التي تتجاذبها خواص السور المكية والمدنية معاً مما يوحي بأن بعض آيات هذه السورة نزلت قبل الهجرة والبعض الآخر بالمدينة. ولكن تعريف الخواص على مستوى الآيات يجعلنا نواجه مشكلة عدم توفر الخصائص في غالبية الآيات مما يحد من قدرات خوارزميات التلقين الآلي.

ومن مصادر ورود الخطأ في البيانات هو اختزال الخاصية في مجموعة من الكلمات الدالة وهو اختيار اجتهادي قابل للمناقشة، ولاحظنا أن الفعل "انطلق" دخل ضمن بحثنا عن آيات الطلاق.

وبشكل عام أظهرت دراستنا هذه أهمية التلقين الآلي في الدراسات القرآنية فمثلاً عند مناقشة الجدول في الشكل رقم (٤) نجد أن التصنيف الآلي جعل سورة الرحمن في قسم السور المكية وإن كانت مصحف المدينة تضعها من ضمن السور المدنية. وفي هذه السورة يقول السيوطي كما نقل

في [رباني ٢٠٠٤] " قال السيوطي: الجمهور على أنها مكية، وهو الصواب وتتميز بمزايا السور المكية أسلوباً وموضوعاً." وبالتالي فإننا نؤمن أن الباحث في الدراسات القرآنية قد يفاجأ باكتشاف بعض الأنماط والعلاقات المثيرة للاهتمام وهو بكونه باحثاً في هذا المجال يستطيع انتقاء ما هو صواب والتغاضي عن ما هو خطأ. وهذه منهجية نريد إثباتها هنا وهي أن نتائج التصنيف الآلي أو أية مخرجات آلية من خوارزميات الحاسوب لا بد أن تمرر على الخبير البشري قبل اعتمادها خصوصاً إذا كانت البيانات المدخلة تمثل نصوصاً مقدسة وحساسة كالنص القرآني. والله من وراء القصد.

٦ الخاتمة

نرجوا أن نكون قد وفقنا في هذه الدراسة لتقريب كيفية الاستفادة من مجال التلقين الآلي في الدراسات القرآنية. والتجربة التي أجريناها هنا في تصنيف سور القرآن هي تجربة هدفها تبسيط المسألة وتقريب المفاهيم لكونها سهلة الصياغة في قالب تجارب ويكا المشروحة في الفصل الرابع، وإن كانت نتائجها لا تكون بالضرورة مثار إعجاب باحث الدراسات القرآنية. ولكن بتتبع نفس الخطوات يمكن صياغة تجارب أكثر تعقيداً وفائدة من هذه التي أجريناها هنا. ونقوم الآن بإجراء أبحاث بجامعة ليدز عن كيفية الاستفادة من التلقين الآلي في الربط التلقائي بين الآيات المتشابهة في القرآن الكريم. ومن المجالات التي يمكن تطبيق التلقين الآلي هي علم القراءات بحيث يستطيع الباحث في علوم القراءات بتحديد خصائص حيوية وصياغة ملف بيانات ومن ثم اكتشاف أنماط مخفية وأسرار مثيرة للاهتمام في ربط قراءات محددة مع خصائص لغوية وأسلوبية.

المراجع

[اسكوديرو من معه ٢٠٠٠]

Escudero, G., Marquez, L. and Rigau, G. 2000. Boosting applied to word sense disambiguation. In Proceedings of ECML-00, 11th European Conference on machine learning (Barcelona, Spain, 2000), 129 – 141.

[أندروتسوبولوس ومن معه ٢٠٠٠]

Androutsopoulos, I., Koutsias, J., Chandrinou, K. V., and Spyropoulos, C. D. 2000. An experimental comparison of naive Bayesian and keywordbased anti-spam filtering with personal e-mail messages. In Proceedings of SIGIR-00, 23rd ACM International Conference on Research and Development in Information Retrieval (Athens, Greece, 2000), 160–167.

