An-Najah University Journal for Research – B
Humanities

Language Processing: The Precedence of Neural Networks on the Account of Hidden Markov Models

Dia Eddin AbuZeina[1,*]
Received: 25th Jan. 2024, Accepted: 19th Jul. 2024, Published: ××××

Accepted Manuscript, In press

ABSTRACT: Background: since its discovery at the beginning of the last century, Markov models gain a great popularity, and have been widely used in different domains. However, the most prominent use was in computational linguistics, or what is known as natural language processing (NLP). Abstractly, Markov models are nothing but a statistical representation of a particular system. The mathematical statistical representation of a given system is the heart of Markov theory. Markov models characterized by solid mathematical representation, which significantly promotes using it. No doubt, Markov models are mainly used in prediction and classification, to serve computational linguistics as well as other scientific applications. They include two main types: observable Markov models (for short, known as Markov models), and hidden Markov models (HMMs). Bearing in mind that the HMMs are an extension of the Markov models, to model certain applications. **Aim:** In this paper, I present a brief description of HMMs, as well as some HMMs-based NLP applications. Then, I shed light the researchers' trend towards artificial neural networks (ANNs). **Methods:** the global digital libraries used to collect the topic related papers. The searching keywords focus on HMMs and ANNs. The collected papers filtered out to prepare the short list computational linguistics related papers. Then, the selected papers sorted based on published date to track the transformation process towards ANNs. **Results:** This study shows that the ANNs have been significantly and promising used in NLP tasks. Simultaneously, the recent NLP-related research shows that HMMs have less contribution for such tasks. That is, the ANNs (especially deep learning) is the de-fact trend in NLP field. **Recommendations:** the NLP research community and the new linguistic researchers should highly consider this transformation.

Keywords: Natural Language Processing, Computational Linguistics, Hidden Markov Models, Artificial Neural Networks, Machine Learning, Features, Classifiers.

[1] Faculty of Information Technology and Computer Engineering, Palestine Polytechnic University, Hebron, Palestine
*Corresponding author email: abuzeina@ppu.edu

مجلة جامعة النجاح للأبحاث – ب
العلوم الإنسانية

حوسبة اللغة: ما بين هيمنة الشبكات العصبية وتراجع نماذج ماركوف المخفية

ضياء الدين أبوزينة[1]

تاريخ التسليم: (2024/1/25)، تاريخ القبول: (2024/7/19)، تاريخ النشر: ××××

مخطوطة مقبولة، قيد النشر

الملخص: **خلفية الدراسة:** منذ اكتشافها في بدايات القرن الماضي، حظيت نماذج (ماركوف) على قبول واستحسان الباحثين، واستخدمت في مجالات متعددة، وكان لها حظ وافر ونصيب بارز من الأبحاث المتعلقة بحوسبة اللغة، أو ما يعرف (بـ معالجة اللغات الطبيعية). يُعد التمثيل الإحصائي الرياضي للعمليات العشوائية الأساس الذي بنيت عليه نظرية ماركوف إذ تعتمد على أساس رياضي متين، منحها شهرة واسعة، وجعلها محط أنظار الباحثين لفترة طويلة، الأمر الذي أحدث نقلة نوعية في عمليات التنبؤ والتصنيف، وفتح المجال لحوسبة اللغة وغيرها من العلوم، مثل الرعاية الصحية، والتسويق، وأسواق المال والاعمال، والتنبؤات الجوية الخ. وتشمل نماذج ماركوف نوعين رئيسين: نماذج ماركوف الظاهرة (اختصاراً يطلق عليها نماذج ماركوف)، ونماذج ماركوف المخفية. ورغم أن النوع الأول أفاد في مجالات بحثية كثيرة، إلا أنه لم يلبّ جميع الاحتياجات، فكانت نماذج ماركوف المخفية امتدادا لها، وتلبية لحاجة شريحة أوسع وأنماط مختلفة من المسائل البحثية. إلا أن السنوات الأخيرة شهدت توجهاً كبيراً نحو استخدام الشبكات العصبية الاصطناعية سواء في مجال حوسبة اللغات وفي غيرها من المجالات. **أهداف الدراسة:** تهدف هذه الدراسة الى رصد عملية انتقال الباحثين التحول نحو استخدام الشبكات العصبية الاصطناعية، وأثر ذلك على نماذج ماركوف المخفية. **منهجية الدراسة:** تم الاستعانة بقواعد البيانات الرقمية ومصادر الأبحاث الالكترونية لجمع أكبر عددٍ ممكن من الأوراق العلمية التي تتعلق بنماذج ماركوف المخفية والشبكات العصبية الاصطناعية، ثم فرزت هذه الأوراق لاختيار قائمة قصيرة من الأوراق النوعية المتعلقة بحوسبة اللغات، ثم صنفت الأوراق المختارة حسب فترة النشر، للخروج بتصور يوضح عملية الانتقال نحو الشبكات العصبية الاصطناعية. **النتائج:** تشير نتائج هذه الدراسة إلى أن العقد الأخير شهد تراجعاً في استخدام نماذج ماركوف المخفية، مقابل زيادة الاهتمام بالشبكات العصبية الاصطناعية (خاصة التعلم العميق)، ويمكن القول إن الشبكات العصبية الاصطناعية تهيمن اليوم على الأعمال البحثية الجديدة في مجال حوسبة اللغات وغيرها من المجالات. **الخلاصة والتوصيات:** إن التوجه الجديد نحو استخدام الشبكات العصبية الاصطناعية يفرض على الباحثين الجدد مواكبة هذا الانتقال، والاطلاع على نتائجه، والمساهمة فيه.

الكلمات المفتاحية: معالجة اللغات الطبيعية، حوسبة اللغة، نماذج ماركوف المخفية، الشبكات العصبية، التعلم العميق، تعلم الآلة، الخصائص، المصنفات.

مقدمة

تعود فكرة نماذج ماركوف إلى عالم الرياضيات الروسي (أندري ماركوف) (1856–1922م) الذي قدم أعماله في هذا المجال في بدايات القرن الماضي، وكانت باكورة أعماله في مجال العمليات العشوائية (random processes) في الأعوام (1905-1906م). تقوم فكرة ماركوف على تمثيل التغيرات العشوائية التي تحدث في نظام معين، من خلال نموذج رياضي إحصائي (count-based models)، بهدف استخدامه في عمليات التنبؤ والتصنيف، ونماذج ماركوف تتعلق بالانتقالات المتسلسلة أو المتتالية التي تحدث في نظام ما، ويمكن النظر إليها على أنها مخرجات النظام (observations). يتم بناء نماذج ماركوف بالاعتماد على ملاحظة المخرجات أو المشاهدات الصادرة من النظام لفترة طويلة، وتسمى هذه العملية عملية التدريب (training). تُستخدم نماذج ماركوف في مجالات كثيرة، إلا أن البداية الفعلية ارتبطت باللغويات، حيث استخدمها ماركوف في مجال الشعر، أحد مجالات اهتمامه، وحظيت نماذجه بعد ذلك بشهرة كبيرة، واهتمام واسع، نظراً لقدرتها على تمثيل الظواهر المختلفة بصورة رياضية، وبالتالي إمكانية الانتفاع منها في نمذجة العمليات المختلفة (Rabiner, 1989).

تركز هذه الدراسة على نماذج ماركوف المخفية (Hidden Markov Models) ويطلق عليها اختصاراً HMMs، وقبل الخوض في تفاصيل هذا النوع (أي نماذج ماركوف المخفية)، لا بد من الإشارة إلى أنه امتداد لفكرة نماذج ماركوف الظاهرة (ويطلق عليها نماذج ماركوف فحسب). تعتبر نماذج ماركوف (الظاهرة) الأساس الذي بنيت عليه نماذج ماركوف المخفية، اذ انها امتداد لنماذج ماركوف الظاهرة، وتعتمد عليها. كلمة نموذج في أعمال ماركوف يقصد بها تمثيل النظام على شكل مجموعة من الحالات (states)، بحيث يتنقل النظام بين هذه الحالات بصورة عشوائية، لكن ضمن احتمالات معينة. كلا النوعين سواء نماذج ماركوف الظاهرة أو المخفية يعتمد على سلاسل البيانات الممتدة على خط الزمن، ويطلق عليها (temporal data). الفرق الرئيس بين النوعين (أي نماذج ماركوف الظاهرة والمخفية) يتعلق بالحالات التي يتنقل بها النظام. فحالات النظام في نماذج (ماركوف الظاهرة) تكون ظاهرة جلية بصورة مباشرة، أما نماذج ماركوف (المخفية) فتتعلق بالظواهر غير المرئية للمشاهد ويطلق عليها (hidden states). ويكون الهدف هذه الحالة هو تحديد هذه الحالات (المخفية) بناء على الظواهر المشاهدة (observations). تحتوي هذه الورقة على عدد من الأمثلة والتطبيقات التي توضح الفرق بين نماذج ماركوف الظاهرة والمخفية.

1 كلية تكنولوجيا المعلومات وهندسة الحاسوب، جامعة بوليتكنك فلسطين، الخليل، فلسطين.
*الباحث المراسل: abuzeina@ppu.edu

المثال التالي يعرض طريقة استخدام نماذج ماركوف الظاهرة والمخفية. لنفترض أن حالة الطقس تأخذ شكلاً واحداً من بين ثلاث حالات وهي: {مشمس، ماطر، غائم}. اما المشاهدات فتنحصر في حالتين: {يحمل مظلة، لا يحمل مظلة}، ويكون المطلوب في هذه حالة الطقس هو الوصول إلى حالة الطقس المخفية (hidden states) بناء على المشاهدات الظاهرة (observations) والمتمثلة باستخدام المظلة من عدمه. ويتضح هذا الامر من خلال تخيل ان شخص يجلس في غرفة مغلقة، ويريد معرفة حالة الطقس في الخارج (وهي مخفية عنه) من خلال وجود مظلة من عدمه مع الأشخاص الداخلون الى الغرفة. ممكن كذلك أن تكون المشاهدات مختلفة، مثل: {ملابس الشخص القادم من الخارج مبللة، ملابس الشخص القادم من الخارج غير مبللة}. فمعرفة حالة ملابس الشخص (observation) سوف تقودنا إلى تحديد حالة الطقس المخفية (hidden state). عدد الحالات وعدد المشاهدات في نماذج ماركوف المخفية ثابت، وتعتمد على النظام الذي نعالجه. في هذه الدراسة. تستخدم نماذج ماركوف المخفية خوارزمية فتيربي (Viterbi) لتحديد الحالات المخفية بناء على المشاهدات الظاهرة. وهذا ما سوف نتطرق اليه في الأقسام اللاحقة.

