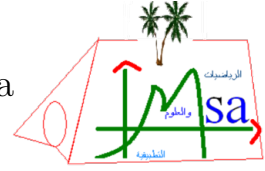




الجمهورية الجزائرية الديمقراطية الشعبية  
People's Democratic Republic of Algeria

وزارة التعليم العالي والبحث العلمي

Ministry of Higher Education and Scientific Research



جامعة غرداية

كلية العلوم والتكنولوجيا

مخبر الرياضيات و العلوم التطبيقية

قسم الرياضيات والإعلام الآلي

مذكرة تخرج لنيل شهادة الماستر

شعبة: الرياضيات والإعلام الآلي

تخصص: الأنظمة الذكية لإستخراج المعارف (SIEC)

الموضوع

تحليل المشاعر: نهج التعلم العميق

Sentiment Analysis: Deep Learning Approach

من إعداد الطالبتين

أميمة رزاق

أروى رزاق

أعضاء اللجنة:

الأستاذ. سليمان أولاد النوي	أستاذ محاضر ب	جامعة غرداية	- رئيسا
الأستاذ. سليمان بالاعور	أستاذ محاضر ب	جامعة غرداية	- مشرفا
الأستاذ. عبد الرحمان عجيلة	أستاذ مساعد أ	جامعة غرداية	- مناقشا
الأستاذ. عبد القادر بوهاني	أستاذ مساعد أ	جامعة غرداية	- مناقشا

السنة الجامعية 2019/2020

## ملخص

تحليل المشاعر أو التنقيب عن الرأي هو مجال يدرس استقطابية الآراء من النصوص ويعد من أكثر المجالات البحثية نشاطا في الوقت الراهن، ويرجع هذا إلى ظهور وسائل التواصل الاجتماعي حيث أصبح تبادل الآراء والتعبير عنها أمرا سهلا، لكن يبقى الإشكال في كيفية تحليل هذه الآراء وتقييمها، وبما أن تحليلها يدويا يستغرق وقتا طويلا إن لم نقل مستحيلا فقد حاول الباحثون إيجاد حلول تساعد في تحليلها آليا حيث ظهرت الأنظمة المستندة على القواعد والأنظمة المستندة على التعلم الآلي وكان للتعلم العميق دورا مهما في تقدم هذه الأبحاث. نقتراح في هذا العمل نموذج تعلم عميق يعتمد على الشبكات العصبية التلافيفية والشبكات العصبية المتكررة لتحليل المشاعر باللغة العربية باستخدام مجموعة بيانات Arabic ULMFiT Model. كانت النتائج المتحصل عليها جيدة فقد حققت دقة (Accuracy) تصل إلى 95% ونسبة خطأ (Loss) قدرت ب 15%.

**الكلمات المفتاحية:** التعلم العميق، الشبكات العصبية التلافيفية، الشبكات العصبية المتكررة، تحليل المشاعر، Arabic ULMFiT Model.

# Abstract

Sentiment analysis or opinion mining is a field that studies the polarity of opinions from texts. It is one of the active research's fields at present time, and this due to the rise of social media, where it became easy to express and exchange opinions. However the problem remains in how to analyze and evaluate these opinions. Since analyzing them manually is time consuming or even impossible, researchers have tried to find solutions that help in analyzing those opinions automatically where rule-based systems and machine learning systems appear, and deep learning play an important role in the advancement of this researches. In this work we propose a deep learning model based on convolutional neural networks and recurrent neural networks to analyze arabic sentiments using 'Arabic ULMFiT Model' dataset. The obtained results are good as they achieve an accuracy up to 95% and a loss of 15%.

**Key words:** Deep learning, Convolutional neural networks, Recurrent neural networks, Sentiment analysis, Arabic ULMFiT Model.

# شكر و عرفان

قال رسول الله صلى الله عليه وسلم :

"مَنْ سَلَكَ طَرِيقًا يَلْتَمِسُ فِيهِ عِلْمًا سَهَّلَ اللَّهُ لَهُ بِهِ طَرِيقًا إِلَى الْجَنَّةِ".

رواه مسلم

إن الفضل والحمد والشكر كله لله تعالى أن وفقنا لإنجاز هذا العمل وإيَّاه سبحانه نسأل أن يجعله من العلم النافع في الدنيا ولنا في الآخرة،  
نصلي ونسلم على أشرف المرسلين  
سيدنا وحبينا محمد صلى الله عليه وسلم وعلى آله وصحبه أجمعين.

إن الاعتراف بالجميل ما هو إلا جزء يسير من رده ولأن الكلمات كل ما نملكه إزاء كل من غمرنا بالجميل ونخص بالذكر المشرف الدكتور الفاضل "سليمان بلعور" الذي لم يخل علينا بالنصح والإرشاد لتسهيل عملنا في إطار هذا البحث، لكم منا أسمى معاني التقدير والاحترام.

كما نتوجه بالشكر الجزيل و العرفان للعزيزة مارية و الأخ عبد الجليل رزاق.  
كما لا ننسى التوجه بأسمى عبارات الشكر إلى كل رفقاء الدرب من أساتذة وزملاء في الدراسة خاصة طلبة ماستر إعلام آلي 2018-2020.  
إلى كل من كان له الفضل في إنجاز هذا العمل ولو بدعاء أو كلمة تشجيع.

إلى كل من تمنى أن يرى عملنا هذا شيئاً ملموساً  
وإلى كل من سره نجاحنا والشكر لله من قبل ومن بعد.

## إهداء

الحمد لله أقصى مبلغ الحمد و الشكر له من قبل و من بعد.  
إلى من أحمل اسمه بكل فخر، إلى الذي سهر على تعليمي بتضحيات جسام  
إلى مدرستي الأولى في الحياة، أبي الغالي أطال الله في عمره.  
إلى من ساندتني في دعائها و صلاتها، إلى من سهرت الليالي تنير دربي  
إلى من لم تذخر جهدا في سبيل نجاحي و إسعادي دوما، أمي الحبيبة  
جزاها الله عني خير الجزاء.

إلى من هم ملاذي و ملجأئي، إلى من تقاسمت معهم أحلى لحظات الحياة  
إخوتي أنار الله دروبهم.

إلى أنسي و أنيستي، إلى من تقاسمت معي مصاعب الحياة  
أختي حبيبتي سدد الله خطاها.

إلى بهجة البيت: مريم، أماني، ريهام، رؤى، عمر، زينم، سجاد، معاذ...  
إلى الغاليات: إيمان، سارة، هاجر، سوسو.

إلى من جمعتنا الدراسة فأصبحنا أعز أخوات MjnAch.

إلى صديقتي: مروة، وفاء، فرح، نبيلة.

إلى الأهل و الأقارب و الأحباب.

إلى من نسيه القلم و لم ينسه القلب.

أميمة.

## إهداء

إلى من ساندتني و دعمتني بدعائها و لم تدخر جهدا في سبيل نجاحي  
إلى نبع الحب و الحنان الغالية أُمي أطال الله في عمرك.

إلى من تعب لأجلي و لم يبخلني بأي شيء إلى أعظم و أعز رجل في الدنيا  
صاحب الوجه الطيب الحنون أبي أدامك الله سندا لي.

إلى من آثروني على أنفسهم و ساندوني طيلة حياتي و مشواري الدراسي الأعرء إخوتي.

إلى من جعلوا حياتي بهجة و فوضى الغوالي أبناء إخوتي.

إلى من تشاركنا مشوار الدراسة و جنون الحياة صديقاتي الغاليات

و أخص بالذكر مجموعة MjnAch.

إلى كل إنسان طيب قابلته في حياتي.

أهدي لكم ثمرة جهدي هذه.

أروى.

# الفهرس

iv	قائمة الأشكال
v	قائمة الجداول
vi	قائمة الخوارزميات

## 1

## مقدمة

3	مدخل للتعلم العميق Deep Learning	1
3	مقدمة	1.1
3	الشبكات العصبية Neural Networks	2.1
4	الشبكات العصبية الحيوية (البيولوجية) Biological Neural Networks	1.2.1
4	الشبكات العصبية الاصطناعية Artificial Neural Networks	2.2.1
6	دوال التنشيط	3.2.1
9	أنواع الشبكات العصبية الاصطناعية	4.2.1
11	تعليم الشبكات العصبية الاصطناعية	5.2.1
11	خوارزميات تعلم الشبكة العصبية	6.2.1
13	الإفراط في التعلم و التفریط في التعلم Overfitting and Underfitting	7.2.1
15	التعلم العميق	3.1
16	التحسين في التعلم العميق Optimization in deep learning	1.3.1
17	التسوية في التعلم العميق Regularization in deep learning	2.3.1
18	الشبكات العصبية التلافيفية Convolutional Neural Networks (CNNs)	3.3.1
22	الشبكات العصبية المتكررة Recurrent Neural Network (RNNs)	4.3.1
30	خاتمة	4.1

## 31

## 2 تحليل المشاعر Sentiment Analysis

31	مقدمة	1.2
31	مفاهيم أولية	2.2
31	ماهي المشاعر و ماهو الرأي؟	1.2.2

32	لماذا تحليل المشاعر؟	2.2.2
32	معايير تحليل المشاعر	3.2.2
33	أنواع الآراء	4.2.2
34	مستويات تحليل المشاعر	5.2.2
35	Natural Language Processing معالجة اللغة الطبيعية	3.2
35	طرق تمثيل البيانات	4.2
37	مقاييس تقييم النموذج	5.2
38	تحليل المشاعر	6.2
39	الأنظمة المستندة على القواعد	1.6.2
43	الأنظمة المستندة على التعلم الآلي	2.6.2
58	الأنظمة الهجينة	3.6.2
58	تحديات تحليل المشاعر	7.2
59	خاتمة	8.2
<b>60</b>	<b>دراسة تجريبية Implementation</b>	<b>3</b>
60	مقدمة	1.3
60	هندسة الشبكة	2.3
62	مجموعة البيانات	3.3
63	معالجة البيانات	4.3
64	بيئة العمل Environment	5.3
65	مناقشة النتائج	6.3
67	خاتمة	7.3
70	المراجع العلمية	



# قائمة الأشكال

4	Biological Neuron	خلية عصبية حيوية	1.1
5	Artificial Neural Network	شبكة عصبية اصطناعية	2.1
6	Artificial Neuron	خلية عصبية اصطناعية	3.1
7	Sigmoid	منحنى الدالة	4.1
7	Tanh	منحنى الدالة	5.1
8	ReLU	منحنى الدالة	6.1
9	Softmax	منحنى الدالة	7.1
10		الشبكات العصبية ذات التغذية الأمامية	8.1
10		الشبكات العصبية ذات التغذية الخلفية	9.1
14		الإفراط في التعلم و التفریط في التعلم	10.1
15		شبكة عصبية عميقة	11.1
19		الشبكة العصبية التلافيفية	12.1
20		عملية الالتفاف بمرشح $3 \times 3$	13.1
21		عملية التجميع باستخدام التجميع الأقصى بمثال لخطوة واحدة و خطوتان	14.1
22		عملية التسطيح	15.1
23		بنية الشبكة العصبية المتكررة	16.1
24		البنية واحد لواحد لشبكة عصبية متكررة	17.1
24		البنية واحد لكثير لشبكة عصبية متكررة	18.1
25		البنية كثير لواحد لشبكة عصبية متكررة	19.1
25		البنية كثير لكثير غير المباشر لشبكة عصبية متكررة	20.1
26		البنية كثير لكثير المباشر لشبكة عصبية متكررة	21.1
26		الشبكة العصبية المتكررة ثنائية الإتجاه	22.1
28		شبكة عصبية متكررة بعمق 3	23.1
29		خلية الشبكة المتكررة LSTM	24.1
37		نموذجي Skip-gram و CBOW	1.2

39	أهم الأساليب المتبعة في عملية تحليل المشاعر	2.2
54	نموذج Kaggle	3.2
56	النموذج المقترح [KSCD19]	4.2
61	هندسة الشبكة	1.3
	الدقة التي حصل عليها نموذجنا في الأعلى وفي الأسفل نسبة خطأ النموذج لكل من	2.3
66	مجموعات الإختبار و التدريب أثناء تدريب النموذج	
66	قيم مقاييس التقييم المتحصل عليها في الإختبار	3.3
67	مصفوفة الإرتباك	4.3

## قائمة الجداول

37	جدول التوافق . . . . .	1.2
41	أمثلة عن بعض القواعد المعمول بها لتحديد قطبية الجمل . . . . .	2.2
42	مقارنة بين أداء محلل البيانات الأساسي ومحلل البيانات المقترح . . . . .	3.2
48	مثال تطبيقي حول عمل بايز البسيطة . . . . .	4.2
51	توزع الفئات في مجموعات البيانات . . . . .	5.2
52	المعطيات العامة لنماذج التعلم العميق . . . . .	6.2
53	دقة نموذج BiLSTM-LSTM للتصنيف الثنائي في مجموعات البيانات . . . . .	7.2
53	مصفوفة الإرتباك لنموذج BiLSTM-LSTM للتصنيف الثنائي في مجموعات البيانات	8.2
54	دقة نموذج Kaggle للتصنيف الثنائي و الثلاثي في مجموعات البيانات . . . . .	9.2
54	مصفوفة الإرتباك لنموذج Kaggle للتصنيف الثنائي و الثلاثي في مجموعات البيانات .	10.2
	دقة النموذج المقترح و النتائج المقارنة مع النموذجين الأساسيين للتصنيفات الثنائية	11.2
57	و الثلاثية . . . . .	
57	دقة نموذج CNN-LSTM لمجموعات البيانات الأربع في مستويات الترميز الثلاث .	12.2
61	المعطيات المعتمدة في النموذج . . . . .	1.3

# قائمة الخوارزميات

47	.....	تدريب بايزز البسيطة	1
47	.....	اختبار بايزز البسيطة	2

## مقدمة

مع التطور التكنولوجي و ظهور وسائل التواصل الإجتماعي أصبح الناس يناقشون آراءهم و يعبرون عنها بأريحية من خلال المدونات أو التغريدات و يشاركون و يهتمون بآراء الأشخاص الآخرين حول موضوعات سياسية أو منتجات أو خدمات ...إلخ. كما أصبحت الشركات تعتمد على هذه الشبكات الإجتماعية من أجل الترويج لخدماتها و منتجاتها و تبحث عن آراء الناس للتحسين منها ما أدى إلى بيانات نصية كثيرة.

