

# هذه الخوارزمية تتصفح الويكيبيديا لكي تولد كتاباً تعليمية بشكل آلي

تمثل موسوعة ويكيبيديا مورداً قيماً للغاية. ولكن من غير الواضح كيف يمكن تجميع المحتوى الخاص بموضوع معين في إطار متكامل.

بقلم مارتن جايلز 20 يناير 2019

“التعلم الآلي-الدليل الكامل” يعادل مجلداً ضخماً. هذا الكتاب - الذي تزيد عدد صفحاته على 6,000 - يُعد مقدمة شاملة إلى التعلم الآلي، يتضمن أحدث الفصول التي تتحدث عن الشبكات العصبونية الاصطناعية، والخوارزميات الوراثة، والرؤية الآلية.

ولكنه ليس نسخة عادية، إنه كتاب ويكي (ويكيبيوك)، عبارة عن كتاب تعليمي يمكن لأيِّ كان قراءته أو تعديله، مكون من مقالات مأخوذة عن ويكيبيديا، الموسوعة الهائلة المتواجدة على الإنترنت.

هنا تكمن قوة الفكرة، فالمعلومات التي يتم جمعها من مصادر جماهيرية متنوعة يتم تحديثها باستمرار بأحدث التطورات، ويتم تعديلها باستمرار لتصحيح الأخطاء والمعلومات المبهمة.

## Example of Wiki-book

Concept  
↓

### Book:Machine Learning - The Complete Guide

From Wikipedia, the free encyclopedia



Copyright Cover Needed

This is a **Wikipedia book**, a collection of Wikipedia articles that can be easily saved, rendered electronically, and ordered as a printed book.

Edit this book
Book Creator · Wikitext

Select format to download
PDF (A4) · PDF (Letter)

Order a printed copy from these publishers:
PedaPress

[About] [Advanced] [FAQ] [Feedback] [Help] [WikiProject] [Recent Changes]

Articles

- Active learning (machine learning)
- Reinforcement learning
- Multi-task learning
- Transduction
- Explanation-based learning

**Introduction and Main Principles**

- Machine learning
- Data analysis
- Occam's razor
- Curse of dimensionality
- No free lunch theorem
- Accuracy paradox
- Overfitting
- Regularization (machine learning)
- Inductive bias
- Data dredging
- Ugly duckling theorem
- Uncertain data

**Search Methods**

- Nearest neighbor search
- Stochastic gradient descent
- Beam search
- Best first search
- Breadth-first search
- Hill climbing
- Grid search
- Brute-force search
- Depth-first search
- Tabu search
- Anytime algorithm

**Statistics**

- Exploratory data analysis
- Covariate
- Statistical inference
- Algorithmic inference
- Bayesian inference
- Base rate
- Bias (statistics)
- Gibbs sampling
- Cross-entropy method
- Latent variable
- Maximum likelihood
- Maximum a posteriori estimation
- Expectation-maximization algorithm
- Expectation propagation
- Kullback-Leibler divergence
- Generative model

**Main Learning Paradigms**

- Supervised learning
- Unsupervised learning
- Offline learning
- Online learning model
- Online machine learning
- Hyperparameter optimization

**Classification Tasks**

- Classification in machine learning
- Concept class
- Features (pattern recognition)
- Feature vector
- Feature space
- Concept learning
- Binary classification
- Decision boundary
- Multiclass classification
- Class membership probabilities
- Calibration (statistics)
- Concept drift
- Prior knowledge for pattern recognition
- Its flower data set (Classic data sets)

**Online Learning**

- Margin Infused Relaxed Algorithm

**Semi-supervised learning**

- Semi-supervised learning
- One-class classification
- Coupled pattern learner



**WIKIPEDIA**  
The Free Encyclopedia



**WIKIBOOKS**  
Open books for an open world

مثال عن كتاب ويكي

ولكن هذا الأمر يمثل نقطة ضعف أيضاً، فويكيبيديا تتضمن كما هائلاً من المحتوى. إن تحديد ما الذي ينبغي تضمينه في مثل هذا الكتاب التعليمي يعد مهمة صعبة، ولعل هذا ما يبرر كون الكتاب ضخماً للغاية. فبوجود أكثر من 550 فصلاً، لا يمكن القول إن قراءته عملية سهلة.