جدول 1 يحتوي على مقارنة بين طبيعة استخدام نماذج ماركوف الظاهرة ونماذج ماركوف المخفية، مع التركيز على التطبيقات الشهيرة في مجال اللغويات الحاسوبية. مع العلم أن تطبيقات نماذج ماركوف، سواء الظاهرة أو المخفية أكبر من ذلك بكثير. أشير إلى ان عدداً من المسائل يمكن معالجتها باستخدام أي من الطريقتين. على سبيل المثال، يمكن بناء نظام لتصنيف البريد الإلكتروني (Spam filter) بالاعتماد على نماذج ماركوف الظاهرة وكذلك نماذج ماركوف المخفية.

جدول (1): عينة من تطبيقات نماذج ماركوف الظاهرة والمخفية.

نماذج ماركوف الظاهرة	نماذج ماركوف المخفية
- تحديد كاتب المقال (*Authorship attribution*)	- وسم الكلام (*Part of speech tagging*)
- تصنيف البريد الإلكتروني (*Spam filter*)	- التعرف على الكلام (*Speech recognition*)
- تحديد اللغة لنص معين (*Language identification*)	- التعرف الضوئي على الحروف (*Optical recognition character*)
- تقسيم الكلمة (*Word-dividing*)	- الترجمة الآلية (*Machine translation*)
- انتاج النصوص (*Text Generation*)	- التشكيل الآلي (*Machine diacritization*)
- تصنيف النصوص (*Text Classification*)	- تلخيص النصوص (*Text summarization*)
- كشف المعلومات السرية (*Steganography detection*)	- التعرف على الكيانات (*Named entity recognition*)
- تحديد عدد العناقيد (*Number of clusters*)	- التعرف على المشاعر (*Speech emotion recognition*)
- التحقق من الوثائق (*Document authentication*)	- التعرف على لغة الإشارة (*Sign language recognition*)
- دمج نتائج محركات البحث (*Rank aggregation*)	- تأليف الكلام (*Speech synthesis*)
- تحديد أسماء المواقع (*recognizing Names location*)	- تجزئة النصوص (*Text Segmentation*)
- اقتراح الكلمة التالية (*Next word prediction*)	- تصنيف الكلام (*Text Classification*)
- ترتيب صفحات النت (*PageRank*)	

تهدف هذه الدراسة الى دراسة طبيعة استخدام نماذج ماركوف المخفية، مع التركيز على تطبيقات معالجة اللغات الطبيعية (Natural language processing)، وتعرف اختصاراً (NLP). ثم نسلط الضوء على التوجه الحالي لدى الباحثين لاستخدام الشبكات العصبية الاصطناعية (Artificial Neural Networks or NNs - ANNs) في نفس المجال. فقد شهدت السنوات الأخيرة إقبالاً كبيراً على استخدام الشبكات العصبية (التعلم العميق)، مع تضاؤل استخدام نماذج ماركوف المخفية، مما حفز على دراسة هذه المسألة، ومتابعة هذا التحول. يهدف هذا البحث الى تسليط الضوء على الطرق الواعدة في مسائل معالجة اللغة، وتنوير الباحثين بزخم التوجيهات الحديثة، وأثر ذلك على الواقع البحثي. في القسم الثاني: أقدم نماذج ماركوف المخفية وعدد من تطبيقات معالجة اللغة، وأستعرض في القسم الثالث الشبكات العصبية، وأعرج على تاريخها وأنواعها، ثم أعرض عينة من استخداماتها، وفي القسم الرابع أستعرض نتائج هذه الدراسة، ثم أقدم في القسم الخامس الخلاصة والتوصيات.

نماذج ماركوف المخفية

تعتبر نماذج ماركوف المخفية امتداد لنماذج ماركوف الظاهرة، فالحالات التي يتنقل بها النظام في نماذج ماركوف الظاهرة واضحة للمشاهد، مثلاً: رمي حجر نرد، ومشاهدة النتيجة التي تكون واحدة من بين الحالات التالية: {1،2،3،4،5،6}، أو إلقاء قطعة نقد، ومشاهدة النتيجة {صورة، رقم}. نماذج ماركوف المخفية تختلف عن ذلك، فالمبدأ العام لنماذج ماركوف المخفية يقوم على تحديد الحالات المخفية من خلال المشاهدات الظاهرة. مثلاً، معرفة حالة الطقس من خلال المظلة، إن كانت مبللة أو غير مبللة، أو التعرف على الكلام المَحكي من خلال الإشارة الصوتية، أو التعرف على وسم (tag) الكلمة من خلال الكلمة (word or token).

رغم تاريخها الطويل، حافظت نماذج ماركوف المخفية على مكانتها، ولاقت قبولاً كبيراً لا يزال أثره حتى يومنا الحاضر، ويظهر لنا ذلك من خلال الأوراق العلمية والرسائل الجامعية القديمة والحديثة. قبل الخوض في تفاصيل الفروع التي أفادت فيها نماذج ماركوف، لا بد من توضيح المبدأ الرياضي لهذه النماذج: فنماذج ماركوف المخفية عبارة عن ثلاث مصفوفات احتمالية، بحيث تحتوي كل مصفوفة على قيم رقمية احتمالية، تمثل طبيعة السلوك او تعبر عن الانتقال داخل النظام المستهدف. تشمل هذه المصفوفات الاحتمالات التالية: مصفوفة الاحتمالية الابتدائية للحالة، ومصفوفة احتمالات الانتقال من حالة إلى أخرى، ومصفوفة احتمالات أن تنتج المشاهدة عن حالة معينة.

يتم حساب القيم الاحتمالية، وتعبئة المصفوفات، بالاعتماد على مجموعة التدريب (training data). الشكل رقم 1 يوضح هذه المصفوفات، ونحتاج أولاً لتحديد عدد الحالات المخفية في النظام (hidden states)، ونرمز له بالرمز (m)، كذلك نحتاج إلى تحديد عدد المشاهدات في النظام (observations)، ونرمز له بالرمز (n). تختلف التسميات لمصفوفات نماذج ماركوف، لكن في الغالب، تأخذ مصفوفة القيم الابتدائية الرمز (π)، ومصفوفة الاحتمالات الانتقالية الرمز (A)، ومصفوفات المشاهدات الرمز (B). هذه المصفوفات تشكل الأساس لجميع العمليات الخاصة بنماذج ماركوف المخفية، مثل عمليات التنبؤ او التصنيف.

مصفوفة القيم الابتدائية (*Initial probability*)	مصفوفة الاحتمالات الانتقالية (*State transition matrix*)
$\pi = \begin{bmatrix} P(1) \\ P(2) \\ \vdots \\ P(m) \end{bmatrix}$	$A = \begin{bmatrix} P(1\|1) & P(2\|1) & \cdots & P(m\|1) \\ P(1\|2) & P(2\|2) & \cdots & P(m\|2) \\ \vdots & \vdots & \vdots & \vdots \\ P(1\|m) & P(2\|m) & \cdots & P(m\|m) \end{bmatrix}$
مصفوفات المشاهدات (*Observation probability matrix*)	
$B = \begin{bmatrix} P(1\|1) & P(1\|2) & P(1\|3) & \cdots & P(1\|m) \\ P(2\|1) & P(2\|2) & P(2\|3) & \cdots & P(2\|m) \\ \vdots & & \vdots & & \vdots \\ P(n\|1) & P(n\|2) & P(n\|3) & \cdots & P(n\|m) \end{bmatrix}$	

شكل (1): تمثيل نموذج ماركوف من خلال المصفوفات.

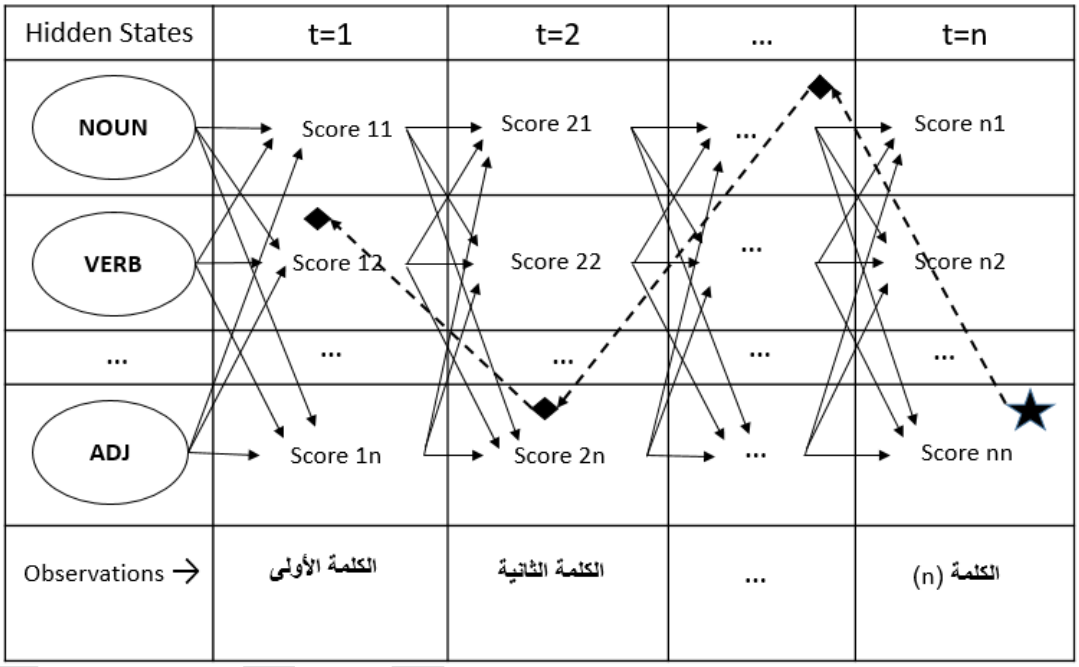
يتم استخدام المصفوفات أعلاه من خلال خوارزمية فتيربي (Viterbi) بحيث يتم التعرف على الحالات المخفية بناء على المشاهدات الظاهرة، من خلال عملية تراجع (backtracking) باستخدام القيم الاحتمالية الأعلى، كما سنرى في التطبيقات اللاحقة. فيما يلي خوارزمية فتيربي:

- For each state, initialize the probability to the initial probability * observation probability
- For each time t:
  For each state:
  Probability = max (probability $_{t-1}$ * transition probability) * observation probability

أفادت نماذج ماركوف المخفية في مجالات متعددة منها: الطب، وعلوم الصحة، وعلم الأحياء والتنبؤ الجيني، وأسواق المال والأعمال، وحالة الطقس والتنبؤات الجوية، وتنبؤات سلوك الأفراد، وغيرها. الا انني سوف أركز في هذه الدراسة على مجال واحد وهو معالجة اللغة، او ما يعرف بحوسبة اللغة. سوف نستخدم مجال واحد لمتابعة عملية التحول في استخدام الأدوات، ولرسم صورة واضحة عن التوجهات الحالية نحو استخدام الشبكات العصبية. نبدأ بالتعرف على التطبيقات الرئيسة في حوسبة اللغة، لاستخدامها في التحقق من ثبات وتقدم الشبكات العصبية، وتراجع نماذج ماركوف المخفية. ان فهم المسائل الرئيسة يسهل علينا إجراء المقارنة ما بين الطريقتين، وهما الأبرز في مسائل اللغويات الحاسوبية. للاختصار، سأعرض مرجعاً واحداً فقط لكل فرع من الفروع البحثية التي تم عرضها، كما اشير الى ان اغلب المراجع قديمة نظراً لاستخدام نماذج ماركوف منذ فترة طويلة، وهذا ما نحاول دراسته في هذا البحث.

وسم الكلام

وسم الكلام (Part of speech tagging) هو عملية تحديد الصيغة الصرفية لكل كلمة في النص، سواء كانت اسماً، أو فعلاً، أو حرفاً، وذلك بالاعتماد على السياق الذي ترد فيه. يفيد وسم النصوص في كثير من تطبيقات معالجة اللغات الطبيعية، فهو يستخدم في تحليل النصوص (syntactic parsing) للتحقق من صحة الجملة وتوافقها مع قواعد اللغة، ويستخدم في تحليل جملة الاستعلام (في محركات البحث) لفهم المعنى المطلوب، وبالتالي الاستجابة بصورة مناسبة. الشكل رقم 2 يوضح شبكة الحسابات (Trellis)، وما تقوم به خوارزمية فتيربي (Viterbi algorithm)، لاختيار الوسم الصحيح. بعد تعبئة القيم الاحتمالية المقابلة لكل كلمة في النص، يتم الرجوع (backtracking) من الكلمة الأخيرة (المشار اليها بالنجمة * الموجودة في العمود الأيمن من الشكل 2) للحصول على الحالات المخفية، والتي تمثل وسم الكلمات في النص. نشير الى أنه في مسألة وسْم الكلام: المشاهدات الظاهرة هي كلمات النص المراد وسمه، بينما الحالات المخفية تمثل الوسوم (tags). من الأعمال البحثية التي استندت إلى نماذج ماركوف المخفية: (Cutting et al., 1992). لمزيد من المعلومات حول وسم الكلام العربي (AbuZeina & Al-Tamimi, 2021).



شكل (2): شبكة حساب الاحتمالات (Trellis) وطريقة فتيربي في وسم سلسلة نصية.

التعرف على الكلام

التعرف على الكلام (Speech recognition) هي عملية تحويل الكلام إلى شيء رقمي مفهوم، سواء على شكل نصي أو على شكل تعليمات معينة، كالأوامر مثلاً. يفيد التعرف على الكلام في تسهيل عملية التواصل بين الإنسان والآلة، كما يستخدم في مجالات مختلفة مثل: أنظمة الرعاية الصحية، ومساعدة ذوي الاحتياجات الخاصة، والخدمات المصرفية، سواء عبر أجهزة الصراف الآلي أو من خلال شبكة الإنترنت، والتحكم في الآلات والمعدات ووسائل النقل، والتحقق من هوية الشخص، وغيرها. انتشر في السنوات الأخيرة نظامان لتنفيذ عملية التعرف على الكلام، وكلاهما يستخدم نماذج ماركوف المخفية، وهما: نظام (CMU Sphinx) الأمريكي، ونظام (HTK) البريطاني، فالمشاهدات الظاهرة هي الإشارة الصوتية للكلام المراد التعرف عليه، بينما الحالات المخفية هي سلسلة الكلمات المطابقة للكلام المحكي، من الأعمال البحثية التي استندت إلى نماذج ماركوف المخفية: (Picone, 1990).

التعرف الضوئي على الحروف

التعرف الضوئي على الحروف (Optical character recognition) ويطلق عليه اختصاراً (OCR) هو عملية التعرف على النص، سواء المطبوع (printed) أو المكتوب بخط اليد (handwritten)، والتعرف على الخط المطبوع يشمل – عادة – النصوص القديمة، وتكون على شكل صور، أما التعرف على النص المكتوب بخط اليد فهو عملية تحويل الخط اليدوي إلى نسخة رقمية قابلة للمعالجة (أي البحث والتعديل). تشمل عمليات التعرف على الخط اليدوي مجالات كثيرة، وتهدف في مجملها إلى تحقيق كفاءة أكبر في عمليات المعالجة اللاحقة، ومن هذه المجالات:

- الأوراق المالية مثل الشيكات، لتحديد رقم الشيك وقيمته وتاريخه وكافة التفاصيل.
- عناوين الرسائل البريدية الورقية، لتوجيه هذه الرسائل إلى عناوينها الصحيحة، دون تدخل بشري.
- التعرف على محتوى نموذج معين والاحتفاظ بكافة المعلومات والتفاصيل بشكل رقمي.

في مسألة التعرف الضوئي على الحروف، المشاهدات الظاهرة هي النص المطبوع أو المكتوب بخط اليد، ومطلوب التعرف عليه، وغالباً ما يكون على شكل صور، بينما الحالات المخفية هي النسخة الرقمية من الكلمات المدخلة. من الأعمال البحثية التي استندت إلى نماذج ماركوف المخفية: (Agazzi & Kuo, 1993).

الترجمة الآلية

الترجمة الآلية (Machine translation) عملية ترجمة نص من لغة إلى أخرى. تسهل هذه العملية التواصل بين الأشخاص المتحدثين بلغات مختلفة، وفهم التعليمات والإعلانات عند الانتقال من بلد إلى آخر. توفر الترجمة الآلية في التكلفة على المصالح التجارية، من خلال اختصار عدد المترجمين، مع تحقيق انتشار عالمي أكبر في مناطق مختلفة. من فوائدها أيضاً اختصار وقت الترجمة إلى لحظات، وبدقة عالية، علاوة على التحكم في الترجمة وتخصيصها لخدمة مجال معين، من خلال اختيار المعاني الملائمة لطبيعة المقاطع المترجمة وفرعها الدقيق. في هذه المسألة: المشاهدات الظاهرة هي كلمات النص المطلوب ترجمته، بينما الحالات المخفية هي الكلمات المقابلة في اللغة المترجم إليها. من الأعمال البحثية التي استندت إلى نماذج ماركوف المخفية: (Vogel, Ney, & Tillmann, 1996)

التشكيل الآلي

تمتاز اللغة العربية بأنها لغة معربة، ما يسمح باستخدام علامات التشكيل (diacritization) لتوضيح المعنى المقصود، وإزالة الغموض قدر الإمكان. تشمل حركات التشكيل (العلامات الإعرابية) المستخدمة في اللغة العربية: الفتحة، والضمة، والكسرة، وتنوين الفتح، وتنوين الضم، وتنوين الكسر، والشدة، والمد، إلا أن نصوص اللغة العربية (في الغالب) تهمل الحركات، ما يؤدي إلى إمكانية تعدد المعاني حسب السياق، فعندما نقول: (ما أحسنَ زيدٌ) فإن المعنى هنا هو نفي الإحسان عن زيد، وعندما نقول: (ما أحسنُ زيدٍ؟) فإننا نسأل عن الشيء الذي أحسن فيه زيد، وإن قلنا: (ما أحسنَ زيدا) أفادت التعجب. من الأمثلة أيضاً على كلمات تأخذ أكثر من معنى تبعاً للتشكيل: (بَر، بُر، بِر)، (جَنّة، جُنّة، جِنّة)، (كَلام، كُلام، كِلام)، الخ. في هذه المسألة: المشاهدات الظاهرة هي كلمات النص المطلوب تشكيله (أي نص غير مشكل)، بينما الحالات المخفية هي الكلمات المشكلة المطابقة للنص المدخل. من الأعمال البحثية التي استندت إلى نماذج ماركوف المخفية (Elshafei, Al-Muhtaseb, & Alghamdi, 2006).