إن توفر هذا الكم الهائل من البيانات جعل من الحاجة لإيجاد طرق لمعالجة هذه البيانات و تصنيفها من حيث استقطاب المشاعر أمرا ضروريا.

كان الإهتمام بمجال تحليل المشاعر منذ القدم، و لكن لم يتم العمل عليه بشكل كبير لقلة البيانات، لذا يمكن ربط ظهور تحليل المشاعر بظهور مواقع التواصل الإجتماعي، حيث بدأ العمل بشكل كبير باستعمال العديد من التقنيات كالأنظمة المستندة على القواعد [NH04] و أنظمة التعلم الآلي (Naive Bayes, SVM, ...). [PLV02]. و كما هو الحال مع العديد من المجالات الأخرى فإن ظهور و تطور أساليب التعلم العميق ساهمت بشكل كبير في حل مشاكل معالجة اللغة الطبيعية التي كانت تسبب معيقات حالت دون تقدم مجال تحليل المشاعر.

و بناء على الدراسة البيليوغرافية التي قمنا بها و التي شملت أمثلة عن نماذج لتحليل المشاعر من بينها نموذج BiLSTM-CNN، قمنا بإقتراح شبكة عصبية عميقة تعتمد على شبكات الذاكرة طويلة-قصيرة المدى ثنائية الإتجاه و العادية و الشبكات العصبية التلافيفية. نتائج الدراسة التجريبية التي قمنا بها و المطبقة على مجموعة بيانات Arabic ULMFiT Model أفضت إلى نتائج جيدة فقد حققت دقة بلغت % 95.

تم تقسيم مذكرة البحث إلى ثلاث فصول:

يهتم الفصل الأول بالشبكات العصبية الاصطناعية و كيفية عملها بصفة عامة، ثم التعلم العميق و بعض أنواعه.

أمّا الفصل الثاني فيتناول بعض المفاهيم الأولية حول تحليل المشاعر و بعض الدراسات التي أجريت حوله، حيث درسنا نمودجا للأنظمة المستندة على القواعد و نمودجي بايز البسيطة و LSTM-CNN في الأنظمة المستندة على التعلم الآلي، ليختتم الفصل بعد ذلك ببعض التحديات التي تواجه مجال تحليل المشاعر.

يناقش الفصل الثالث التجارب المجرأة لإنجاز نمودج تحليل المشاعر للغة العربية القائم على التعلم العميق بالإضافة للنتائج المتحصل عليها.

و أخيراً، الخاتمة تلخص كل ما تطرقنا إليه.

# الفصل الأول

## مدخل للتعلم العميق Deep Learning

### 1.1 مقدمة

التعلم العميق هو فرع من فروع التعلم الآلي يعتمد على الشبكات العصبية الاصطناعية المستوحاة من هيكل ووظائف الدماغ البشري، أحدث ضجة كبيرة في مجال الذكاء الاصطناعي منذ ظهوره خاصة عام 2006 [HOT06]، فقد أتى بحلول لمشاكل النهج التقليدي الذي يعتمد على التصميم اليدوي للميزات.

التعلم العميق هو عبارة عن شبكة عصبية عميقة، يتمتع بأداء متميز في العديد من المجالات كالتعرف على الكلام، معالجة اللغة الطبيعية و تحليل الصور، كما أنه مناسب لتحليل البيانات الضخمة لأنه يتطلب القليل من الهندسة اليدوية [LBH15].

في هذا الفصل سنشرح مفهوم الشبكات العصبية الاصطناعية و كيفية عملها ثم سنتطرق للحديث عن التعلم العميق و بعض انواعه.

### 2.1 الشبكات العصبية Neural Networks

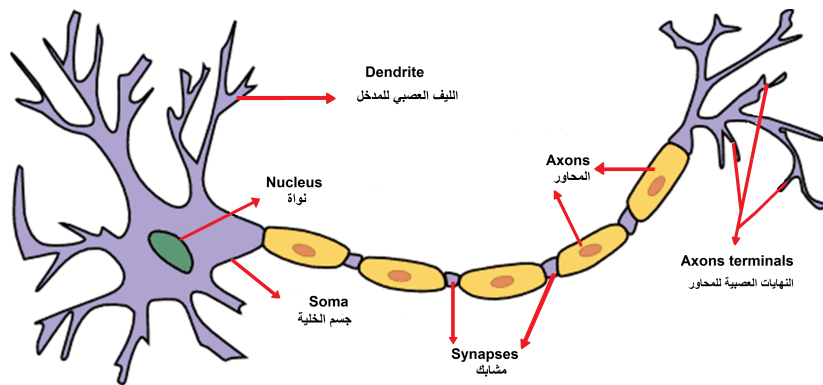
الشبكات العصبية هي نموذج من نماذج التعلم الآلي مستوحاة من بنية ونشاط الدماغ البشري لإنشاء برنامج كمبيوتر يتعلم من البيانات. ظهر مفهومها لأول مرة في أواخر القرن التاسع عشر (1890) من قبل عالم النفس W. James وتم بدء العمل عليها في أربعينيات القرن العشرين من طرف العالمين J. Mc Culloch و W. Pitts [Tou92] و مازالت المفاهيم

التي طورها الرواد تشكل مجال بحث ليومنا هذا.

نعلم أن الدماغ البشري يتكون من  $10^{11}$  خلية عصبية (Neurons) [Kri07] ترسل إشارات فيما بينها، تتصل كل خلية على الأقل بعشرة آلاف أخرى. من هنا وضع العلماء الفرضيات و الخوارزميات لمحاولة محاكاة وظيفة الخلايا العصبية فكان لابد من الربط بين الشبكة العصبية الحيوية و الشبكة العصبية الاصطناعية [MP43].

### 1.2.1 الشبكات العصبية الحيوية (البيولوجية) Biological Neural Networks

تتكون الشبكة العصبية الحيوية من مجموعة من الخلايا العصبية حيث كل خلية تتكون من جسم خلية (Soma) و ليف عصبي لمدخل الخلية (Dendrite) و الذي يستقبل معلومات من الخلايا المجاورة و ينقلها إلى مخارج هذه الخلية عن طريق المشابك (Synapses) التي هي عبارة عن تلامس بين هذا الليف و المحاور (Axons) التي تحمل النبضة الخارجة من الخلية إلى خلايا أخرى (الشكل 1.1).



شكل 1.1: خلية عصبية حيوية Biological Neuron

إذا كان مجموع النبضات اللحظي عند المدخل لا يصل إلى مستوى حد معين (Threshold) لإثارة الخلية فإن الخلية تظل خامدة، و إذا زاد المجموع عن الحد المعين فإنها تبدي نشاطا عن طريق إعطاء نبضة عند المخرج.

### 2.2.1 الشبكات العصبية الاصطناعية Artificial Neural Networks

الخلية العصبية الاصطناعية هي نموذج حسابي يعتمد على هيكل و وظائف الشبكات العصبية الحيوية. تتشكل الشبكات العصبية عادة من طبقات (Layers) متعددة كما هو



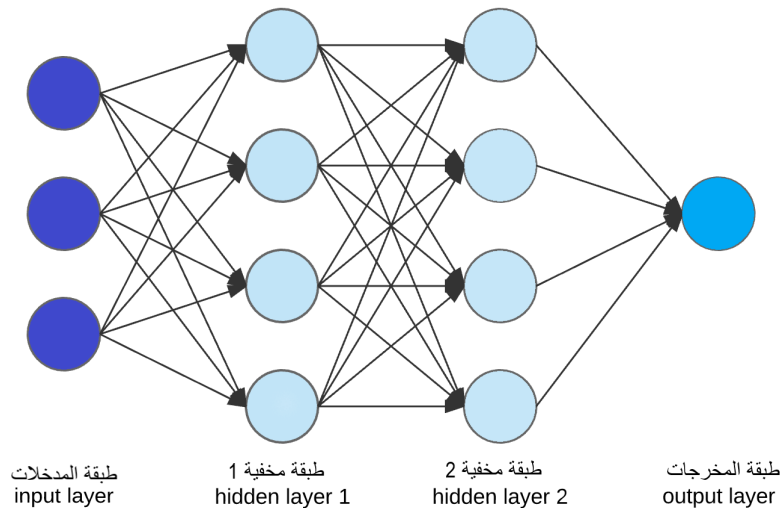
موضح في الشكل (2.1).

تتكون كل من هذه الطبقات من مجموعة من العقد (Nodes)، حيث تتلقى كل عقدة بيانات من العقد في الطبقة السابقة، وتمر تلك البيانات إلى عقد الطبقات الموالية.

**طبقة المدخلات Input layer:** وهي البيانات الأولية للشبكة العصبية، تقوم هذه الطبقة بنقل المعلومات (الميزات) إلى الطبقات المخفية.

**الطبقات المخفية Hidden layers:** هي الطبقات الوسيطة بين طبقة المدخلات و طبقة المخرجات و هي المكان الذي يتم فيه إجراء جميع العمليات الحسابية.

**طبقة المخرجات Output layer:** هي الطبقة التي تعطي نتائجاً لمدخلات معينة.



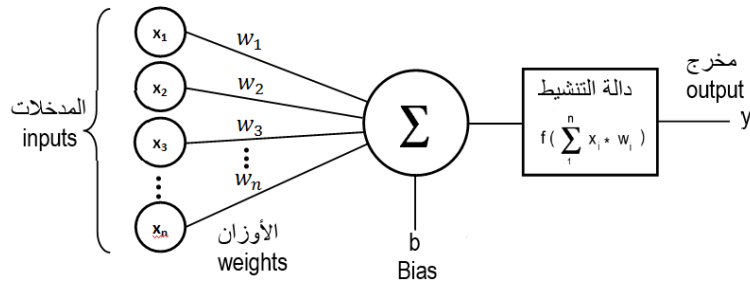
شكل 2.1: شبكة عصبية اصطناعية Artificial Neural Network

تجدر الإشارة أنه لكل مدخل من مدخلات الشبكة العصبية  $x_i$  وزن مقترن ( $w_i$  Weight) يتم تعيينه على أساس أهميته النسبية للمدخلات الأخرى و الذي يشير إلى حجم الدور الذي سيلعبه مخرج العقدة الواحدة في الحساب الذي تنفذه العقدة التالية (الشكل 3.1).