[ثابت ٢٠٠٥]

N. Thabet. 2005. Understanding the thematic structure of the Qur'an: an exploratory multivariate approach. Proceedings of the annual meeting of Association of Computational Linguistics. Pp. 7 – 12.

[ديزا وديزا ٢٠٠٩]

Elena Deza & Michel Marie Deza (2009) Encyclopedia of Distances, page 94, Springer.

[ديوكس وحبش ٢٠١٠]

Kais Dukes and Nizar Habash. 2010. Morphological Annotation of Quranic Arabic. Language Resources and Evaluation Conference (LREC 2010). Valletta, Malta.

[رباني ٢٠٠٤]

محمد شفاعت رباني. المكي والمدني. مجمع الملك فهد لكبابة القرآن الكريم. البحث متوفر على الرابط الآتي:

<http://www.qurancomplex.org/Tree.asp?section=2&TabID=2&SubItemID=1&l=arb&SecOrder=2&SubSecOrder=1>

[زيراس و هارتمان ١٩٩٣]

Tzeras, K. and Hartmann, S. 1993. Automatic indexing based on Bayesian inference networks. In Proceedings of SIGIR-93, 16th ACM International Conference on Research and Development in Information Retrieval (Pittsburgh, PA, 1993), 22-34.

[سيبستاني ٢٠٠٢]

Fabrizio Sebastiani. Machine Learning in Automated Text Categorization, in ACM Computing Surveys, Vol. 34, No. 1, March 2002, pp. 1-47.

[عبد الباقي ١٩٤٣]

محمد فؤاد عبد الباقي. ١٩٤٣. المعجم المفهرس لألفاظ القرآن الكريم. دار الكتب المصرية. القاهرة

[لاركي ١٩٩٩]

Larkey, L.S. 1999. A patent search and classification system. In Proceedings of DL-99, 4th ACM Conference on Digital Libraries (Berkeley, CA, 1999), 179-187.

[مويصل ٢٠٠٩]

H. Moisl, 2009. Sura Length and lexical probability estimation in cluster analysis of the Qur'an. ACM Transactions on Asian language information processing (TALIP) 8, 4 (Dec. 2009), 1-19.

[ويتن وفرانك ٢٠٠٥]

Witten, Ian and Frank, Eibe. 2005. Data Mining, 2nd Edition. (Morgan Kaufmann).

[يانغ ومن معه ٢٠٠٢]

Yang, Y., Slattery, S., and Ghani, R. 2002. A study of approaches to hypertext categorization. Journal of Intelligent Information Systems. 18, 2/3 (March-May), 219-241.

المخلص باللغة الانجليزية

Title: Automatic Categorization of the Qur'anic Chapters

Abstract. Machine learning is a scientific discipline that allows computers to make decision and find patterns from a huge amount of data which would be very tedious for manual inspection. In this study we have shown how to make this discipline useful for studying characteristics of the Qur'anic chapters and performing automatic classification and clustering based on pre-defined attributes identified by Quranic scholars. We have performed our experiments using open source tool called Weka, where we interpreted the results.

Keywords: Machine Learning, Text Categorization, Natural Language Processing, Text Mining, Artificial Intelligence.

المصطلحات

التلقين الآلي	Machine learning
تجميعات	clusters
شجرة قرارات	Decision trees
الرؤية الحاسوبية	Computer vision
معالجة اللغات الطبيعية	Natural language processing
محركات البحث	Search engines
الأنماط	patterns
علم الأحياء المعلوماتية	Bio informatics
تلقين آلي مشرف عليه	Supervised learning
تلقين آلي غير مشرف عليه	Un-supervised learning
استرجاع المعلومات	Information retrieval
ترشيح النص	Text filtering
الرسائل الدعائية	Spam messages
فك اللبس	Word sense disambiguation
علم الإحصاء التوزيعي	Statistical sampling distribution
المسافة الأوكليدية	Euclidean distance
بؤر التجميع	Cluster centroid