تلخيص النصوص

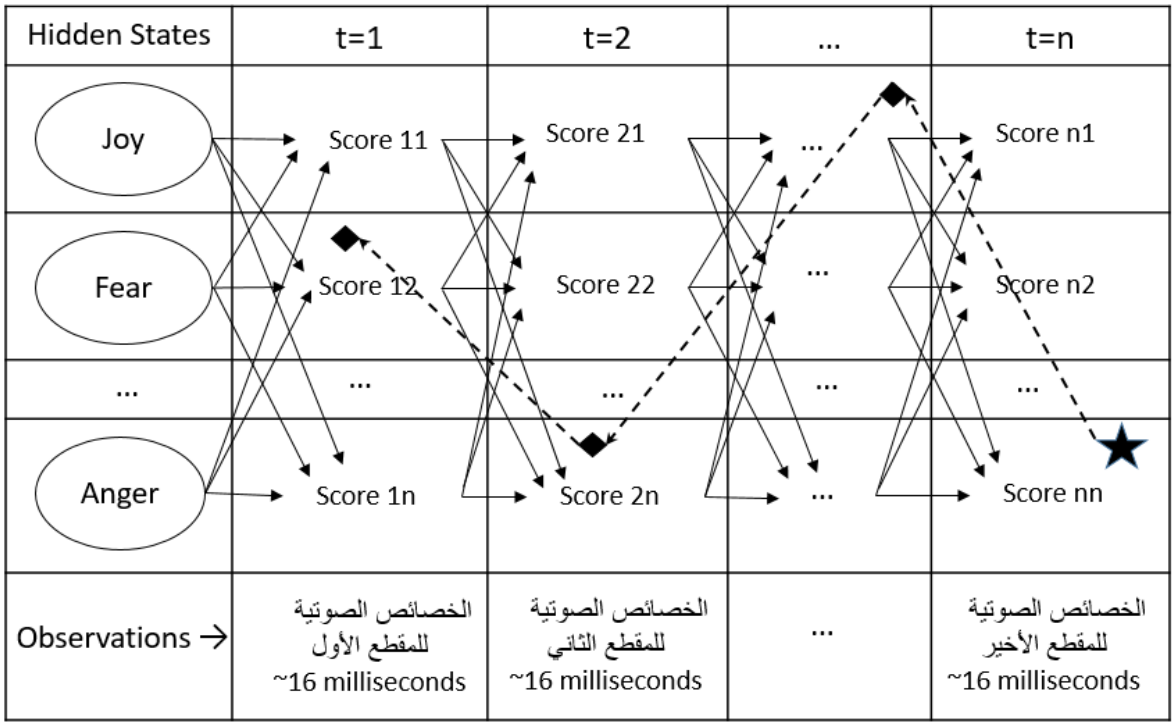
مع الانتشار الهائل للمعلومات الرقمية، سواء على شبكة الإنترنت، أو غيرها من المستودعات الرقمية، ازدادت الحاجة إلى برامج لاختصار الأفكار الرئيسة الموجودة في نص معين، وهو ما يعرف بـ (تلخيص النصوص) (Text summarization)، فاستخلاص الأفكار الرئيسة يختصر الوقت، سواء على صعيد الافراد (للاطلاع بصورة سريعة)، أو الأجهزة (للمعالجة والنقل بصورة سريعة). على سبيل المثال، تستخدم محركات البحث هذه الطريقة لعرض جزء من النص، بدلاً من عرضه كاملاً، ويزيد هذا الامر في الإنتاجية، ويساعد في سرعة انتشار نتائج الأبحاث. في هذه المسألة: المشاهدات الظاهرة هي جمل النص المراد اختصاره، بينما الحالات المخفية هي حالتان: "الإضافة إلى الملخص" و"عدم الإضافة إلى الملخص"، فإذا حصلت الجملة على صنف "الإضافة إلى الملخص"، تكون من ضمن الجمل التي سيتم إضافتها للملخص، والجمل التي تأخذ صنف "عدم الإضافة إلى الملخص" يتم استثنائها. الخصائص التي يتم قياسها لاتخاذ القرار بخصوص إضافة الجملة من عدمه تشمل: موقع الجملة في النص، وعدد الكلمات في الجملة، والعلاقة بين كلمات الجملة وكلمات النص كاملاً. من الأعمال البحثية التي استندت إلى نماذج ماركوف المخفية (Conroy & O'leary, 2001).

التعرف على الكيانات

تُستخدم برامج التعرف على الكيانات (Named entity recognition) في تحديد وتصنيف الكيانات داخل النص، وتشمل الأشخاص، والمؤسسات، والأماكن، والأوقات، والتواريخ، والقيم العددية، الخ. تفيد هذه العملية في كثير من تطبيقات معالجة اللغة، مثل الترجمة الآلية، ومحركات البحث، وتأليف الكلام، وأنظمة الاختبارات (Question answering)، الخ. يحتوي نظام التعرف على الكيانات على مجموعة أنواع أو أصناف (tags) معدة مسبقاً لهذا الغرض. في هذه المسألة: المشاهدات الظاهرة هي كلمات النص المطلوب تحديد الكيانات فيه، بينما الحالات المخفية هي الفئات المحددة مسبقاً في النظام. ممكن يحتوي النص على كيانات فحسب، وممكن ان يحتوي على رموز واشكال معينة لها فئاتها الخاصة. من الأعمال البحثية التي استندت إلى نماذج ماركوف المخفية (Zhou and Su, 2002).

التعرف على المشاعر

يستخدم التعرف على المشاعر (Speech emotion recognition) في تحديد الحالة العاطفية للشخص من خلال صوته، بشكل عام، تشمل الحالة العاطفية للشخص على مشاعر مختلفة مثل: {السعادة، الحزن، الغضب، الخوف، الانبهار، الاشمئزاز، عدم الاكتراث، الخ}. من تطبيقات التعرف على المشاعر ضبط الجودة في المؤسسات، ورفع مستوى الخدمة في المصالح التجارية، ويكون ذلك من خلال تحليل مشاعر العملاء بعد تواصلهم مع مراكز الخدمة، للتأكد من مستوى رضى الزبائن. من التطبيقات ايضاً تحديد حالة السائق، وتقديم المساعدة له عند الحاجة، تجنباً للأخطاء أثناء القيادة. في هذه المسألة: المشاهدات الظاهرة هي الصوت الخاص بالمتكلم، بينما الحالات المخفية هي الحالات العاطفية المعرفة مسبقاً في النظام. من الأعمال البحثية التي استندت إلى نماذج ماركوف المخفية (Schuller, Rigoll, & Lang, 2003). الشكل رقم 3 يوضح طريقة نماذج ماركوف في التعرف على المشاعر.



شكل (3): طريقة تحديد المشاعر بالاعتماد على صوت الانسان.

التعرف على لغة الإشارة

تستخدم لغة الإشارة (Sign language recognition) وهي إشارات من خلال اليد، كوسيلة للتواصل بين أفراد المجتمع ممن يعانون إعاقات سمعية (فئة الصم والبكم). تعرف هذه المسألة أيضاً بـ (Gesture recognition) وتعني التعرف على الإيماءات (حركة في الجسم وغالباً في اليد والرأس للتعبير عن المعنى). من خلال لغة الإشارة، يستطيع الصم والبكم التواصل فيما بينهم، وكذلك التواصل مع الفئات الأخرى من المجتمع. إلا أن تعلم هذه الرموز أمر شاق، يستدعي استخدام التكنولوجيا لتوفير حلول مناسبة. في هذه المسألة: المشاهدات الظاهرة هي صور تمثل حركات اليد، بينما الحالات المخفية هي معنى كل حركة وتكون معرفة مسبقاً في النظام. من الأعمال البحثية التي استندت إلى نماذج ماركوف المخفية (Liang & Ouhyoung , 1996).

تأليف الكلام

تهدف عملية تأليف الكلام (Speech synthesis) إلى تحويل النص المكتوب الى كلام منطوق بشكل آلي، ويطلق على هذه العملية أيضاً (Text-to-speech). ويساعد تأليف الكلام في قراءة التعليمات أو الإرشادات عند مواجهة صعوبة في قراءة هذه التعليمات نظراً لاختلاف الكلام في كل مرة، مثل توجيه السائق أثناء القيادة. كذلك، يساعد ذوي الإعاقة (السمعية أو البصرية) في الاطلاع على محتوى معين، من خلال الحصول على الكلام المنطوق المطابق للمحتوى النصي. كما يمكن الانتفاع من تأليف الكلام للاستماع إلى المقالات والكتب أثناء القيام بالأعمال الروتينية، مثل الأعمال المنزلية، أو قيادة السيارة. من الأنظمة التي تستخدم تقنية تأليف الكلام أنظمة المحاورة (Dialog systems) وتهدف إلى تحقيق التواصل الكلامي بين المستخدم والآلة، مثل نظام (Siri). في أنظمة تأليف الكلام: المشاهدات الظاهرة هي النص المطلوب تحويله إلى صيغة كلامية، بينما الحالات المخفية هي الصوت المطابق للنص المدخل. من الأعمال البحثية التي استندت إلى نماذج ماركوف المخفية (Masuko, Tokuda, Kobayashi, & Imai, 1996).

تجزئة النصوص

تهدف عملية تجزئة النصوص (Text Segmentation) إلى تقسيم النص لمقاطع متجانسة حسب الموضوع. فالعملية الرئيسة في هذه المسألة هي إيجاد حدود المقطع المتجانس داخل النص. يفيد هذا الأمر في الحصول على معلومات خاصة بموضوع أو حدث معين من مصادر مختلفة، مثل نشرات الأخبار المكتوبة أو الصوتية بعد إجراء عملية التعرف على الكلام. يفيد تجزئة النصوص في أنظمة استرجاع المعلومات، من خلال إرجاع المقطع المناسب لجملة الاستعلام، بدل إرجاع النص كاملاً. ويفيد في تتبع موضوع معين وتحديد المصادر التي ورد فيها، وتلخيص النصوص للحصول على أفضل تمثيل للنص المراد تلخيصه. في أنظمة تجزئة النصوص: المشاهدات الظاهرة هي الكلمات أو الجمل للنص المطلوب تجزئته، بينما الحالات المخفية هي حالتان، وهما أن الكلمة أو الجملة "حد فاصل" أو "ليست حدا فاصلا". من الأعمال البحثية التي استندت إلى نماذج ماركوف المخفية (Van Mulbregt, Carp, Gillick, Lowe, & Yamron, 1998).