إن تدريب الشبكة يتمحور حول عملية تحديد تلك الأوزان، حيث أنه في الخطوة الأولى يتم جمع المدخلات المضروبة في أوزانها مع قيمة الإنحياز ( $b$ ) Bias الذي يمثل الثابت الذي يساهم في جعل النموذج أكثر مرونة بحيث يتناسب بشكل أفضل مع البيانات المعطاة،

أما في الخطوة الثانية فيكون تطبيق دالة من دوال التنشيط  $f$  على الناتج، كما هو موضح في المعادلة (1.1).

$$Y_{output} = f\left(\sum_1^n (x_i * w_i) + b\right) \quad (1.1)$$



شكل 3.1: خلية عصبية اصطناعية Artificial Neuron

### 3.2.1 دوال التنشيط

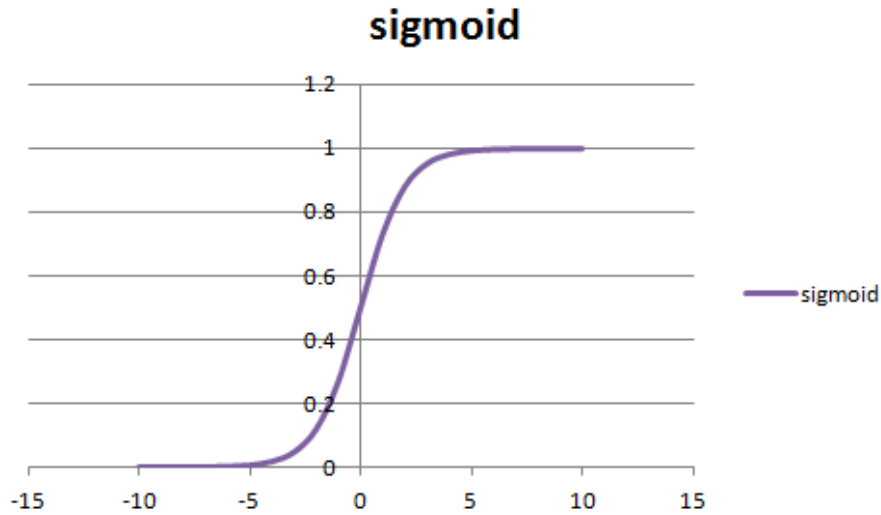
تقوم دوال التنشيط بحديد ما إذان كانت المعلومات (الوزن، الإنحياز) التي تلقاها العصب من الطبقة السابقة كافية لتنشيط الخلايا العصبية أم لا أي أن قيمتها ضمن مجال العتبة (Threshold) المحددة لتنشيط الخلية، ثم يتم إرسال الناتج إلى الطبقة التالية من الخلايا العصبية كمدخلات.

يتم اختيار دالة التنشيط بناء على ماهية المشكل، و يوجد عدة دوال تنشيط من بينها:

#### الدالة الأسية أو السجماوية Sigmoid

تعتبر من أشهر الدوال إستعمالا (المعادلة (2.1))، الدالة الأسية هي دالة غير خطية تستخدم في حل مشاكل التصنيف الثنائية، حيث أنها تعطي قيما بين ال0 و ال1. إذا وصل وزن القيمة للحد الأدنى للعتبة (Threshold) فإنها تعطي القيمة 0 و إذا وصل للحد الأعلى تعطي القيمة 1 (الشكل 4.1).

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2.1)$$

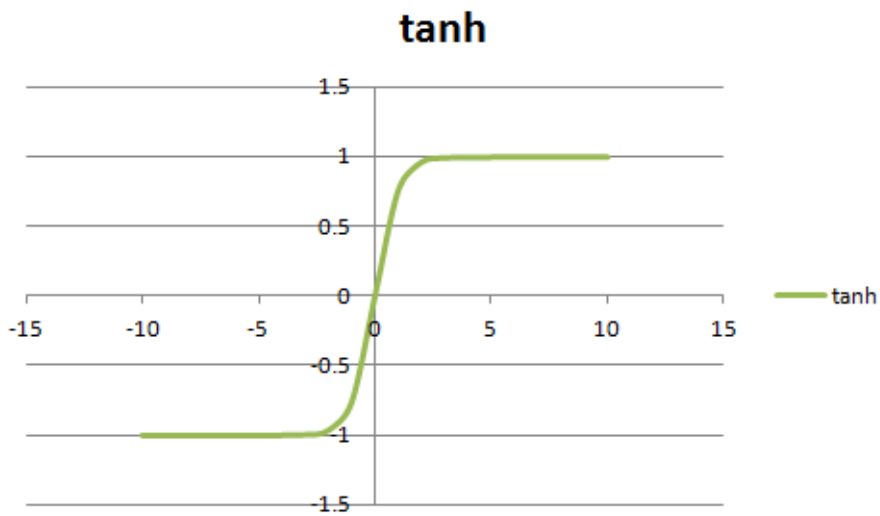


شكل 4.1: منحنى الدالة Sigmoid

## دالة الظل الزائدي (Tanh) Tangent Hyperbolic (Tanh)

دالة تنشيط غير خطية (المعادلة (3.1))، مشابهة للدالة الأسية إلا أن قيمها تتراوح بين 1- و 1 (الشكل 5.1) و بالتالي فإن المدخلات السلبية تعطي نتائج سلبية و يتم تعيين المدخلات ذات القيم الصفرية فقط إلى مخرجات قريبة من الصفر.

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{1 - e^{-2x}}{1 + e^{-2x}} \quad (3.1)$$

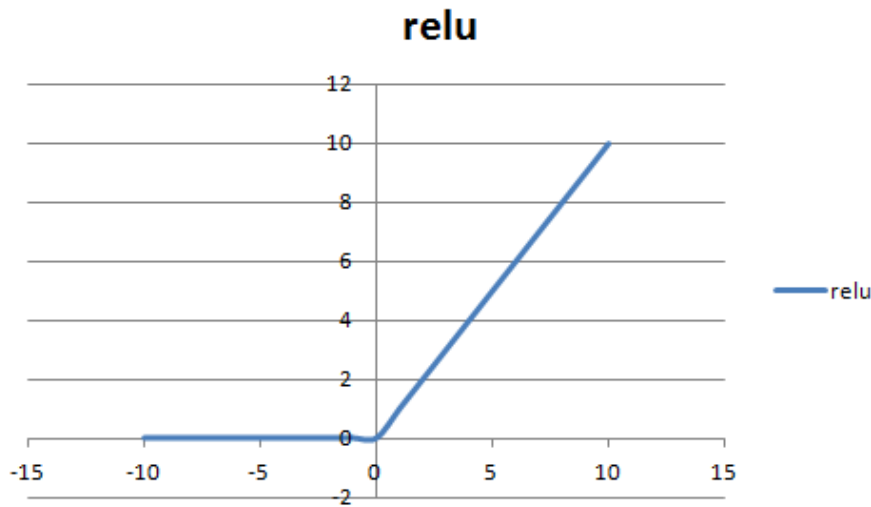


شكل 5.1: منحنى الدالة Tanh

## الدالة ReLU(Rectified Linear Unit)

دالة خطية بالنسبة للقيم الموجبة و منعدمة بالنسبة للقيم السالبة (الشكل 6.1)، بمعنى آخر فإنها تعطي القيمة الأكبر بين 0 و  $(\sum_1^n(x_i * w_i) + b)$ . تلغي الدالة ReLU أهمية القيم السالبة، و تعطي أهمية متباينة للقيم الموجبة.

$$f(x) = \max(0, x) \quad (4.1)$$



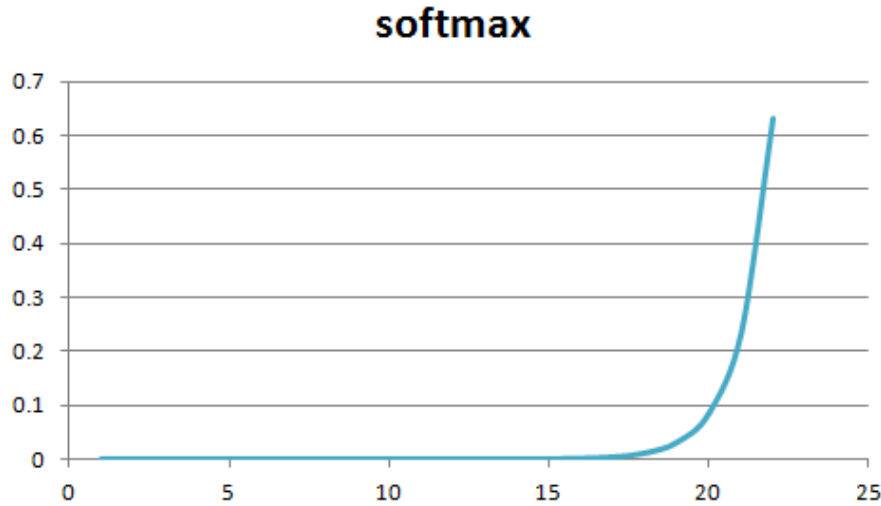
شكل 6.1: منحنى الدالة ReLU

## الدالة Softmax

هي تحسين للدالة الأسية، تستخدم في حالات التصنيف المتعددة ( التي تحتوي على أكثر من فئتين ) لحساب توزيع الاحتمالات لكل فئة، حيث تعطي ناتجا محصورا بين ال0 و ال1 (الشكل 7.1) تتم قسمته على مجموع المخرجات. تعطي عبارتها في المعادلة (5.1).

$$f(x_i) = \frac{e^{x_i}}{\sum_{c=1}^k e^{x_c}} \quad (5.1)$$

k: عدد الفئات.



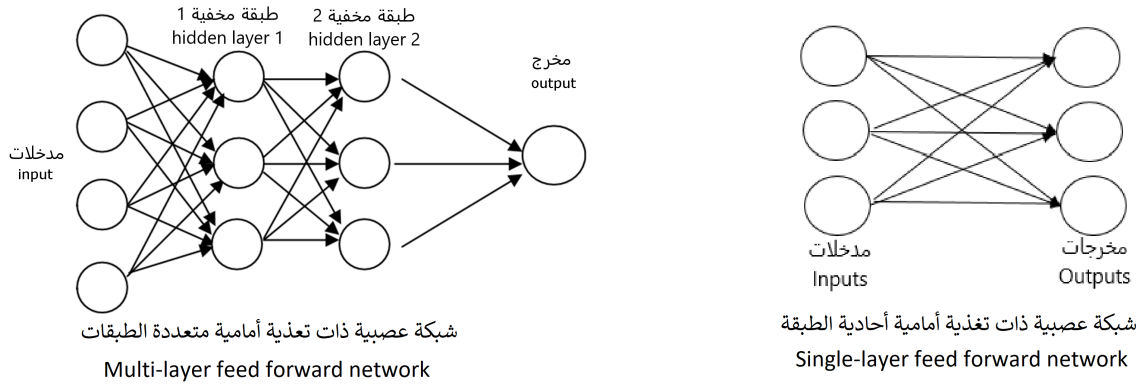
شكل 7.1: منحني الدالة Softmax

#### 4.2.1 أنواع الشبكات العصبية الاصطناعية

هندسة الشبكة العصبية مهمة للغاية في الحصول على أداء جيد لها. يلعب اختيار نوع الشبكة المناسب لحل مشكل معين دورا هاما في فعاليتها حيث توجد عدة أنواع للشبكات العصبية الاصطناعية نذكر منها:

##### الشبكات العصبية ذات التغذية الأمامية Feed-Forward networks

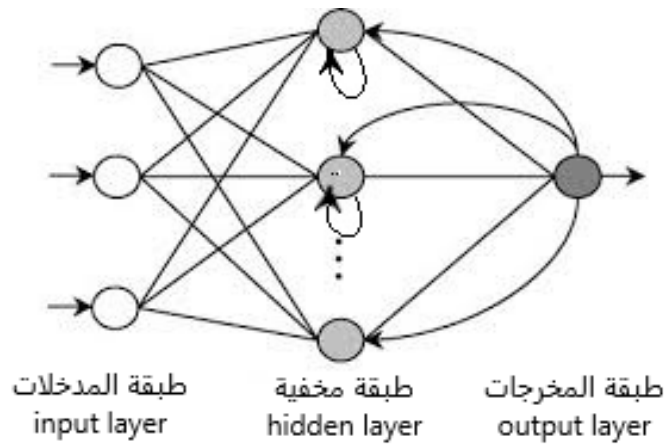
تعد من أكثر الشبكات العصبية إستخداما. هذا النوع من خوارزميات الشبكة العصبية يتكون على الأقل من طبقتين (الشكل 8.1). يمرر المعلومات  $x_i$  في إتجاه واحد إلى الأمام من طبقة الإدخال عبورا بالطبقات المخفية التي تجرى فيها الحسابات وصولا إلى طبقة الإخراج منتجتا المخرجات  $y_i$ . تتجنب هذه الشبكات الوقوع في حلقة مغلقة.



شكل 8.1: الشبكات العصبية ذات التغذية الأمامية

### الشبكات العصبية ذات التغذية الخلفية Feed-back networks

هذا النوع من الشبكات يحوي على الأقل طريقا خلفيا لمخرجاته  $y_i$  لتصبح مدخلات  $x_i$ ، مكونا بذلك حلقة تغذية خلفية لتعطي نتائج أفضل، وقد يكون هناك تغذية ذاتية أي أن مخرجات الخلية العصبية تعود لتصبح مدخلات لها (الشكل 9.1).



شكل 9.1: الشبكات العصبية ذات التغذية الخلفية

للإطلاع على أنواع أخرى من الشبكات العصبية الإصطناعية يمكن النظر في [AS98, ش13].

### 5.2.1 تعليم الشبكات العصبية الاصطناعية

نستخدم بيانات التدريب (Training set  $\{(x_1, y_1), \dots, (x_i, y_i)\}$ ) لكي ندرّب مجموعة الشبكات العصبية. هذه البيانات عبارة عن أمثلة مختارة بعناية من أجل المساهمة في سرعة تعلم الشبكة.