هذا الأمر يطرح سؤالاً مثيراً للاهتمام. بالنظر إلى التقدم الذي تم إحرازه في الذكاء الاصطناعي في السنوات الأخيرة، هل هناك وسيلة لتعديل محتوى ويكيبيديا بشكل آلي بغية إنشاء محتوى كلي متسق مفيد بصيغة كتاب تعليمي؟

هنا يأتي دور شاهار أدماتي وزملائهما في جامعة بن-غوريون في النقب في إسرائيل. قام هذا الفريق بتطوير طريقة لتوليد كتب ويكي بشكل تلقائي باستخدام التعلم الآلي، ويطلقون على آلتهم البرمجية اسم ويكيوك-بوت. يقول الفريق: "يتمثل عنصر الابتكار في تقنياتها في أنها تهدف إلى توليد كتاب ويكي موسوعي كامل، دون تدخل بشري".

إن الأسلوب المتبع وضح نسبياً. حيث بدأ الباحثون بتحديد مجموعة من كتب الويكي التي يمكنها أن تلعب دور مجموعة بيانات للتدريب. بدأوا العمل باستخدام 6,700 من كتب الويكي المدرجة في مجموعة بيانات توفرها ويكيبيديا لهذا النوع من الدراسات الأكاديمية.

بما أن كتب الويكي هذه تشكل إلى حد ما معياراً ذهبياً لكل من التدريب والاختبار، فقد احتاج الفريق إلى وسيلة للتأكد من جودتها. يقول الفريق: "لقد اخترنا التركيز على كتب الويكي التي جرى تصفحها 1000 مرة على الأقل، وذلك بناء على الافتراض القائل بأن كتب الويكي الشائعة تتمتع بجودة معقولة".

خلقت هذه الطريقة 490 كتاب ويكي قاموا بإخضاعها للمزيد من الفلترة، وذلك بناء على عوامل مثل وجود أكثر من 10 فصول في الكتاب الواحد. ليبقى بعد ذلك 407 كتاب ويكي استخدمها الفريق في تدريب آلتهم البرمجية.

بعد ذلك قام الفريق بتقسيم مهمة إنشاء كتاب ويكي إلى عدة أجزاء، يتطلب كل منها مهارة مختلفة من التعلم الآلي. تبدأ المهمة بعنوان يؤلفه شخص بشري، يصف مفهوماً من نوع ما، مثل التعلم الآلي - الدليل الكامل.

تتمثل المهمة الأولى في تدقيق مجموعة مقالات ويكيبيديا بأكملها لتحديد ما هو ملائم بما يكفي لتضمينه. يقول أدماتي وزملاؤه: "تمثل هذه المهمة تحدياً صعباً نظراً للحجم الهائل للمقالات الموجودة في ويكيبيديا، والحاجة إلى اختيار المقالات الأكثر صلة بالموضوع من بين ملايين المقالات المتوفرة".

للحصول على المساعدة في هذه المهمة، استخدم الفريق البنية الشبكية للويكيبيديا، فغالباً ما تشير المقالات إلى مقالات أخرى باستخدام الارتباطات التشعبية. الافتراض بأن المقال المرتبط تشعبياً من المرجح أن يكون ذا صلة هو أمر منطقي.

لذا بدأوا بنواة صغيرة من المقالات التي تشير إلى المفهوم الأساسي في العنوان. ثم قاموا بعد ذلك بتحديد جميع المقالات التي ترتبط بهذه المفاهيم الأساسية على بعد 3 مستويات من الارتباط التشعبي على الشبكة.

ولكن، ما هو العدد الذي ينبغي تضمينه من هذه المقالات المرتبطة؟ لمعرفة ذلك، بدأوا بالعناوين التي وضعها أفراد من البشر لكتب الويكي الأربعمئة والسبعة التي بقيت، وأجروا تحليل الارتباط ثلاثي المستويات. ثم قاموا بتحليل كم من المحتوى الموجود في الكتب التي ألفها البشر الذي قام الأسلوب المؤتمت بتضمينه.