تصنيف النصوص

تهدف عملية تصنيف النصوص (Text classification) إلى تحديد موضوع النص بناء على مجموعة مواضيع أو فئات محددة مسبقاً. تُفيد هذه العملية في تنظيم المعلومات لتسهيل عمليات البحث في المستودعات الرقمية الكبيرة. ترتيب الوثائق ضمن فئات معينة حسب الموضوع يقدم للمستخدم نتائج أفضل، ووصول أسرع للمعلومة. من خلال تتبع الاعمال السابقة، نرى ان نماذج ماركوف المخفية لم تستخدم بصورة كبيرة في مسألة تصنيف النصوص، كما هو الحال في التطبيقات اللغوية الأخرى، مثل وسم النصوص أو التعرف على الكلام، ويرجع ذلك إلى أن تصنيف النصوص لا يعتمد بصورة كبيرة على تسلسل البيانات كما هو الحال في وسم الكلام مثلاً. فالوثيقة أو النص يقدم بشكل كامل ليتم تصنيفه دون الالتفات إلى مسألة تتابع الكلمات، ينظر الى الكلمات في مسألة تصنيف النصوص على انها حقيبة من الكلمات (bag of words) دون النظر الى ترتيبها او تسلسلها، وهذا يخالف المبدأ الذي قامت عليه نماذج ماركوف

وهو السلاسل او الاحداث او الكلمات المتتابعة. في أنظمة تصنيف النصوص: المشاهدات الظاهرة هي النص المطلوب تصنيفه، بينما الحالات المخفية هي الفئات أو المواضيع المعرفة في النظام، مثل الاقتصاد، والتعليم، والصحة، الخ. من الأعمال البحثية التي استندت إلى نماذج ماركوف المخفية (Yi, 2005).

الشبكات العصبية الاصطناعية

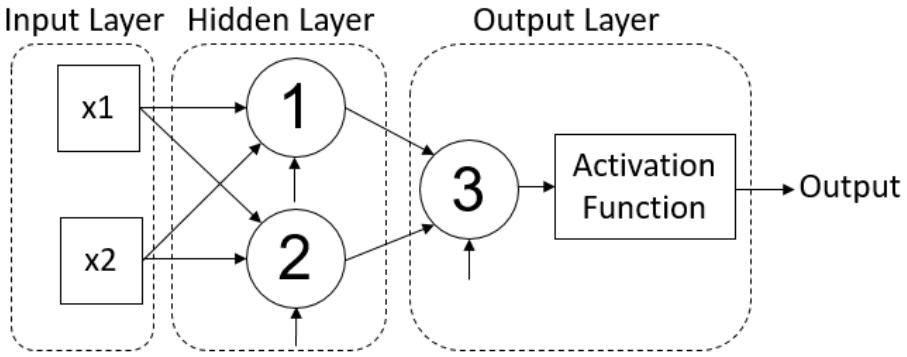
تعتبر الشبكات العصبية الاصطناعية (NNs)، واحدة من المكونات البرمجية التي تعمل بطريقة تحاكي طريقة عمل العقل البشري. تفيد الشبكات العصبية في كثير من التطبيقات مثل عمليات التنبؤ، والتحكم، والتصنيف، والألعاب، والعنقدة (تجميع المكونات المتشابه) وهو ما يعرف بـ (clustering)، إلى غير ذلك من المسائل البحثية التي تظهر بين الفينة والأخرى. يمكن النظر إلى الشبكات العصبية على أنها وحدة معالجة تقوم باستقبال القيم المدخلة، ومن ثم تقوم بتقديم المخرجات أو النتائج بعد إجراء الحسابات اللازمة، بناء على الأوزان المحددة في توصيلات الشبكة. فالشبكة العصبية تقوم على مبدأ مدخلات (inputs)، ومعالجة (processing)، ونتيجة (output).

بدأ العمل في مجال الشبكات العصبية في خمسينات القرن الماضي عندما قام (Rosenblatt, 1958) بتقديم الشبكة العصبية ذات الطبقة الواحدة والتي أطلق عليها اسم (Perceptron). بشكل عام، تهدف الشبكات العصبية إلى بناء نظام معالجة لديه القدرة على التعلم، بحيث يصبح مدرباً (trained) وقادراً على التنبؤ والتصنيف وتقديم النتائج للقيم المدخلة (أي عملية تصنيف)، وغالباً ما تكون هذه المدخلات جديدة (أي تعرض على الشبكة لأول مرة). تدريب الشبكات العصبية يكون من خلال عملية تكرارية تشمل تقديم عدد كبير من الأمثلة مع الإجابات الصحيحة لكل مثال. يطلق على هذه العملية (أي تقديم الأمثلة والإجابات) عملية التعلم لتقدير الأوزان (weights) بين الوصلات المختلفة داخل الشبكة، بعدها تكون الشبكة مدربة وقادرة على إعطاء مخرجات صحيحة للمدخلات الجديدة (لم يسبق عرضها على الشبكة). الجدول 2 يوضح التسلسل الزمني لتطور الشبكات العصبية.

جدول (2): تطور الشبكات العصبية الاصطناعية.

الفترة	الموضوع
فترة الأربعينات	البداية الحقيقية للشبكات العصبية الاصطناعية بناء على فهم طريقة عمل الخلية العصبية الطبيعية. قام (McCulloch & Pitts, 1943) بتطوير نموذج لخلية عصبية اصطناعية، وكان هذا النموذج محصورا فى العمليات المنطقية البسيطة وغير قادر على التعلم. قدم (Hebb, 1949) أفكاره فيما يتعلق بقدرة الشبكة العصبية على التعلم في نهاية الأربعينات.
فترة الخمسينات	تم تقديم نموذج (Rosenblatt, 1958) وهو نموذج بسيط لخلية عصبية اصطناعية تحاكى طريقة تخزين المعلومات فى الذاكرة مع القدرة على التعلم. هذا النموذج يحتوى على مدخلات ومخرجات مع القدرة على التحكم في الأوزان الخاصة بالمدخلات بناء على المخرجات.
فترة الستينات	قدم (Widrow & Hoff, 1960) طريقة (Least mean squares-LMS) لتطوير عملية التعلم وحساب الأوزان فى الشبكات العصبية.
فترة السبعينات	شهدت بداية السبعينات هدوءا نسبيا بعد أن صدر كتاب (Perceptrons) لمؤلفيه (Minsky & Papert, 1969)، حيث وضحا عدداً من المآخذ على الشبكات العصبية. إلا أن الكتاب قدم أفكاراً جديدة تتعلق بالطبقات المتعددة، وخوارزمية الاتجاه الأمامى (feed-forward). فى العام 1974 قدم (Werbos,1974) خوارزمية الاتجاه العكسى (backpropagation) التي تسمح بالتعلم فى الطبقات المتعددة. وكانت هذه خطوة هامة لإعادة النشاط البحثى للشبكات العصبية الاصطناعية.
فترة الثمانينات	شهدت فترة الثمانينات نشاطا بحثيا مميزا حيث قدم (Rumelhart et al., 1986) فكرة المعالجة الموزعة المتوازية. كما تم تطوير نظريات التعلم بشكل أفضل مما كان عليه فى نموذج (perceptron)، حيث تم الاعتماد على خوارزمية الاتجاه العكسى المقدمة فى العام 1974. كما تم التركيز الشبكات العصبية متعددة الطبقات (Network Multi-Layer Neural). وظهرت الشبكات العصبية التكرارية Recurrent Neural Network (RNN) المصممة للتعامل مع التطبيقات التى تحتوى على سلاسل البيانات مثل التعرف على الصوت والترجمة الآلية. فى هذه الفترة ظهرت أيضاً شبكة هوبفيلد (Hopfield, 1982)، وهى شكل من أشكال شبكات RNN، لكنها مختصة بأنماط معينة من المسائل، مثل القيم القصوى أو الاستمثال (إيجاد الأمثل من بين مجموعة من القيم ضمن الموارد المتاحة) وهو ما يعرف بـ (optimization). من الأمثلة على هذه المسائل مسألة البائع المتجول (Travelling Salesman Problem). ظهرت فى هذه الفترة أيضاً شبكات Self-organized map (SOM) المخصصة للعنقدة (Clustering).
فترة التسعينات	انخرط الكثير من الباحثين فى أبحاث الشبكات العصبية فى مجالات مختلفة كعلم الأعصاب، وعلم النفس، والرعاية الصحية، والرياضيات، والفيزياء، والهندسة، وأسواق المال، وصناعة الدوائر المتكاملة، والبيئة، والمناخ، والزراعة، والفنون، واللغويات الحاسوبية، وغير ذلك من العلوم المختلفة. ظهرت شبكة Long Short Term Memory (LSTM) وهي تطوير لشبكة (RNN) لكنها تتمتع بذاكرة طويلة الأمد، مما عزز أداء الشبكات العصبية، خاصة فيما يتعلق بسلاسل البيانات. ظهرت الشبكات العصبية الالتفافية Convolutional neural networks (CNN) لمعالجة المسائل التى تحتوى على مصفوفات ذات أبعاد مختلفة، واستخدمت بشكل أساسي لأغراض التصنيف (classification)، والأبحاث المتعلقة بمعالجة الصور والصوت والفيديو.
بعد 2000	ظهرت شبكات المعتقد العميق Deep Belief Network (DBN) فى (2006) لتعزيز السرعة والأداء فى الشبكات العصبية متعددة الطبقات (Hinton, Osindero, Teh, & 2006) كما ظهرت الشبكات التنافسية المولدة Generative Adversarial Networks (GANs) فى (2014) بإمكانيات إضافية ومنها التعامل مع البيانات غير المنتظمة (Goodfellow et al., 2014). توسع استخدام الشبكات العصبية بصورة كبيرة وشمل مجالات متعددة. الدراسة (Abiodun, 2018) قدمت عرضاً شاملاً عن الأبحاث الحالية والناشئة فى مجال الشبكات العصبية. قدم (LeCun, Bengio, & Hinton, 2015) دراسة شاملة عن التعلم العميق وإمكانياته، فيما قدم (Minar & Naher 2018) التطورات الحديثة في موضوع التعلم العميق.

الشكل رقم 4 يمثل شبكة عصبية اصطناعية بسيطة. مع العلم أن الشبكات العصبية تكون أكبر من ذلك بكثير من حيث عدد المدخلات، وعدد الطبقات الداخلية، وعدد المخرجات. يظهر في الشكل دالة التفعيل (Activation Function) لتفعيل مخرج محدد على طبقة الإخراج، بالاعتماد على قيم المدخلات على طبقة الادخال.



شكل (4): شبكة عصبية اصطناعية متعددة الطبقات مع مدخلين.