تنقسم طرق تعليم الشبكات العصبية إلى قسمين:

#### التعلم بإشراف Supervised learning

يمكن الإشارة إلى التعلم بإشراف على أنه تصنيف، وهو أسلوب تعليمي يحدد معطيات الشبكة الاصطناعية من بيانات التدريب التي تعتبر "المشرف" في هذه الحالة. تتمثل مهمة الشبكة العصبية في تحديد قيم معطيات الإخراج  $y_i$  بناء على قيم معطيات الإدخال  $x_i$ . يتم تزويد الشبكة بالإجابة الصحيحة (المخرجات) لكل قيمة من قيم معطيات الإدخال. تضبط الأوزان بحيث يكون الفرق بين الإجابات المتوقعة و الإجابات الحقيقية (دالة الخطأ Loss function) ضئيل قدر الإمكان أو معدوم. من أشهر الخوارزميات شيوعاً في هذه الفئة، خوارزمية الانتشار الخلفي Back propagation.

#### التعلم بدون إشراف Unsupervised learning

نسعى في التعلم بدون إشراف إلى تحديد كيفية تنظيم البيانات. هذه الفئة من التعلم لا تتطلب إجابات صحيحة مرتبطة ببيانات الإدخال في بيانات التدريب.

تقوم الشبكة باستكشاف الميزات و الأنماط في البيانات و الإرتباطات بين الأنماط اعتماداً على بيانات الإدخال مشكلة فئات من هذه الإرتباطات. لذلك لا يمكن الجزم بأصحية الناتج لأنه لا يوجد لدينا خصائص عن الناتج المرغوب.

### 6.2.1 خوارزميات تعلم الشبكة العصبية

في التعلم بإشراف، تدريب الشبكة هو عبارة عن عملية تكرار للذهاب و العودة في الشبكة، نقوم خلالها بمقارنة النتيجة المتحصل عليها و النتيجة الحقيقية، و تكون هذه المقارنة باستخدام دالة الخطأ Loss function لقياس مدى جودة التنبؤ في الشبكة.

يعبر عن عملية الذهاب بالانتشار الأمامي Forward propagation و عملية العودة بالانتشار الخلفي Back propagation.

### عملية الانتشار الأمامي Forward propagation

هي إنتشار بيانات الإدخال الأولية (Inputs) في طبقات الشبكة المخفية، بعد تطبيق الخلايا العصبية حساباتها على هذه البيانات نصل إلى الطبقة النهائية كنتيجة للتنبؤ (Output) كما هو مبين في المعادلة (1.1).

### دالة الخطأ Loss function

تعتبر دالة الخطأ عاملاً مهماً في تحديد مدى قرب الشبكة العصبية من الوزن المثالي، فهي تساعدنا على حساب الفرق بين النتائج الحقيقية و النتائج المتحصل عليها، و الذي نعبر عنه بنسبة الخطأ و هو ما نهدف إلى تقليله.

هناك عدة طرق لحساب دالة الخطأ منها متوسط مربع الخطأ (Mean Square Error) الموضحة في المعادلة (6.1) و جذر متوسط مربع الخطأ (Root Mean Square Error) الموضحة في المعادلة (7.1) و Binary cross entropy الموضحة في المعادلة (8.1).

$$MSE = \frac{1}{n} \sum_{j=1}^n (y_i - \hat{y}_i)^2 \quad (6.1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_i - \hat{y}_i)^2} \quad (7.1)$$

$$BCE = -(y * \log(\hat{y}) + (1 - y) * \log(1 - \hat{y})) \quad (8.1)$$

$n$ : عدد بيانات الإدخال Training set.

$y_i$ : النتائج الحقيقية،  $i$  هو عدد الفئات.

$\hat{y}_i$ : النتائج المتحصل عليها بعد التنبؤ



## عملية الانتشار الخلفي Back propagation

هو طريقة لتدريب الشبكات العصبية الاصطناعية تعتمد على ضبط الأوزان إستنادا على نسبة الخطأ. يستخدم الانتشار الخلفي بعض خوارزميات التحسين كخوارزمية نسب التدرج (gradient descent) من أجل الوصول لأقل قيمة ممكنة لدالة الخطأ . تعرف معادلة نسب التدرج كالاتي :

$$w = w - \alpha \frac{\partial j(w)}{\partial w} \quad (9.1)$$

$w$  : الوزن

$j(w)$  : دالة الخطأ

$\alpha$  : معدل التعلم

معدل التعلم هو عامل مهم يتحكم في مقدار ضبط أوزان الشبكة، كلما انخفضت قيمته زادت فترات التدريب لأن تغييرات الأوزان تكون بطيئة، و كلما ارتفعت كانت التغييرات في الاوزان أسرع و بالتالي تقل فترات التدريب.

تكون قيمة معدل التعليم موجبة و يعتبر اختيارها خطوة صعبة للغاية لذا غالبا ما تؤخذ ما بين 0 و 1.

تقوم خوارزمية نسب التدرج بحساب الوزن الجديد عن طريق إضافة قيمة للوزن السابق، حيث تمثل هذه القيمة مشتقة دالة الخطأ بالنسبة للوزن مضروبة بمعدل التعلم. يرتبط تغير قيمة الوزن بإشارة المشتقة، بحيث إذا كانت إشارة المشتقة سالبة فإن قيمة الوزن تزداد و العكس.

## 7.2.1 الإفراط في التعلم و التفريط في التعلم Overfitting and Underfitting

لبناء نموذج دقيق قدر الإمكان يجب تجنب الوقوع في مشاكل الإفراط و التفريط في التعلم من أجل ذلك يجب فهم ماهية أخطاء التنبؤ و المتمثلة في الإنحياز و التباين (variance)

التي تساعدنا في تقييم النموذج و بالتالي قد نستطيع تفادي هذه المشاكل.  
الانحياز: هو الفرق بين متوسط القيم المتحصل عليها من التنبؤ و القيم الحقيقية.

$$Bias(\hat{y}) = E[\hat{y}] - y \quad (10.1)$$

التباين: هو الفرق في التوافق بين بيانات التدريب و بيانات التحقق.

$$Var(\hat{y}) = E[(E[\hat{y}] - \hat{y})^2]. \quad (11.1)$$

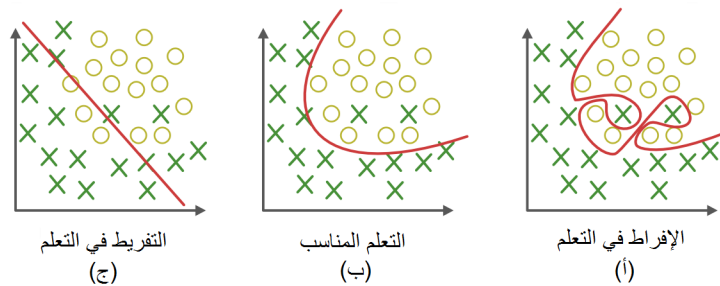
بيانات التحقق هي البيانات التي يقيم بها النموذج أثناء التدريب.

### الإفراط في التعلم

هو الحالة التي يتعلم فيها النموذج بشكل مفرط من بيانات التدريب (الحالة أ) في الشكل (10.1)، بحيث لا يكون قادرا على تعميم التنبؤ كونه يلتقط التقلبات العشوائية و الضوضاء لهذه البيانات و يتعلمها كمفاهيم فيؤثر ذلك على أداء النموذج لأن هذه المفاهيم لا تنطبق على البيانات الجديدة أي في حالة إعطاء النموذج مثال غير موجود في بيانات التدريب فإنه لا يستطيع تصنيفه.

### التفريط في التعلم

عكس الإفراط في التعلم، التفريط في التعلم يحدث عندما لا يكون النموذج قادرا على استخراج الميزات من بيانات التدريب بشكل صحيح و لا تعميم التنبؤ، أي أنه لم يتعلم بالقدر الكافي من هذه البيانات (الحالة ج) الشكل (10.1). قد ينتج التفريط في التعلم عن محاولة استعمال نموذج بسيط للغاية.



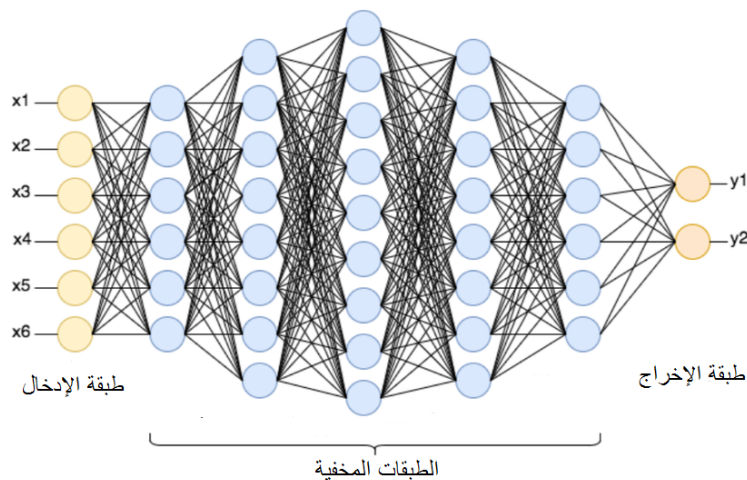
شكل 10.1: الإفراط في التعلم و التفريط في التعلم

يمكن أن يؤدي ارتفاع قيمة الإنحياز إلى الوقوع في مشكل التفريط في التعلم، لذلك إذا كانت قيمته عالية نستخدم بعض الطرق لتخفيضها كزيادة عدد الطبقات المخفية في الشبكة، تدريب النموذج لفترة أطول أو تغيير هندسة الشبكة بما يلائم. كذلك بالنسبة لارتفاع قيمة التباين فإنه يؤدي إلى الوقوع في مشكل الإفراط في التعلم، لذا نحاول تخفيضه في حالة ارتفاعه وهذا إما عن طريق زيادة عدد البيانات أو التسوية (regularization)، أو البحث عن هندسة أخرى ملائمة.

قد تؤدي المحاولات في إخفاض أحدهما إلى ارتفاع قيمة الآخر، على سبيل المثال إذا أردنا تخفيض التباين عن طريق زيادة البيانات قد يؤدي بنا هذا إلى ارتفاع قيمة الإنحياز، لذلك لا نستطيع الجزم بتأثير الطرق السابقة على قيمتي الإنحياز و التباين، لذا من الأفضل لتفادي مشكلتي التفريط و الإفراط في التعلم تؤخذ بعين الإعتبار الزيادة و النقصان في قيمتي الإنحياز و التباين معا.

### 3.1 التعلم العميق

التعلم العميق هو ترجمة قريبة من deep neural network أو الشبكات العصبية العميقة، و نقصد بالعميقة أنها تحتوي على طبقات مخفية عديدة (الشكل 11.1). كما أظهرت أبحاث أنه كلما زاد العمق كان أداء النموذج أفضل، لأن الزيادة في الطبقات تسمح باستخراج ميزات أكثر من البيانات وبالتالي التعلم بشكل أحسن [GBC16].



شكل 11.1: شبكة عصبية عميقة

### 1.3.1 التحسين في التعلم العميق Optimization in deep learning

يوجد عدة عوامل يعتمد عليها أداء نموذج الشبكة العصبية، و المتطرق إليه بشكل كبير عادة هو اختيار هندسة أو هيكل لهذه الشبكة، إلا أنه يمكن مواجهه عدة مشاكل كذلك من بينها صعوبة تدريب هذه الشبكة و التي تعتبر عملية مهمة و مكلفة في الوقت نفسه فقد تستغرق من أيام إلى شهور، لذلك عادة ما يتم البحث عن عامل من شأنه تقليل هذا الوقت قدر الإمكان من جهة و تحسين أداء النموذج من جهة أخرى، يمكن أن يكون هذا العامل هو المحسنات و الهدف منه تخفيض قيمة دالة الخطأ.

هناك عدد كبير من المحسنات نذكر منها:

#### محسن التدرج النسبي Gradient descent

أكثر خوارزميات التحسين شيوعاً يتمثل دوره في إيجاد أفضل القيم لدالة الخطأ. تستخدم خوارزمية التدرج النسبي كمحسنات صندوق أسود حيث يصعب الحصول على تفسيرات عملية لنقاط القوة و الضعف لها، الفكرة العامة لهذا المحسن هي تهيئة المعطيات إلى قيم عشوائية ثم اتخاذ خطوات صغيرة في اتجاه "المنحدر" في كل تكرار.

يوجد عدة أنواع لخوارزمية التدرج النسبي ك stochastic gradient descent, batch gradient descent, mini batch gradient descent.

#### محسن Adam

خوارزمية "Adam" ليست إختصاراً و ليست مكتوبة بإسم Adam، فإسمها مشتق من "تقدير اللحظة التكيفية adaptive moment estimation" [GBC16] أثبتت الدراسات أن خوارزمية التحسين هذه عملية أكثر و ذات أداء جيد و تقارن بشكل إيجابي مع أنواع المحسنات العشوائية (stochastic optimization) الأخرى [KB14]، و يمكن استخدامها بدلا من إجراء التدرج النسبي العشوائي (stochastic gradient descent) لتحديث أوزان الشبكة التكرارية بناء على بيانات التدريب، فهي طريقة سهلة التنفيذ و ذات كفاءة حسابية جيدة و لا تأخذ متطلبات كبيرة للذاكرة كما أنها مناسبة للمشكلات الكبيرة من حيث

البيانات و / أو المعطيات [KB14].