تبين أن الأسلوب المؤتمت غالباً ما تضمن الكثير من محتوى كتب الويكي الأساسي مع الكثير من المحتوى الإضافي. لذا فقد احتاج الفريق إلى طريقة أخرى لتشذيب المحتوى بشكل أكبر.

مرة أخرى، يأتي دور علم الشبكات. فكل كتاب ويكي ألفه أحد البشر لديه بنية شبكية خاصة به، يتم تحديدها من خلال عدد الروابط التشعبية التي تشير إليه انطلاقاً من المقالات الأخرى، وعدد الروابط التشعبية التي تشير انطلاقاً منه إلى غيره من المقالات، وقائمة بالمقالات المضمنة في هذا الكتاب والتي تحوي ترتيب الصفحات، وهكذا دواليك.

لذلك قام الفريق بإنشاء خوارزمية تتفحص كل مقالة تم اختيارها بشكل آلي بحثاً عن موضوع معين، ثم تحدد ما إذا كان تضمينها في كتاب ويكي سوف يجعل البنية الشبكية أكثر شبهاً بالكتب التي ألفها البشر أم لا. وإذا كانت النتيجة سلبية، يتم تجاهل المقال.

الخطوة التالية هي تنظيم المقالات ضمن فصول، وهي مهمة تنطوي في جوهرها على فكرة التجميع، حيث يجري تفحص الشبكة التي تشكلها مجموعة المقالات بأكملها ومعرفة كيفية تقسيمها ضمن تجمعات مترابطة بشكل متين. هناك العديد من خوارزميات التجميع المتوفرة لإنجاز هذا النوع من المهام.

الخطوة النهائية هي تحديد الترتيب الذي ينبغي للمقالات أن تظهره وفقه في كل فصل من الفصول. لفعل ذلك، يقوم الفريق بتنظيم المقالات ضمن أزواج ويستخدمون نموذجاً يستند إلى مفهوم الشبكة لتحديد أي منهما ينبغي أن يظهر أولاً. ومن خلال تكرار ذلك على مجموعة أزواج المقالات كلها، حيث تتولى الخوارزمية إيجاد الترتيب المفضل للمقالات وبالتالي الفصول.

وبهذه الطريقة تمكن الفريق من إنتاج إصدارات آلية من كتب الويكي التي تم إنشاؤها مسبقاً على يد البشر. من الصعب الحكم على مدى جودة الكتب المؤتمتة مقارنة بالكتب التي ألفها البشر. من المؤكد أنها تحتوي على قدر كبير من المواد نفسها، وبترتيب مماثل غالباً، وهو ما يعد بداية جيدة.

لكن أدامتي وزملاءه لديهم خطة لتحديد مدى الفائدة من مقاربتهم. فهم يخططون لإنتاج مجموعة من كتب الويكي التي تتحدث عن مواضيع لم تقم أي من الكتب البشرية بتغطيتها بعد.

سيقومون بعد ذلك بمراقبة مشاهدات الصفحات والتعديلات التي تجري على هذه الكتب لمعرفة مدى الشعبية التي ستحققها، ومدى كثافة التعديل الذي تخضع له، مقارنة بالكتب التي ألفها البشر. يقول الفريق: "سيكون هذا اختباراً حقيقياً لمقاربتنا".

إنه عمل مثير للاهتمام لديه القدرة على إنتاج كتب تعليمية قيّمة تغطي طيفاً واسعاً من المواضيع، وحتى على إنشاء نصوص أخرى مثل وقائع المؤتمرات.

لم يتحدد بعد كم ستكون هذه الكتب قيّمة للقراء من البشر، ولكننا سوف نواصل المراقبة لمعرفة ذلك.

مرجع: [arxiv.org/abs/1812.10937](https://arxiv.org/abs/1812.10937): إنتاج ويكيبيوك-بوت الآلي لكتب ويكيبيديا