فيما يلي أعرض عدداً من تطبيقات معالجة اللغة التي استندت إلى الشبكات العصبية، وأبدأ بما انتهيت به في القسم السابق وهو تصنيف النصوص (Text classification).

تصنيف النصوص

أشرنا في القسم السابق إلى أن نماذج ماركوف المخفية لم تستخدم بصورة كبيرة في مسألة تصنيف النصوص، وهذا يشير إلى توفر البديل الجيد من خوارزميات تعلم الآلة، مثل مصنف بيز (Bayes classifier)، ومصنف متجه الدعم الآلي (Support Vector Machine)، ومقياس التشابه cosine similarity)

measure)، وغيرها الكثير. إلا أن الشبكات العصبية قد حققت نجاحاً كبيراً في مسألة تصنيف النصوص. شهد العام (2013) انطلاقة قوية في تصنيف النصوص بالاعتماد على الشبكات العصبية، عندما استخدم (Mikolov, Chen, Corrado, & Dean, 2013) الشبكات العصبية في انتاج خصائص نصية بالغة الثراء في المعنى والتركيب (syntactic and sematic rich)، وقد أطلق على هذه الخصائص مصطلح (Word embeddings). أحدثت هذه الطريقة (أي طريقة الـ Word embeddings) نقلة نوعية في مسألة تصنيف النصوص والتطبيقات المرتبطة بها، خاصة فيما يتعلق باسترجاع المعلومات في محركات البحث ومنها محرك البحث الشهير (جوجل). وقد توالت التحسينات على طريقة عمل وإنتاج خصائص الكلمات (Word embeddings)، وشهدت تحسينات كثيرة وقفزات نوعية مما حقق أداء عالي في أنظمة استرجاع المعلومات. من هذه التحسينات تقديم طرق للتعامل مع الاطوال المختلفة للمدخلات مثل Bidirectional Encoder Representations from) Transformers) ويطلق عليه (BERT) (Devlin, Chang, Lee, & Toutanova, 2018) وكذلك (OpenAI - GPT) اختصاراً لـ (Generative Pre-trained Transformer) (Radford et al. , 2019) . هذا التحول في تصنيف النصوص من الطرق التقليدية، ومنها طريقة نماذج ماركوف، يشير إلى التوجه الجديد، ودور الشبكات العصبية في تصنيف الأنماط، فضلاً عن استخدامها في الأغراض البحثية المختلفة. الفكرة الرئيسية ومبدأ عمل الشبكات العصبية هو التعامل المباشر مع الخصائص لإنتاج أوزان التوصيلات، بخلاف طرق تعلم الآلة (machine learning) الأخرى التي تلجأ إلى العد الإحصائي، أو إنشاء دوال الكثافة الاحتمالية (probability density function) واختصاراً يطلق عليه (PDF).

التعرف على الكلام

على الرغم من الاستخدام الواسع لنماذج ماركوف المخفية في مجال التعرف على الكلام، إلا أن العقد الأخير شهد تقدماً في استخدام الشبكات العصبية في مسألة التعرف على الكلام. فالمسح الأدبي للأعمال السابقة أظهر أن عدداً من الأعمال تحولت بشكل جزئي إلى استخدام الشبكات العصبية (أي أنها تدمج نماذج ماركوف المخفية والشبكات العصبية)، فيما استخدمت أعمال أخرى الشبكات العصبية بصورة كاملة. فيما يلي عينة من الأعمال التي استخدمت الشبكات العصبية بصورة جزئية. استخدم (Hinton et al., 2012) الشبكات العصبية في بناء النماذج الصوتية (Acoustic models) عوضاً عن الطريقة الشائعة وهي طريقة (Gaussian mixture models) ويطلق عليه اختصاراً (GMM)، وقد تبين أن هذا التعديل يحقق تحسناً في أداء أنظمة التعرف على الصوت، واستبدال دوال الكثافة الاحتمالية (Probability density function PDF) التي تعتمد على (GMM) بالشبكات العصبية أظهر أيضاً تحسناً في الأداء في أعمال (Pan, Liu, Wang, Hu, & Jiang, 2012).

من الأعمال التي استبدلت نماذج ماركوف بالشبكات العصبية بصورة كاملة ما قدمه (Mustafa, Allen, & Appiah, 2019)، حيث بين أن أداء الشبكات العصبية في التعرف على الكلام أفضل من أداء نماذج ماركوف المخفية، وقد تركزت التجربة على عملية التعرف من خلال الأجهزة الذكية. كذلك قدم (Wazir & Chuah, 2019) نظاماً للتعرف على أصوات الأرقام العربية باستخدام الشبكات العصبية، حيث استخدمت الخصائص الصوتية مباشرة كمدخلات للشبكة العصبية.

استخدم (Graves, Mohamed,& Hinton, 2013) الشبكات العصبية Long short-term memory (LSTM) في التعرف على الكلام المتصل، بالاعتماد على مدونة صوتية خاصة باللغة الإنجليزية يطلق عليها (Massachusetts Institute / Texas Instruments of Technology - TIMIT) للكلام المتصل، واستخدم (Chorowski, Bahdanau, Cho, & Bengio, 2014) الشبكات العصبية التكرارية RNN في التعرف على الكلام المتصل، مدونة (TIMIT) للكلام المتصل. قدم (Zen, 2015) دراسة تشير إلى عملية الانتقال نحو الشبكات العصبية RNN في بناء النماذج الصوتية (Acoustic models)، أحد المكونات الرئيسية في أنظمة التعرف على الكلام. ما تم الحديث عنه مجرد عينة محدودة من الأعمال السابقة التي تبين شيوع الشبكات العصبية في مجال التعرف على الصوت، وأغلبها دراسات حديثة نسبيا.

وسم الكلام

يعد وسم الكلام من التطبيقات المهمة في مجال حوسبة اللغة، لأنه مرتبط بتطبيقات أخرى لا تقل أهمية عن عملية الوسم ذاتها. الدراسات السابقة أشارت إلى استخدام الشبكات العصبية في مسألة وسم الكلام، فيما يلي عينة من هذه الأعمال. استخدم (Meftah, Semmar, & Sadat, 2018) الشبكات العصبية في وسم الكلام المتداول عادة في شبكات التواصل الاجتماعي، والمدونات الشخصية، والرسائل النصية. استخدم (Wang, Qian, Soong, He, & Zhao, 2015) الشبكات العصبية في وسم الكلام، وأظهرت النتائج أداء متقدماً مقارنة مع الطرق الأخرى، وكان ذلك لمدونة (corpus) قياسية معروفة في هذا المجال. استخدم (Deshmukh & Kiwelekar, 2020) الشبكات العصبية في وسم الكلام المتحدث به في منطقة معينة غرب الهند، فيما استخدم (Can, Karaoğlan, Kışla, & Metin, 2021) الشبكات العصبية لوسم الكلام للغة التركية.

الترجمة الآلية

لفترة طويلة، استخدمت الطريقة الإحصائية في الترجمة الآلية، والتي تعتمد على عَدِّ الكلمات والمقاطع لاعتماد الترجمة الأعلى احتمالياً. إلا أن الأمر اختلف خلال السنوات الأخيرة، بعد أن هيمنت الشبكات العصبية على أبحاث الترجمة. فيما يلي عينة من الأبحاث المتعلقة بعملية الترجمة:

قدم (Stahlberg, 2020) دراسة شاملة توضح النقلة النوعية التي حصلت في السنوات الأخيرة، مما يكشف عن تقدم الشبكات العصبية وتراجع الطرق الإحصائية مثل نماذج ماركوف. المسح الأدبي يشير إلى أن أغلب أعمال الترجمة تتعلق باللغة الإنجليزية، مما يحفز التوجه نحو استخدام الشبكات العصبية في الأعمال البحثية الخاصة بترجمة اللغة العربية. استخدم (Bahdanau, Cho, & Bengio, 2014) الشبكات العصبية (Encoder-decoder architecture) للترجمة الآلية بين اللغات الإنجليزية والفرنسية. استخدم (Datta, David, Mittal, & Jain, 2020) الشبكات العصبية التكرارية RNN للترجمة الآلية بين اللغات الإنجليزية والفرنسية، فيما استخدم (Mi, Sankaran, Wang, & Ittycheriah, 2016) الشبكات العصبية للترجمة بين اللغة الإنجليزية واللغة الصينية.

نتائج الدراسة

قدمت في الأقسام السابقة تصوراً عاماً عن طبيعة استخدام كل من نماذج ماركوف المخفية والشبكات العصبية. في هذا القسم، أقدم عدداً من الفروقات العملية التي تبين تفوق الشبكات العصبية على نماذج ماركوف المخفية، وسيكون التركيز بالتحديد على نماذج ماركوف المخفية HMMs والشبكات العصبية التكرارية RNN، لأن كلا الطريقتين تعتمد على سلاسل البيانات الممتدة عبر الزمن (sequence of random observations)، وهذا النوع من المشاهدات موجود ومرتبط بصورة كبيرة بتطبيقات معالجة اللغة، مثل التعرف على الصوت، والترجمة الآلية، ووسم الكلام.

قبل الخوض في المقارنة بين نماذج ماركوف المخفية والشبكات العصبية التكرارية، لا بد من الإشارة إلى أن نماذج ماركوف المخفية تنتمي إلى مسار تعلم الآلة، في حين أن الشبكات العصبية التكرارية تنتمي إلى مسار التعلم العميق، وهو أيضاً ينضوي تحت مظلة تعلم الآلة كما أشار الى ذلك (Shinde & Shah, 2018)، وكلاهما من فروع الذكاء الاصطناعي. الجدول التالي يقدم فروقاً بين هذين المسارين، وبعض هذه الفروق مقتبس من (Elastic Platform Team, 2024)

جدول (3): بعض الفروقات بين تعلم الآلة والتعلم العميق.