### 2.3.1 التسوية في التعلم العميق Regularization in deep learning

يعد الإفراط في التعلم من أكثر المشاكل خطورة في عملية تدريب الشبكات العصبية العميقة ذات البيانات الكبيرة و الطبقات الكثيرة. أحد حلول هذه المشكلة هو التسوية، و قد عرفت في [GBC16] بأنها "أي تعديل نقوم به على خوارزمية تعلم يهدف إلى تقليل خطأ الإختبار و ليس خطأ التدريب". و نجد تعريفاً آخر لها في [KGC17] على أنها "أي تقنية تكميلية تهدف إلى جعل النموذج العام يعمم بشكل أفضل، أي تحقيق نتائج أفضل على مجموعة الإختبار، و يمكن أن يشمل ذلك خصائص مختلفة لدالة الخطأ و خوارزميات تحسين الخطأ أو تقنيات أخرى"، و سنعالج فيما يلي بعض تقنيات التسوية.

#### التسرب Dropout

واحد من أكثر تقنيات التسوية استخداماً و إثارة للإهتمام. هو عملية تكرارية تختار فيها في كل مرة عقداً معينة بشكل عشوائي، إما من طبقات الإدخال أو من الطبقات المخفية و يحذف جميع اتصالاتها الواردة و الصادرة ما ينتج عنه في كل مرة مجموعة مختلفة من العقد و بالتالي مجموعة مختلفة من المخرجات. يتم تعيين احتمالية الإحتفاظ بالوحدة عن طريق قيمة احتمالية  $p$ ، بحيث كلما كثر العقد نعطي قيمة أقل ل  $p$  و من ناحية أخرى يمكن تعيين  $p$  أعلى للطبقات التي تحتوي على عدد أقل من العقد، لأنها أقل عرضة للإفراط في التعلم، و قد تصل هذه القيمة إلى ال 1 عندما نريد الإحتفاظ بجميع الوحدات كما هو الحال بالنسبة لطبقات الإدخال لأننا لا نريد حذف الكثير من ميزات الإدخال [cou].

#### التوقف المبكر Early stopping

هي طريقة بسيطة، فعالة و كثيرة الإستخدم في تدريب الشبكات العصبية، الهدف منها هو تدريب البيانات لفترة كافية تجعل النموذج قادراً على التعلم دون الوقوع في مشكل الإفراط في التعلم. يتوقف التدريب عندما يبدأ الأداء في التدهور في مجموعة التحقق و يبدأ خطأ التعميم في الزيادة. نقوم باختيار مقياس للأداء و يتم حسابه في كل فترة (epoch) من فترات التدريب، و ليس بالضرورة أن تكون الفترة التي يحدث فيها أول تدهور لهذا المقياس هي الوقت المناسب لإيقاف التدريب فبعض التأخر في التوقف يعد فكرة جيدة.

نقوم بتخزين معطيات النموذج في كل فترة تدريب يتحسن فيها المقياس. عند انتهاء خوارزمية التدريب نعيد هذه المعطيات بدل معطيات آخر فترة تدريب قبل التوقف. [Bro18]

### زيادة البيانات Data augmentation

هي الزيادة في بيانات التدريب عن طريق إحداث تغييرات في هذه البيانات (مثال: تدوير، تقلبات أو تحجيم في الصور الأصلية،...) دون اللجوء إلى البحث عن بيانات جديدة فهذا يمكن أن يكون مكلفا للغاية أو غير متوفر فذلك لا يكون خيارا متاحا في أغلب الأحيان.

يمكن اعتبار زيادة البيانات طريقة مستقلة عن بنية الشبكة ذلك لأنه يتم تطبيقها على بيانات التدريب كما توضح بعض الدراسات أنها قادرة على التكيف مع تغييرات بنية الشبكة و كمية بيانات التدريب و أنه يمكن أن يكون لها أداء جيد بنفس أداء النماذج المستوية (regularized models) أو أفضل [HGK18].

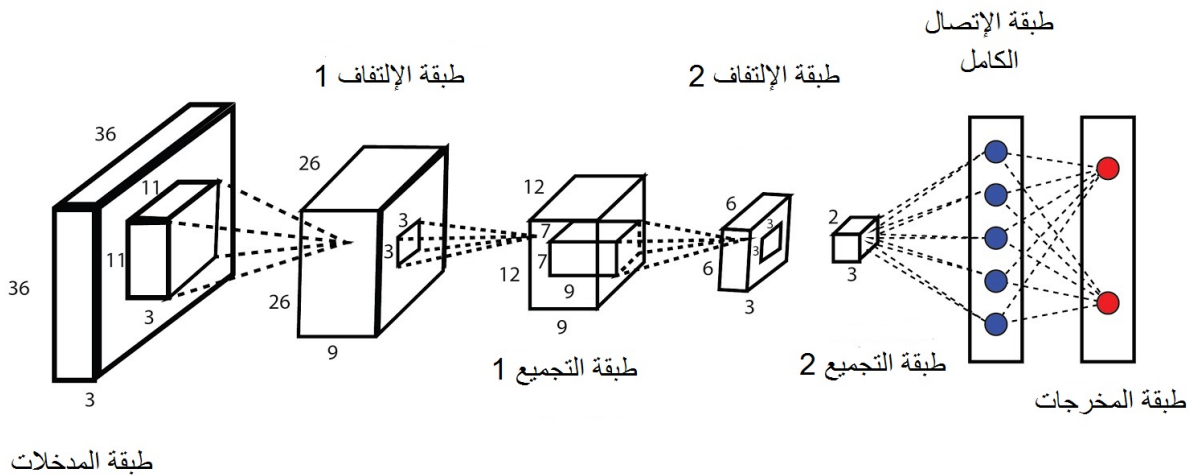
### 3.3.1 الشبكات العصبية التلافيفية Convolutional Neural Networks (CNNs)

الشبكات العصبية التلافيفية او (CNN) هي مجموعة فرعية من التعلم العميق تستخدم في معالجة البيانات التي لها طوبولوجيا تشبه الشبكة كالصور على سبيل المثال لا الحصر، و هي شبكة ذات تغذية أمامية و تعتبر من أولى النجاحات في مجال التعلم الآلي، استلهمت من تجارب Hubel Wiesel على القشرة البصرية للقط حيث لوحظ أن أجزاء معينة من المجال البصري تثير خلايا عصبية معينة [HW59].

برزت الشبكات العصبية التلافيفية في مسابقة ImageNet السنوية عام 2012 (و التي يشار إليها أيضا ب [ILSVRC] ImageNet Large Scale Visual Recognition Challenge) عند ظهور شبكة AlexNet التي حققت نجاحا بفارق كبير عن نظيراتها في هذه المسابقة [KSH12]، و أثبتت دراسات عام 2016 بأن الشبكات العصبية التلافيفية أسهل في التحسين و أنه كلما كانت الشبكة أكثر عمقا أعطت نتائج أدق [HZRS16].

سميت الشبكات العصبية التلافيفية بهذا الإسم لأنها تستخدم عملية رياضية تدعى الالتفاف (convolution) تسمح بتداخل وظيفتين أثناء تفاعلها و ذلك في الطبقات الأولى من الشبكة،

تتبعها طبقة التجميع (pooling) و يتم استخدام طبقات متصلة بالكامل (fully connected layers) في الطبقات الأخيرة (الشكل 12.1).



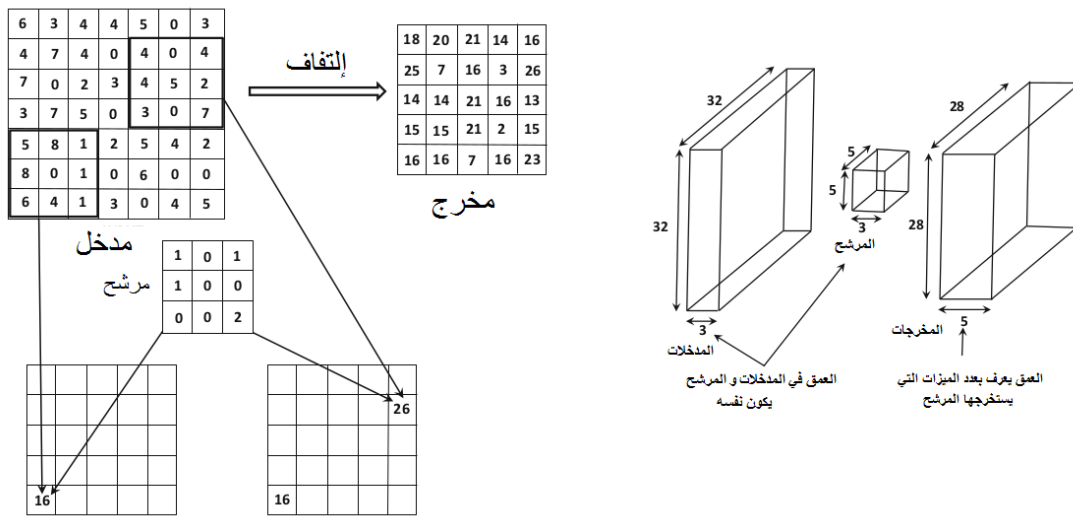
شكل 12.1: الشبكة العصبية التلافيفية

### طبقة الالتفاف Convolutional layer

هي أهم عملية في الشبكة العصبية التلافيفية حيث يتم فيها تقسيم المدخل إلى ثلاث أبعاد (الطول، العرض والعمق)، على سبيل المثال في الصور يمثل الطول و العرض عدد البيكسل أما العمق فيعبر عن عدد خرائط الميزات في الطبقات المخفية، عدد قنوات الألوان الأساسية مثلاً.

يتم تطبيق مجموعة من المرشحات (filters) صغيرة الحجم من الناحية المكانية (الطول و العرض) لكنها تمتد خلال العمق الكامل لحجم الإدخال. نقوم بلف كل مرشح عبر حجم الإدخال مع تطبيق الحسابات اللازمة (الشكل 13.1) فينتج خريطة ميزات خاصة بكل مرشح.

لا يتم إنشاء الخاصية التي تساعد المرشح في التعرف على الميزة يدويا، بل يتم استخراجها من البيانات من خلال خوارزمية التعلم.



شكل 13.1: عملية الالتفاف بمرشح 3\*3

ReLU: هي خطوة من عملية الالتفاف، لا تختلف كيفية تطبيقها في الشبكات العصبية التلافيفية عنه في الشبكات العادية.

تعتبر دالة التنشيط relu الأكثر استخداما مقارنة بالدوال الأخرى (الظل الزائدي و الأسية).

تدريب الشبكة التلافيفية باستخدام relu يكون أسرع و أكثر دقة من استخدام نظيراتها [KSH12]، لذلك أصبح المستخدمون يطبقونها بشكل افتراضي في نماذجهم.

### طبقة التجميع Pooling layer

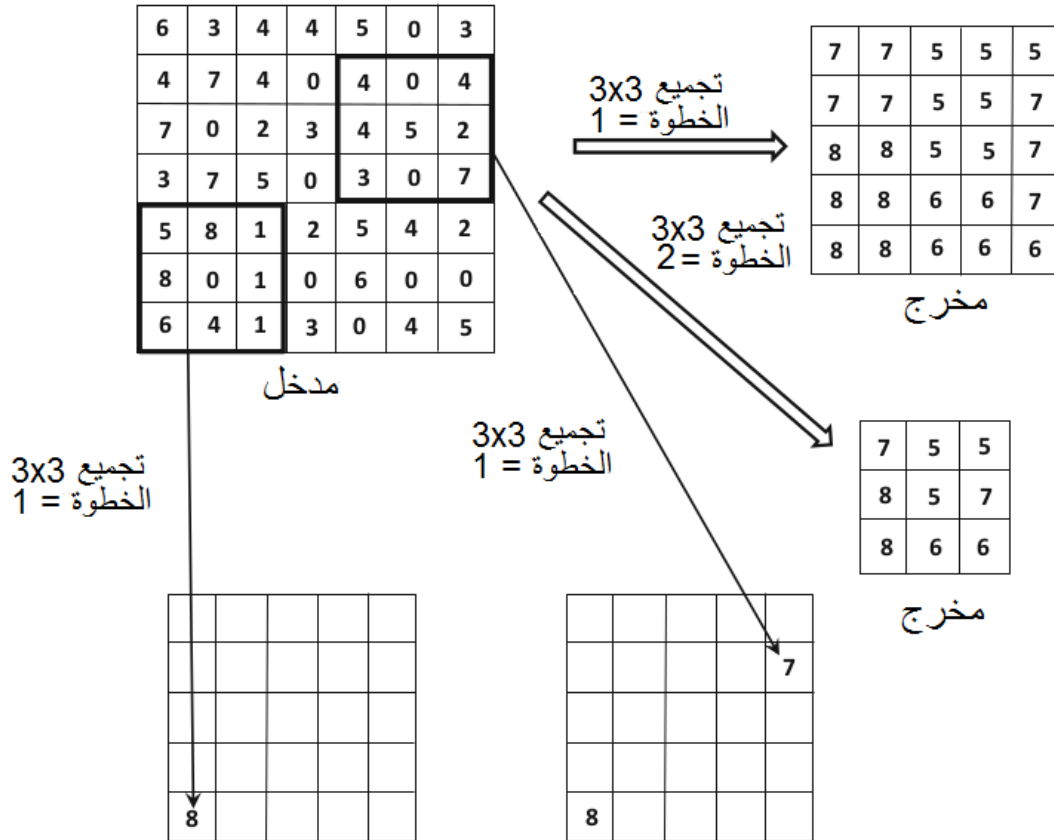
يطبق التجميع على مساحات مربعة  $P*P$  من خريطة الميزات من أجل تقليل كمية الميزات و الحساب في الشبكة، منتجا خريطة ذات بعد مكاني أقل مخفضا بذلك احتمالية الوقوع في مشكل الإفراط في التعلم (الشكل 14.1).

هناك أنواع عديدة من التجميع، فيما يأتي الأكثر استعمالا:

التجميع الأقصى Max pooling: تقوم بإرجاع أكبر قيمة في المساحة المربعة.



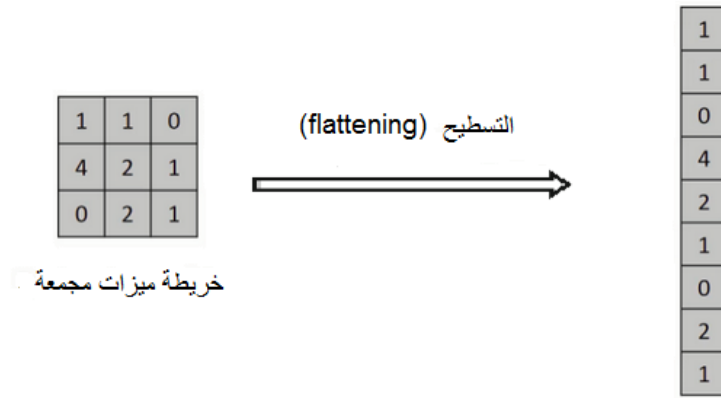
متوسط التجميع Average pooling: يقوم بإرجاع متوسط القيم في المساحة المربعة.  
 التجميع الأقصى الشامل Global Max Pooling: تقوم بإرجاع أكبر قيمة في حجم الإدخال.



شكل 14.1: عملية التجميع باستخدام التجميع الأقصى بمثال لخطوة واحدة و خطوتان

### طبقة الإتصال الكامل Fully connection layer

بعد الإنتهاء من مرحلتي الإلتفاف و التجميع نقوم بإنشاء شعاع (vector) من الميزات الناتجة، و هذا ما يسمى بالتسطيح (flattening) (الشكل 15.1)، يسمح لنا هذا بإدراج هذه الميزات في الطبقة المتصلة بالكامل و التي تمثل الطبقات الأخيرة في الشبكة التي تقود القرار النهائي للتصنيف و تعمل بنفس طريقة شبكة التغذية الأمامية العادية.



شكل 15.1: عملية التسطيح

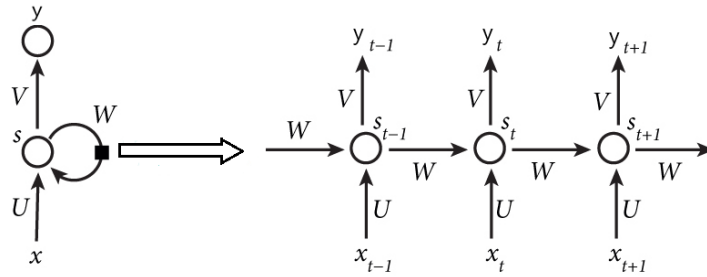
### 4.3.1 الشبكات العصبية المتكررة (RNNs) Recurrent Neural Network

هناك أنواع من البيانات التي تحتوي على تبعيات متسلسلة فيما بينها، و التي يجب أن تؤخذ بعين الإعتبار في معالجة و فهم هذه البيانات مثل (السلاسل الزمنية، النصوص، البيانات البيولوجية،...)، تفيد الشبكات العصبية المتكررة في معالجة هذا النوع من البيانات [LBH15]، لأنها تختلف عن الشبكات العصبية التقليدية التي تكون المدخلات و المخرجات فيها مستقلة عن بعضها، فهي تتميز بخاصية تعرف بالطبقات المتكررة و التي تمكنها من الحفاظ على الحالة بين استخدامات الشبكات، حيث يكون إخراج الخطوة السابقة مدخلا للخطوة الحالية.

يتم تخزين معلومات الخطوات بفضل طبقات الشبكة المتكررة المخفية الخاصة التي تحتوي على خاصية الذاكرة.

#### بنية الشبكة العصبية المتكررة Architecture of RNN

في الشبكات العصبية العادية يكون لكل طبقة مخفية أوزان و إنحيازات خاصة بها لذلك فإن طبقاتها تتصرف بشكل مستقل، على عكس الشبكات العصبية المتكررة التي توفر نفس الأوزان و الإنحيازات للطبقات المخفية، بحيث يمكن دمجها في طبقة متكررة واحدة، هذا يعني أننا نقوم بنفس العملية في كل خطوة مع مدخلات مختلفة (الشكل 16.1).



شكل 16.1: بنية الشبكة العصبية المتكررة

$t$  : الخطوة الزمنية للانتقال من حالة  $s$  إلى أخرى.

$T$  : العدد الإجمالي للخطوات و يمثل أيضا عدد الحالات.

$x_t$  : المدخل في اللحظة  $t$ .

$s_t$  : هي الحالة المخفية في الخطوة  $t$  و تحسب ب:

$$s_t = f(W * s_{t-1} + U * x_t + b_s) \quad (12.1)$$

$f$  : دالة تنشيط عادة ما تكون غير خطية مثل  $\tanh$  أو  $\text{relu}$ .

$W$  : وزن الحالة.

$U$  : وزن المدخل.

$b$  : الإنحياز.

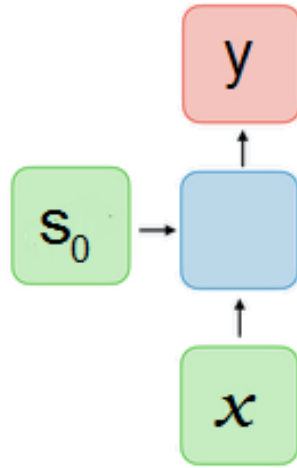
$y_t$  : هو الإخراج عند الخطوة  $t$  و يحسب ب:

$$y_t = f(V * s_t + b_y) \quad (13.1)$$

$V$  : وزن المخرج.

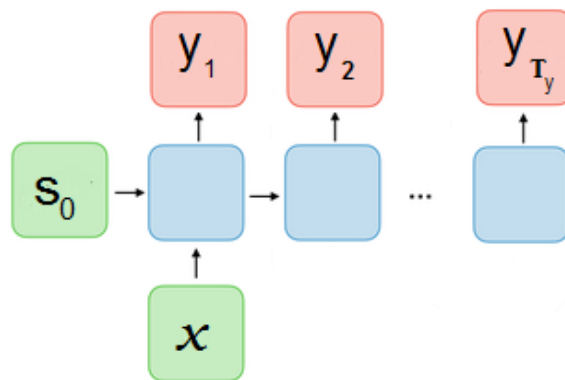
تعالج الشبكات العصبية المتكررة مدخلات ذات حجم متغير، لذا فإن إمكانية ما يمكن لهذه الشبكات معالجته واسعة جدا [ZSSR17]. و تختلف بنية الشبكة على حسب المشكل المراد معالجته، فقد لا يكون ضروريا أن تحتوي كل خطوة على مدخل و مخرج، لذا هناك عدة أنواع لبنية الشبكات العصبية المتكررة [VD18]:

- واحد لواحد **one to one**: هي شبكة عصبية متكررة ذات خطوة زمنية واحدة، تشبه بنيتها بنية الشبكات ذات التغذية الأمامية، حيث يتم تعيين إدخال واحد إلى إخراج (الشكل 17.1). تستعمل لمعالجة البيانات غير المتسلسلة.



شكل 17.1: البنية واحد لواحد لشبكة عصبية متكررة

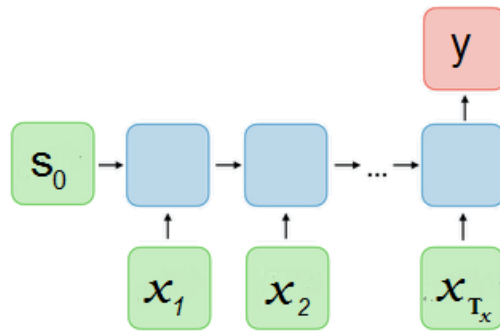
- واحد لكثير **one to many**: تستند هذه البنية على إدخال واحد لعدد متغير من المخرجات (الشكل 18.1)، على سبيل المثال في إنشاء محتوى موسيقي يكون الإدخال إما النوتة الأولى من المقطوعة أو نوعها مثلا، و يكون الإخراج عبارة عن سلسلة نوتات لبقية المقطوعة.



شكل 18.1: البنية واحد لكثير لشبكة عصبية متكررة

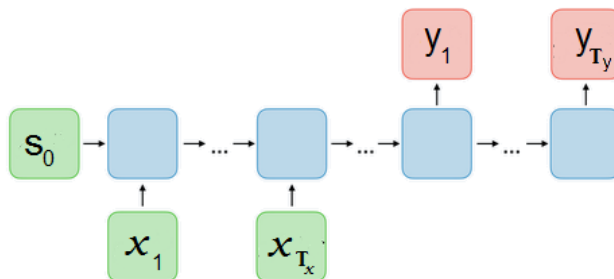
- كثير لواحد **many to one**: تبني هذه الشبكة على عدد متغير من المدخلات منتجة مخرجا واحدا (الشكل 19.1)، كما هو الحال في تحليل المشاعر حيث يكون

الإدخال عبارة عن سلسلة نصية برأي ما و يكون المخرج قطبية هذا الرأي.



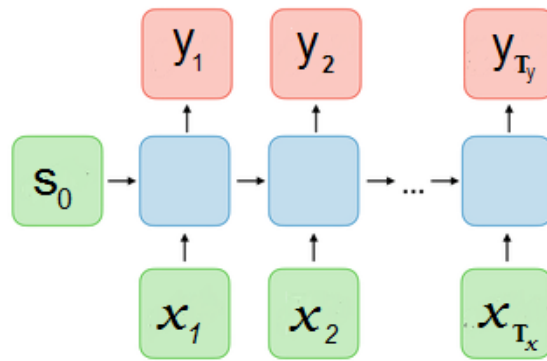
شكل 19.1: البنية كثير لواحد لشبكة عصبية متكررة

- كثير لكثير غير المباشر many to many indirect: تستخدم هذه البنية عددا متغيرا من المدخلات لعدد متغير من المخرجات، يبدأ إنتاج هذه المخرجات بعد إدخال جميع المدخلات (الشكل 20.1). أشهر التطبيقات التي استخدم فيها هذا النوع هي الترجمة الآلية، حيث تكون المدخلات عبارة عن جملة بلغة ما، و تكون المخرجات ترجمة هذه الجملة بلغة أخرى.



شكل 20.1: البنية كثير لكثير غير المباشر لشبكة عصبية متكررة

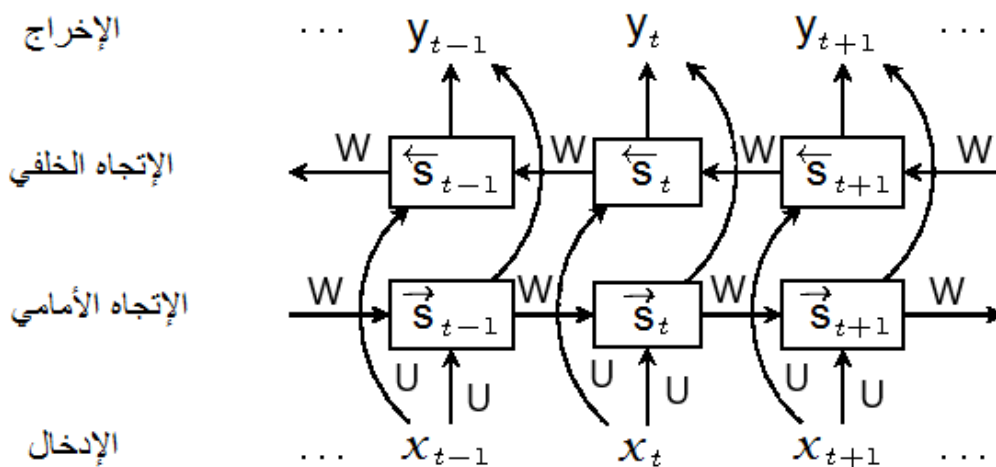
- كثير لكثير المباشر many to many direct: يكون لهذه الشبكة عدد متغير من المدخلات لنفس العدد من المخرجات، حيث تشكل أزواج من المدخلات و المخرجات (الشكل 21.1)، و كمثال على استخدامات هذه البنية تصنيف صور الفيديو.