تعلم الآلة (Machine Learning)	التعلم العميق (Deep Learning)
يدعم التعلم بالإشراف (supervised) وبدون إشراف	يدعم التعلم بالإشراف (supervised) وبدون إشراف
متخصصة في البيانات المتسلسلة	متخصصة في البيانات المتسلسلة
لا يحتاج إلى كمية كبيرة من البيانات للتدريب	يحتاج إلى كمية كبيرة من البيانات للتدريب
وقت التدريب أقل من التعلم العميق	يحتاج إلى وقت أكبر للتدريب
إمكانية ضبط وتعديل المتغيرات محدودة	إمكانية ضبط وتعديل المتغيرات متاحة
يمكن التنفيذ على معالجات تقليدية	يحتاج إلى معالجات عالية الأداء
غير مناسب للأبعاد الكبيرة (high dimensionality)	يتعامل بصورة أفضل مع الأبعاد الكبيرة للبيانات
يحتاج معالجة قبلية (preprocessing) بصورة أكبر	الحاجة إلى خبراء اللغة أقل، وكذلك المعالجة القبلية أقل
ارتباط محدود بالسياق (limited context independence)	اعتماد سياق أوسع للكلمة (long-distance dependency)
دقة الأداء أقل من التعلم العميق	يعطي دقة عالية في الأداء
يخدم تطبيقات تقليدية	يخدم تطبيقات معقدة
استخدام ثابت وإمكانيات ثابتة	يتطور ويتقدم نحو المستقبل بقوة

أعرض فيما يلي عدداً من الأبحاث التي قدمت نتائج عملية تظهر تفوق الشبكات العصبية على نماذج ماركوف المخفية، الجدول رقم 4 يعرض عدداً من هذه الأعمال، وللاختصار، يعرض الجدول عشرة أعمال لستة مجالات. المعلومات المقدمة في الجدول تشير إلى أن توجه الباحثين في السنوات الأخيرة يتركز في توظيف التعلم العميق في مسائل حوسبة اللغة، وقد بينت النتائج صحة هذا التوجه وأنه يحقق أداء أفضل.

جدول (4): مقارنة بين نماذج ماركوف والمخفية والشبكات العصبية التكرارية.

المجال	العمل البحثي
التعرف على الكلام	قدم (Jin, Zhang, Huang, & Miao, 2018) دراسة تتعلق بالتعرف على الكلام، وقد بيّن من خلال النتائج العملية أن الشبكات العصبية RNN قدمت نتائج أفضل من الطرق التقليدية مثل نماذج ماركوف المخفية HMM وخوارزمية فتيربي.
التعرف على الكلام	على الرغم من أن تركيزنا في هذا الجدول على مقارنة أداء الشبكات العصبية التكرارية، إلا أن الدراسة (Karita et al., 2019) قدمت طريقة (Transformer) وهي طريقة تعتمد على التعلم العميق، وتشير الدراسة إلى أنها تفوقت على طريقة RNN في تطبيقات معالجة اللغات الطبيعة، وبضمنها تطبيقات التعرف على الصوت وتأليفه، وهذا الأمر يعزز فهمنا أن التعلم العميق يهيمن على المشهد البحثي.
تأليف الكلام	قدم (Adam, 2020) دراسة تتعلق بتأليف الكلام، وقد بيّن من خلال النتائج العملية أن الشبكات العصبية RNN قدمت نتائج أفضل من الطرق التقليدية مثل نماذج ماركوف المخفية HMM وخوارزمية فتيربي.
الترجمة الآلية	قدم (Liu, Yang, Li, & Zhou, 2014) دراسة تتعلق بالترجمة الآلية، وقد بيّن من خلال النتائج العملية أن الشبكات العصبية RNN قدمت نتائج أفضل من الطرق الإحصائية.
وسم الكلام	قدم (Baghdasaryan, 2023) دراسة تتعلق بوسم نصوص اللغة الأرمنية، وتشير الدراسة إلى أن الشبكات العصبية RNN تتفوق بصورة كبيرة على الطريقة التقليدية الممثلة بنماذج ماركوف المخفية HMM وخوارزمية فتيربي.
وسم الكلام	قدم (Pradhan & Yajnik, 2024) دراسة تتعلق بوسم نصوص اللغة النيبالية، تشير الدراسة إلى أن الشبكات العصبية RNN تتفوق بصورة كبيرة على الطريقة التقليدية الممثلة بنماذج ماركوف المخفية HMM وخوارزمية فتيربي.
وسم الكلام	قدم (Mishra, Shaikh, & Sanyal, 2022) دراسة تتعلق بوسم نصوص اللغة الهندية، وتشير الدراسة إلى أن الشبكات العصبية RNN تتفوق بصورة كبيرة على الطرق التقليدية، ومنها نماذج ماركوف المخفية (HMMs).
وسم الكلام	قدم (AlKhwiter & Al-Twairesh, 2021) دراسة تتعلق بوسم نصوص اللغة العربية، وتشير الدراسة إلى أن الشبكات العصبية RNN تتفوق بصورة كبيرة على طريقة Conditional Random Fields (CRF). وعلى الرغم من أن تركيزنا يتعلق بنماذج ماركوف المخفية، إلا أنني أضيف هذه الدراسة، للإشارة إلى تقدم الشبكات العصبية على نماذج ماركوف وعلى غيرها من طرق الوسم المعروفة.
تصنيف النصوص	قدم (Zhou, Sun, Liu, & Lau, 2015) دراسة تتعلق بتصنيف النصوص، باستخدام الشبكات العصبية التكرارية RNN ومقارنتها مع الشبكات العصبية الالتفافية. (يفهم ضمناً أن الموضوع تجاوز نماذج ماركوف المخفية).
لغة الإشارة	قدم (Lee et al., 2021) دراسة تتعلق بالتعرف على لغة الإشارة، باستخدام الشبكات العصبية التكرارية، مع نسبة دقة تصل 99.4%. (يفهم ضمناً ان الموضوع تجاوز نماذج ماركوف المخفية).
عمليات التنبؤ بشكل عام	قدم (Almqvist, 2021) دراسة تفيد بأفضلية الشبكات العصبية على نماذج ماركوف المخفية في عمليات التنبؤ بشكل عام.

ان افضلية الشبكات العصبية على نماذج ماركوف المخفية تكمن في طريقة العمل لكلا الطريقتين، فالشبكات العصبية تعتمد على طبقات متصلة بين المدخلات والمخرجات، تحاكي طريقة عمل العقل البشري، وقادرة على التعلم من الأخطاء اثناء مرحلة التدريب. بينما تعتمد طريقة نماذج ماركوف المخفية على العد الاحصائي وحساب الاحتمالات للوصول الى النتائج. علاوة على ذلك فان الشبكات العصبية قادرة على التعامل مع سياق ممتد للسلاسل، بينما تعتمد نماذج ماركوف على عدد محدد من العناصر المتجاورة.

الخلاصة والتوصيات

تتبعت هذه الدراسة مسارين رئيسيين في مجال معالجة اللغات الطبيعية: مسار نماذج ماركوف المخفية، ومسار الشبكات العصبية بأنواعها المختلفة، وكلا المسارين على درجة كبيرة من الأهمية، إلا أن المؤشرات المبنيّة على رصد وتتبع الأبحاث العلمية في مجال التنبؤ والتصنيف تشير إلى أن هناك توجهاً واضحاً من قبل الباحثين لتوظيف الشبكات العصبية الاصطناعية. بطبيعة الحال، فإن هذه التوجه قلل من نصيب الطرق الأخرى، مثل نماذج ماركوف المخفية وغيرها من الطرق التقليدية في تعلم الآلة.

عرضت في الورقة عينة يسيرة من الأبحاث التي تعزز التصور بنجاح الشبكات العصبية وهيمنتها على المشهد البحثي، وهذا التحول نحو الشبكات العصبية يتطلب ترتيب الأولويات، من خلال زيادة الاهتمام بالتعلم العميق، ليشمل جميع الفئات المرتبطة بالبحث العلمي، وبضمنهم الأبحاث الممولة في الجامعات، وطلبة الدراسات العليا. هذه الدراسة ربما تكون من باكورة الأعمال التي تسلط الضوء على هذا التحول، وبالتالي، فالأمر بحاجة إلى مزيد من البحث والتمحيص، لوضع الأمور في نصابها، وتحديد فيما إذا كان هذا التحول شاملاً وكاملاً، أم أنه جزئي، وضمن جوانب محددة. من خلال هذا البحث يتضح لنا أن الأعمال القوية في مجال الشبكات العصبية كانت في السنوات الأخيرة، وهذا يعزز فرص البحث العلمي من خلال توظيف الطرق الجديدة في اثراء حوسبة اللغة العربية. لكن دراستنا هذه لن تغلق الباب بصورة مطلقة امام نماذج ماركوف المخفية، فقد اشارت الدراسة (Ky & Tuyen, 2018) الى تفوق نماذج ماركوف المخفية على الشبكات العصبية، وقد تكون هذه النتيجة مرتبطة بطبيعة المسألة التي عرضت في الورقة (التنبؤ بأسعار الأسهم)، فربما تكون بعض المسائل ملائمة أكثر للطريقة الإحصائية (ماركوف). نشير أخيراً الى ورقة (Alshafei, 2007) حيث تربط بين معالجة اللغات الطبيعية والأنظمة الخبيرة والشبكات العصبية، مما يعز التوجه نحو الشبكات العصبية.

في الختام، يتضح لنا أهمية الانخراط مبكراً إلى عالم التعليم العميق، فهي منطقة بحثية واعدة، خاصة مع توفر المصادر البرمجية المفتوحة (open-source software) التي يمكن الانتفاع منها في خدمة أبحاث اللغة العربية. وبشكل محدد، فإن لغة البرمجة (Python) تحتوي على مكتبات قوية لخدمة حوسبة اللغة مثل (Gensim and spaCy)، لكنها مخصصة بشكل أساسي لمعالجة اللغة الإنجليزية.

الموافقة الأخلاقية والموافقة على المشاركة: البحث لا يحتوي على عينات للدراسة.

توافر البيانات والمواد: تم جمع بيانات الدراسة من خلال مصادر المعلومات الرقمية، ولم تستخدم أي مواد في الدراسة.

مساهمة المؤلفين: الباحث منفرد.

تضارب المصالح: لا يوجد تضارب في المصالح مع أي جهة أخرى. كما أن هذا البحث ليس مستل من رسالة ماجستير أو دكتوراة.

التمويل: لم يحصل البحث على أي تمويل مادي.