شكل 21.1: البنية كثير لكثير المباشر لشبكة عصبية متكررة

### الشبكات العصبية المتكررة ثنائية الإتجاه Bidirectional Recurrent Neural Network

نعلم مما تطرقنا له سابقا أن الشبكات العصبية المتكررة تفيدنا بشكل ملموس في معالجة المشاكل ذات البيانات المتسلسلة، إلا أنه قد تواجهنا مشاكل يتطلب حلها معرفة بالمستقبل أو بعضا منه، هذا ما أدى إلى ظهور الشبكات العصبية ثنائية الإتجاه (الشكل 22.1) التي تقوم بوضع شبكتين عصبيتين متكررتين معا بالتوازي، حيث يتم تغذية إحداهما في الإتجاه الأمامي و الثانية في الإتجاه الخلفي دون وجود إرتباط بين مخرجات الحالات الأمامية و مدخلات الحالات الخلفية و العكس [SP97]. تأخذ الشبكات العصبية المتكررة التسلسل الكامل بعين الإعتبار مما يجعل معلومات إدخال الماضي و المستقبل متاحة عند التنبؤ بالمخرج المستهدف [SAWN14].



شكل 22.1: الشبكة العصبية المتكررة ثنائية الإتجاه

**الاتجاه الأمامي:** يتم حساب الحالات في الاتجاه الأمامي بنفس طريقة الشبكات العصبية المتكررة العادية، بإستعمال الحالة في الخطوة  $t-1$  لحساب الحالة في الخطوة  $t$ .

$$\vec{s}_t = U_{\vec{s}} * x_t + W_{\vec{s}} * \vec{s}_{t-1} + b_{\vec{s}} \quad (14.1)$$

**الاتجاه الخلفي:** في الاتجاه الخلفي نبدأ بحساب الخطوة  $t = T$  أي أننا نأخذ الحالة في الخطوة  $t+1$  لحساب الحالة في الخطوة  $t$ .

$$\overleftarrow{s}_t = U_{\overleftarrow{s}} * x_t + W_{\overleftarrow{s}} * \overleftarrow{s}_{t+1} + b_{\overleftarrow{s}} \quad (15.1)$$

و يحسب المخرج النهائي ب:

$$y_t = V * \vec{s}_t + V * \overleftarrow{s}_t + b_y \quad (16.1)$$

### الشبكات العصبية المتكررة العميقة Deep Recurrent Neural Network

يفهم العمق في الشبكات العصبية العادية بزيادة عدد الطبقات المخفية، أما بالنسبة للشبكات العصبية المتكررة فهو تكديس عدة طبقات متكررة معا [PGCB13]، حيث ترتبط الحالة  $s$  في الخطوة  $t$  بالحالة السابقة في نفس طبقتها و الحالة لنفس الخطوة في الطبقة التي تسبقها، و بالتالي يكون لديها ذاكرة للمدخلات [HS13]، تستعمل هذه البنية في السلاسل الزمنية مثلا [PP14].

و لكل طبقة الوزن و الإنحياز الخاص بها (الشكل 23.1).  
تحسب قيمة الحالة  $s_{t_l}$  بالمعادلة (17.1) و قيمة المخرج  $y_t$  بالمعادلة (18.1).

$$s_{t_l} = \begin{cases} f(W_l * s_{t-1_l} + U_l * s_{t_l-1} + b_l) & \text{إذا } l > 0 \\ f(W_l * s_{t-1_l} + U_l * x_t + b_l) & \text{إذا } l = 0 \end{cases} \quad (17.1)$$

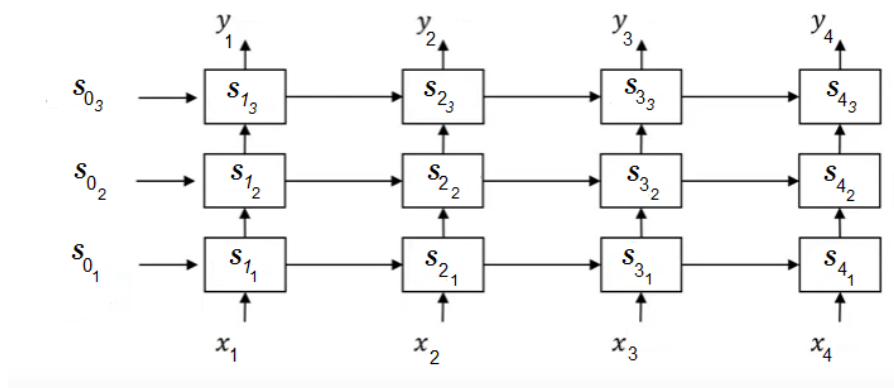
$t$ : يمثل الخطوة.

$l$ : يمثل رقم الطبقة.

$U$ : هي الأوزان من الطبقة السفلية أو الإدخال.

$W$ : هي الأوزان المخرجة من الطبقة 1.

$$y_t = f\left(\sum_{l=1}^L (V_l * s_{t_l} + b_l)\right) \quad (18.1)$$



شكل 23.1: شبكة عصبية متكررة بعمق 3

### إختفاء/انفجار التدرج Vanishing/Exploding gradient

تعتبر الشبكات العصبية المتكررة أنظمة ديناميكية قوية و لكن تبين أن تدريبها يمثل مشكلا، و هذا يعود لاختفاء أو انفجار التدرج من خطوة زمنية لأخرى يحدث هذا عندما تكون الخطوة الزمنية طويلة جدا و عندما تحتوي الشبكة على عدد كبير من الطبقات العميقة [BSF94].

**إنفجار التدرج:** يعبر انفجار التدرج عن الزيادة في التدرج و هذا ينشأ عند انفجار المكونات طويلة المدى، و يمكن حل هذه المشكلة عن طريق قطع التدرجات [PMB12].

**اختفاء التدرج:** يحدث نتيجة الإنخفاض في التدرج و هذا ما يجعل من الصعب تدريب الأوزان، فبدلك لا يستطيع النموذج التعلم من المكونات طويلة المدى، و الحل الأكثر شيوعا لهذا المشكل هو استخدام الذاكرة طويلة قصيرة المدى.

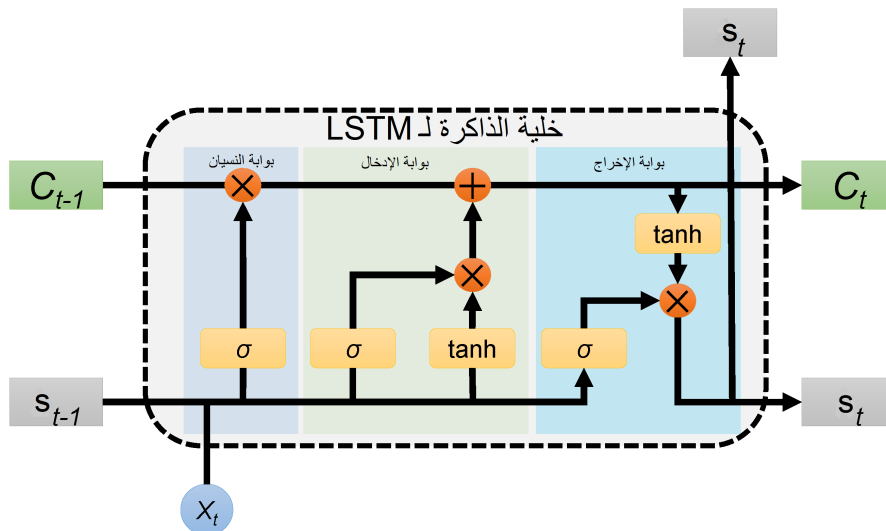
### الذاكرة طويلة-قصيرة المدى Long-Short Term Memory

يمكن للشبكات العصبية المتكررة الإستفادة من معلومات التسلسل، ولكن لبضع خطوات فقط لذا قد تضيع المعلومة أحيانا أو يختفي التدرج [BSF94]، و هذا كان السبب في ظهور نوع جديد من الشبكات العصبية المتكررة يدعى الذاكرة طويلة-قصيرة المدى (LSTM)



[HS97]. يحتوي هذا النوع على خلية (cell) إضافية في الطبقة المخفية تستطيع تذكر المدخلات لفترة طويلة.

تحتوي هذه الخلية على بوابات (الشكل 24.1) بوابة الإدخال (input gate)، بوابة النسيان (forget gate)، بوابة الإخراج (output gate).



شكل 24.1: خلية الشبكة المتكررة LSTM

تقوم بوابة الإدخال بتحديد المعلومات التي ستخزن في حالة الخلية  $C_t$ ، حيث تنقسم هذه العملية إلى جزأين، الأول يتم فيه تطبيق دالة أسية (sigmoid) لتقرير المعلومات التي سنحدثها (المعادلة (19.1))، والثاني تطبق فيه دالة tanh من أجل التخلص من التباين في المعلومات (المعادلة (20.1)).

$$i1_t = \sigma(W_i * s_{t-1} + U_i * x_t + b_{i1}) \quad (19.1)$$

$$i2_t = \tanh(W_i * s_{t-1} + U_i * x_t + b_{i2}) \quad (20.1)$$

و تقوم بوابة النسيان بالتخلص من المعلومات غير المرغوبة من الإدخال، و يتم هذا بتطبيق الدالة الأسية عليها (المعادلة (21.1))، حيث ينظر للقيمة 1 على أن هذه المعلومة مرغوبة و

يجب الإحتفاظ بها، و القيمة 0 على أنه يجب التخلص من هذه المعلومة.

$$f_t = \sigma(W_f * s_{t-1} + U_f * x_t + b_f) \quad (21.1)$$

نقوم بتحديث حالة الخلية  $c_t$  (المعادلة (22.1)) قبل المرور إلى بوابة الإخراج، حيث تعتمد في حسابها على نتائج بوابتي الإدخال و النسيان و حالة الخلية السابقة.

$$c_t = f_t * c_t + i1_t * i2_t \quad (22.1)$$

تقوم بوابة الإخراج بتحديد ما سيتم إخرجه من الخلية و ذلك بتطبيق دالة tanh على حالة الخلية  $c_t$  لحصر القيم بين -1 و 1 ثم ضربها (المعادلة (24.1)) في إخراج الدالة الأسية (المعادلة (23.1)).

$$y_t = \sigma(W_y * s_{t-1} + U_y * x_t + b_y) \quad (23.1)$$

$$s_t = y_t * \tanh(c_t) \quad (24.1)$$

## 4.1 خاتمة

في هذا الفصل تم التطرق إلى المفاهيم الأساسية للشبكات العصبية الإصطناعية، أنواعها و كيف يتم تعليمها، إضافة إلى المشاكل التي قد تواجهها. كما تم الحديث عن بعض الطرق لحل هذه المشاكل في الشبكات العصبية العميقة، كما تطرقنا إلى الشبكات العصبية التلافيفية و الشبكات العصبية المتكررة كأنواع للشبكات العصبية العميقة.

# الفصل الثاني

## تحليل المشاعر Sentiment Analysis

### 1.2 مقدمة

تحليل المشاعر أو التنقيب عن الرأي هو مجال بحث نشط في معالجة اللغة الطبيعية (NLP)، يهدف إلى دراسة مواقف وآراء الناس حول كيانات أو مواضيع معبر عنها في نص معين. يركز تحليل المشاعر على اكتشاف قطبية الآراء المعبر عنها بشكل مباشر أو ضمني ما إذا كانت هذه الآراء إيجابية، سلبية أو محايدة [Liu15].

في هذا الفصل سنتطرق إلى بعض المفاهيم الأولية لتحليل المشاعر ومعالجة اللغة الطبيعية وبعض الدراسات التي أجريت حول تحليل المشاعر، أخيرا نتطرق إلى بعض التحديات التي تواجه البحث في هذا المجال.

### 2.2 مفاهيم أولية

#### 1.2.2 ماهي المشاعر و ماهو الرأي؟

في قاموس ميريام ويبستر (Merriam Webster) [Mer] تُعرّف المشاعر على أنها موقف أو فكر أو حكم ناتج عن شعور، في حين يتم تعريف الرأي على أنه وجهة نظر أو حكم أو تقييم يتكون في العقل حول مسألة معينة.

كان هناك بعض اللبس حول الفرق بين الكلمتين وما إذا كان يجب تسمية المجال بتحليل المشاعر أو التنقيب عن الرأي، لكن لم يتم التطرق لهذا الأمر كثيرا لأن المجال نشأ من

علوم الكمبيوتر و ليس من اللغويات. في الواقع يمكن القول أن الفرق بين المشاعر و الرأي طفيف، فكل منهما يحتوي بعضا من عناصر الآخر لذا يمكن أن يكون الرأي نتيجة لشعور ما [Liu15].