الشكر والتقدير: يتقدم الباحث بالشكر والتقدير لجامعة بوليتكنك فلسطين، على توفيرها بيئة العمل المناسبة لإنجاز هذا البحث.

2024, from https://data-flair.training/blogs/deep-learning-vs-machine-learning/

- Elshafei, M., Al-Muhtaseb, H., & Alghamdi, M. (2006). Statistical methods for automatic diacritization of Arabic text. In *The Saudi 18th National Computer Conference. Riyadh* (Vol. 18, pp. 301-306).
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, *27*.
- Graves, A., Mohamed, A. R., & Hinton, G. (2013, May). Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing* (pp. 6645-6649). Ieee.
- Hebb, D. O. (1949). The first stage of perception: growth of the assembly. *The Organization of Behavior*, *4*(60), 78-60.
- Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, *18*(7), 1527-1554.
- Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., ... & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing magazine*, *29*(6), 82-97.
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, *79*(8), 2554-2558.
- Jin, X., Zhang, K., Huang, X., & Miao, M. (2018, December). On continuous speech recognition of Indian English. In *Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence* (pp. 1-6).
- Karita, S., Chen, N., Hayashi, T., Hori, T., Inaguma, H., Jiang, Z., ... & Zhang, W. (2019, December). A comparative study on transformer vs rnn in speech applications. In *2019 IEEE automatic speech recognition and understanding workshop (ASRU)* (pp. 449-456). IEEE.
- Ky, D. X., & Tuyen, L. T. (2018). A Higher order Markov model for time series forecasting. *International Journal of Applied Mathematics and Statistics*, *57*(3), 1-18.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436-444.
- Lee, C. K., Ng, K. K., Chen, C. H., Lau, H. C., Chung, S. Y., & Tsoi, T. (2021). American sign language recognition and training method with recurrent neural network. *Expert Systems with Applications*, *167*, 114403.
- Liang, R. H., & Ouhyoung, M. (1996, July). A sign language recognition system using hidden markov model and context sensitive search. In *Proceedings of the ACM symposium on virtual reality software and technology* (pp. 59-66).
- Liu, S., Yang, N., Li, M., & Zhou, M. (2014, June). A recursive recurrent neural network for statistical machine translation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1491-1500).
- Masuko, T., Tokuda, K., Kobayashi, T., & Imai, S. (1996, May). Speech synthesis using HMMs with dynamic features. In *1996 ieee international conference on acoustics, speech, and signal processing conference proceedings* (Vol. 1, pp. 389-392). IEEE.
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, *5*, 115-133.
- Meftah, S., Semmar, N., & Sadat, F. (2018, May). A neural network model for part-of-speech tagging of social media texts. In *LREC 2018-Eleventh International Conference on Language Resources and Evaluation*.
- Mi, H., Sankaran, B., Wang, Z., & Ittycheriah, A. (2016). Coverage embedding models for neural machine translation. *arXiv preprint arXiv:1605.03148*.

References

- Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A., & Arshad, H. (2018). State-of-the-art in artificial neural network applications: A survey. *Heliyon*, *4*(11).
- AbuZeina, D. E., & Al-Tamimi, T. (2021). Arabic Part of Speech Tagging by Using the Stanford System: Prepositions as a Case Study. *An-Najah University Journal for Research* - B (Humanities), 35(5), 743–768. https://doi.org/10.35552/0247-035-005-003
- Adam, E. E. B. (2020). Deep learning based NLP techniques in text to speech synthesis for communication recognition. *Journal of Soft Computing Paradigm (JSCP)*, *2*(04), 209-215.
- Agazzi, O. E., & Kuo, S. S. (1993). Hidden Markov model based optical character recognition in the presence of deterministic transformations. *Pattern recognition*, *26*(12), 1813-1826.
- AlKhwiter, W., & Al-Twairesh, N. (2021). Part-of-speech tagging for Arabic tweets using CRF and Bi-LSTM. *Computer Speech & Language*, *65*, 101138.
- Almqvist, O. (2019). A comparative study between algorithms for time series forecasting on customer prediction: An investigation into the performance of ARIMA, RNN, LSTM, TCN and HMM.
- Alshafei, B. (2007). The Treatment of Some Deviation Cases in Non-Standard Data Using Artificial Intelligence and Natural Language Processing Applied to Arabic Language. *An-Najah University Journal for Research* - B (Humanities), 21(4), 1187–1197. https://doi.org/10.35552/0247-021-004-010
- Baghdasaryan, V. H. (2023, August). Comparative Analysis of Hidden Markov Model and Bidirectional Long Short-Term Memory for POS Tagging in Eastern Armenian. *International Journal of Scientific Advances*, Volume: 4, Issue: 4.
- Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.
- Can, Ş., Karaoğlan, B., Kışla, T., & Metin, S. K. (2021). Using Word Embeddings in Turkish Part of Speech Tagging. *International Journal of Machine Learning and Computing*, *11*(5), 367-372.
- Chorowski, J., Bahdanau, D., Cho, K., & Bengio, Y. (2014). End-to-end continuous speech recognition using attention-based recurrent nn: First results. *arXiv preprint arXiv:1412.1602*.
- Conroy, J. M., & O'leary, D. P. (2001, September). Text summarization via hidden markov models. In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 406-407).
- Cutting, D., Kupiec, J., Pedersen, J., & Sibun, P. (1992, March). A practical part-of-speech tagger. In *Third conference on applied natural language processing* (pp. 133-140).
- Datta, D., David, P. E., Mittal, D., & Jain, A. (2020). Neural machine translation using recurrent neural network. *International Journal of Engineering and Advanced Technology*, *9*(4), 1395-1400.
- Bhat, P., Behal, S., & Dutta, K. (2023). Machine learning and deep learning techniques for detecting malicious android applications: An empirical analysis. *Proceedings of the Indian National Science Academy*, *89*(3), 429-444.
- Deshmukh, R. D., & Kiwelekar, A. (2020, March). Deep learning techniques for part of speech tagging by natural language processing. In *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)* (pp. 76-81). IEEE.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Elastic Platform Team (2024), *Deep Learning vs Machine Learning – Demystified in Simple Words*, Retrieved July 25,

Committee on Applied Mathematics, Harvard University, Cambridge, MA.

- Widrow, B., & Hoff, M. E. (1960, August). Adaptive switching circuits. In *IRE WESCON convention record* (Vol. 4, No. 1, pp. 96-104).
- Yi, K. (2005). *Text classification using a hidden Markov model*. (Unpublished Doctoral dissertation). McGill University.
- Zen, H. (2015). Acoustic modeling in statistical parametric speech synthesis-from HMM to LSTM-RNN. *Proc. MLSLP*, *15*.
- Zhou, C., Sun, C., Liu, Z., & Lau, F. (2015). A C-LSTM neural network for text classification. *arXiv preprint arXiv:1511.08630*.
- Zhou, G., & Su, J. (2002, July). Named entity recognition using an HMM-based chunk tagger. In *Proceedings of the 40th annual meeting of the association for computational linguistics* (pp. 473-480).
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Minar, M. R., & Naher, J. (2018). Recent advances in deep learning: An overview. *arXiv preprint arXiv:1807.08169*.
- Minsky, M., & Papert, S. (1969). An introduction to computational geometry. *Cambridge tiass., HIT*, *479*(480), 104.
- Mishra, A., Shaikh, S. H., & Sanyal, R. (2022). Context based NLP framework of textual tagging for low resource language. *Multimedia Tools and Applications*, *81*(25), 35655-35670.
- Mustafa, M. K., Allen, T., & Appiah, K. (2019). A comparative review of dynamic neural networks and hidden Markov model methods for mobile on-device speech recognition. *Neural Computing and Applications*, *31*, 891-899.
- Pan, J., Liu, C., Wang, Z., Hu, Y., & Jiang, H. (2012, December). Investigation of deep neural networks (DNN) for large vocabulary continuous speech recognition: Why DNN surpasses GMMs in acoustic modeling. In *2012 8th International Symposium on Chinese Spoken Language Processing* (pp. 301-305). IEEE.
- Picone, J. (1990). Continuous speech recognition using hidden Markov models. *IEEE Assp magazine*, *7*(3), 26-41.
- Pradhan, A., & Yajnik, A. (2024). Parts-of-speech tagging of Nepali texts with Bidirectional LSTM, Conditional Random Fields and HMM. *Multimedia Tools and Applications*, *83*(4), 9893-9909.
- Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257-286.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, *1*(8), 9.
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, *65*(6), 386.
- Rumelhart, D. E., McClelland, J. L., & PDP Research Group. (1986). *Parallel distributed processing, volume 1: Explorations in the microstructure of cognition: Foundations*. The MIT press.
- Schuller, B., Rigoll, G., & Lang, M. (2003, April). Hidden Markov model-based speech emotion recognition. In *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03).* (Vol. 2, pp. II-1). Ieee.
- Shinde, P. P., & Shah, S. (2018, August). A review of machine learning and deep learning applications. In *2018 Fourth international conference on computing communication control and automation (ICCUBEA)* (pp. 1-6). IEEE.
- Stahlberg, F. (2020). Neural machine translation: A review. *Journal of Artificial Intelligence Research*, *69*, 343-418.
- Van Mulbregt, P., Carp, I., Gillick, L., Lowe, S., & Yamron, J. (1998, December). Text segmentation and topic tracking on broadcast news via a hidden Markov model approach. In *ICSLP*.
- Vogel, S., Ney, H., & Tillmann, C. (1996). HMM-based word alignment in statistical translation. In *COLING 1996 Volume 2: The 16th International Conference on Computational Linguistics*.
- Wang, P., Qian, Y., Soong, F. K., He, L., & Zhao, H. (2015). Part-of-speech tagging with bidirectional long short-term memory recurrent neural network. *arXiv preprint arXiv:1510.06168*.
- Wazir, A. S. M. B., & Chuah, J. H. (2019, June). Spoken Arabic digits recognition using deep learning. In *2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS)* (pp. 339-344). IEEE.
- Werbos, P. (1974). Beyond regression: New tools for prediction and analysis in the behavioral sciences. *PhD thesis,*

DRAFT