### 2.2.2 لماذا تحليل المشاعر؟

تعتبر الآراء مهمة جدا لإتخاذ القرارات عند البشر، ففي القديم إذا أراد الشخص اتخاذ قرار ما يقوم بسؤال العائلة أو الأصدقاء أما الآن و مع ظهور وسائل التواصل الإجتماعي و المدونات أصبحت هذه الآراء و المعلومات متاحة للجميع، و لكن مع هذا الكم الهائل من الآراء في الإنترنت قد يجد هذا الشخص صعوبة في تحليلها، هنا نجد الحاجة لأنظمة تحليل المشاعر و التي اعتمدها العديد من الشركات من أجل تطوير منتجاتها بمراقبة آراء عملائها، إضافة إلى استعمال هذه الأنظمة أيضا في مجالات أوسع مثل القضايا السياسية، مدى ملاءمة مقاطع الفيديو للأطفال بناء على التعليقات، مراقبة العلامات التجارية...

كانت هناك العديد من الأبحاث حول تأثير تحليل المشاعر على هذه المجالات حيث وجد Tumasjan et al. [TSSW10] أن التعليقات السياسية في تويتر تظهر توجهات الناخبين و تتوافق مع نتائج الانتخابات، كما أظهرت نتائج دراسات Gilad و Glace [06+MG] أن منشورات الآراء حول الأفلام تلعب دورا في نجاحها و زيادة إيراداتها المالية.

### 3.2.2 معايير تحليل المشاعر

لدراسة رأي ما يجب أن يتوفر على الأقل على بعض المعايير الآتية:

- الكيان Object: و هو الفرد أو المنظمة أو المنتج المعني بالرأي مثل: حاسوب.
- الجوانب Attributes: و هي خصائص أو مكونات الكيان مثل: الذاكرة العشوائية (RAM)، البطارية، مساحة التخزين، حجم الشاشة...
- صاحب الرأي Opinion holder: و هو الشخص أو المنظمة التي تعبر عن الرأي مثل: الشخص الذي اشترى الحاسوب.
- قطبية الرأي (Opinion orientation (polarity): أو الإتجاه و هو الموقف العام للرأي، و عادة ما يكون إيجابيا أو سلبيا أو محايدا.
- شدة الرأي Opinion strength: هو مستوى أو درجة المشاعر المعبر عنها مثل: أناأكرهه

جدا جدا جدا، فهي تعبر عن مشاعر سلبية قوية.

#### 4.2.2 أنواع الآراء

**الرأي المنتظم Regular Opinion:** يُعبّر عنه على أنه رأي عادي و ينقسم إلى نوعين:

- الرأي المباشر: يعبر مباشرة عن كيان ما أو جوانبه مثل القول في جملة "طعم القهوة رائع".

- الرأي غير المباشر: على عكس الرأي المباشر، يعبر الرأي غير المباشر عن كيان أو جوانبه بشكل غير مباشر بناء على الآثار الإيجابية أو السلبية لبعض الكيانات الأخرى مثل "بعد شرب الدواء شعرت بمغص في معدتي"، تصف هذه الجملة شعورا لتأثير الدواء على المعدة يستنتج منه رأيا سلبيا حول هذا الدواء.

**الرأي المقارن Comparative Opinion:** هو عبارة عن علاقة تشابه أو اختلاف بين كيانين أو أكثر أو علاقة تفضيل بين كيانين بناء على جوانب مشتركة بينهما على سبيل المثال: "متصفح موزيلا أفضل من متصفح كروم".

**الرأي الذاتي Subjective Opinion:** هو رأي منتظم أو مقارن يظهر في تصريحات شخصية و كمثال على ذلك: "أعتقد أن هذه الشركة لا تهتم بما يفكر به عملاؤها عنها".

**الرأي الضمني Fact-Implied Opinion:** هو رأي منتظم أو مقارن في تصريح موضوعي أو واقعي يفهم من سياق الكلام، يعبر عن حقيقة أو فعل مرغوب أو غير مرغوب فيه. و ينقسم إلى نوعين:

- رأي ضمني شخصي Personnal Fact-Implied Opinion: هو رأي ضمني في تصريح حول تجربة شخصية لشخص ما. على سبيل المثال "إشترت الهاتف بالأمس و تعطل اليوم".

- رأي ضمني غير شخصي Non Personal Fact-Implied Opinion: يعبر هذا الرأي عن حدث واقعي غير معبر عنه من منظور شخص ما على سبيل المثال "ارتفع عدد الإصابات بفيروس كورونا عن الأسبوع الماضي"

**رأي الشخص الأول First Person Opinion:** يوضح موقف شخص أو ممثل مجموعة أو منظمة تجاه كيان ما مثل "نحن قلقون بشأن قرار رفع الحجر الصحي".

**رأي غير الشخص الأول Non First Person Opinion:** هو رأي شخص حول كيان ما يصرح به من طرف شخص آخر مثل "أعتقد أن أختي لا تحب هذا الفيلم".

**الآراء الفوقية Meta Opinion:** هو رأي حول رأي آخر و قد يكون هذا الأخير رأي ضمنى أو ذاتي، و كمثال على ذلك "انا مستاءة لأن أختي لم تعجب بالفلم الذي اقترحتة عليها".

للإلمام أكثر بأنواع الآراء يمكن الإطلاع على [Liu15] ، [JL06] و [ZL11].

## 5.2.2 مستويات تحليل المشاعر

يوجد ثلاث مستويات لتحليل المشاعر [WWH05]:

### مستوى الوثيقة Document Level

يتم تصنيف الوثيقة حسب الشعور العام إذا ما كان إيجابى أو سلبى أو حيادى، و يفترض أن تحتوي الوثيقة على رأي حول موضوع أو كيان واحد [PLV02 ,Tur02].

### مستوى الجملة Sentence Level

في مستوى الجملة يكون التصنيف بتحديد ما إذا كان الرأي في هذه الجملة إيجابيا او سلبيا أو محايدا.

### مستوى الجانِب Aspect Level

يطلق عليه أيضا تحليل المشاعر القائم على الجانِب، يدرس هذا المستوى الجمل متعددة الآراء فهو يسعى إلى إكتشاف المشاعر على الكيانات و/أو جوانبها بالنظر مباشرة إلى الرأي و هدفه (الكيان أو أحد جوانبه)، تعطي الجملة "هذا العطر غال جدا لكنني أحبه" إنطباع إيجابى لكن لا يمكن القول عنها أنها إيجابية تماما إذ أنها تحتوي على مشاعر سلبية حول

سعر العطر ويعتبر هذا المستوى الأكثر إستخداما في أنظمة تحليل المشاعر الواقعية في الصناعة [Liu15].

## 3.2 معالجة اللغة الطبيعية Natural Language Processing

تطورت اللغات الطبيعية عبر الأجيال فأصبح من الصعب ضبطها بأحكام وأساليب واضحة. بالنسبة للبشر من السهل فهم اللغات المتداولة بينهم ولكن يختلف الأمر بالنسبة للآلة لذلك مع ظهور معالجة اللغة الطبيعية بإعتبارها تقاطعا للذكاء الإصطناعي مع اللغويات أصبحت للآلة القدرة على فهم اللغة البشرية أو على الأقل فهمها بنسبة تسمح بإعطاء ردود لها.

عرفت إيزابيث ليدي معالجة اللغة الطبيعية ب : "مجموعة من التقنيات الحسائية بدافع نظري لتحليل و تمثيل النصوص التي تحدث بشكل طبيعي على مستوى واحد أو أكثر من التحليل اللغوي لغرض تحقيق معالجة لغة تشبه الإنسان لمجموعة من المهام أو التطبيقات [Lid01]."

معالجة اللغة الطبيعية مهمة للغاية في العديد من المجالات و تعتبر من أوسع مجالات الذكاء الإصطناعي، فقد ظهرت لها العديد من التطبيقات كالتعرف على الكلام وتلخيص النص، الترجمة الآلية وما إلى ذلك.

يعتبر تحليل المشاعر من أكثر المواضيع التي أثارت الأبحاث في معالجة اللغة الطبيعية لأنها أضافت العديد من التحديات الصعبة التي لم يتم التطرق لها من قبل لدرجة أن بعض الباحثين أصبحوا يعتبرونها أكثر من مجرد فرع من فروع معالجة اللغة الطبيعية بل ينظرون إليها على أنها نسخة مصغرة منها [Liu15].

## 4.2 طرق تمثيل البيانات

يمكن القول أن المشكلة في العمل على النصوص هو صعوبة الحصول على المعلومات من النص الذي هو عبارة عن سلاسل طويلة وإدخالها في واجهة برمجة تطبيقات API. لذلك

هناك طرق لجعل هذه النصوص مدخلات يمكن استعمالها.

### (أ) حقيبة الكلمات Bag Of Words

وهي طريقة شائعة لازالت تستخدم حالياً. تعمل طريقة التمثيل هذه بأخذ كل كلمة و حساب عدد المرات التي تظهر فيها في كل وثيقة أو نص، أي أننا نحول سلسلة الكلمات إلى شعاع (vector) حيث يمثل عدد الكلمات طول الشعاع. مساوئ هذه الطريقة هي غياب ترتيب الكلمات و هو أمر مهم في أي لغة.

### (ب) One Hot Encoding

يستخدم هذا التمثيل الأحرف بشكل فردي أي يأخذ ترميز واحد لكل حرف في النص، لكن في أي لغة مفهوم الكلمة مهم و مفيد جداً لذلك تمرير المدخلات كحروف يجعل النموذج يتعلم عن اللغة أكثر، ما يجعله خام أكثر ما لم يكن لدينا كميات هائلة من البيانات، هذا ما أدى إلى ظهور تحويل آخر بين هذا التحويل و حقيبة الكلمات و هو ترميز الكلمة (Word Encoding/Word Embedding).

### (ج) ترميز الكلمة

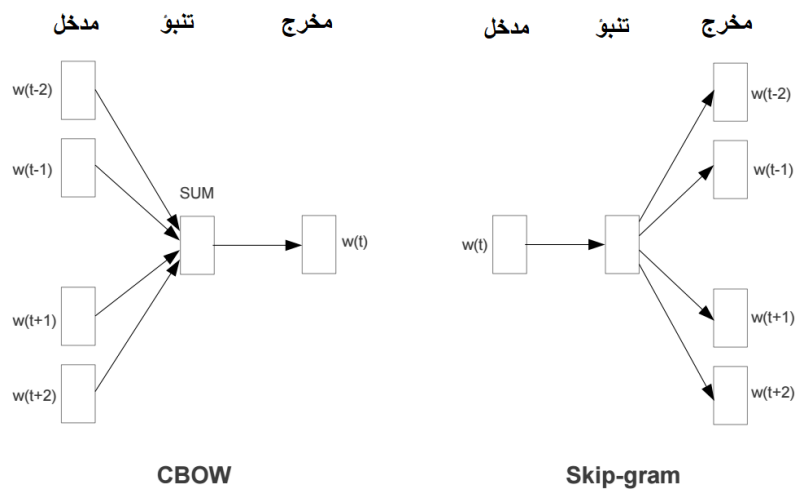
يقوم بتحويل كل كلمة من الوثيقة أو النص إلى شعاع من الأرقام و تكون جميع الأشعة بنفس الطول، على سبيل المثال في الجملة "أنا أحب السفر"، اذا افترضنا أن طول الأشعة هو 4، سيكون لكلمة "أنا" شعاع بنفس الأرقام دوماً و نفس الشيء لباقي الكلمات أي أن تمثيل كل كلمة يبقى ثابت. بدل قيامنا بحساب الترميزات يمكننا استخدام ترميزات محسوبة مسبقاً، و من أشهر هذه الترميزات Word2Vec.

### (د) Word2Vec

هي خوارزمية لتمثيل الكلمات بأشعة تأخذ بعين الاعتبار سياق الكلمة بحيث يتم تمثيل الكلمات المتشابهة في السياق بأشعة متقاربة. تحتوي خوارزمية Word2Vec على نموذجين CBOW و Skip-gram (الشكل 1.2)



حيث يقوم الأول بالتنبؤ بمعنى الكلمة بناء على سياقها لكي يتمكن من تمثيلها. أما الثاني فيقوم بالتنبؤ بالكلمات المجاورة للكلمة [MCCD13].



شكل 1.2: نموذجي Skip-gram و CBOW

## 5.2 مقاييس تقييم النموذج

هناك العديد من المقاييس لتقييم جودة نماذج التعلم العميق اعتمدنا في تقييم نموذجنا على المقاييس التالية:  
ليكن لدينا الجدول التالي:

جدول 1.2: جدول التوافق

الفئة السلبية	الفئة الإيجابية	
tn	tp	التوقع الصحيح
fp	fn	التوقع الخاطئ

### Precision

هي نسبة توقع النموذج للفئة بشكل صحيح أي عدد الأمثلة المصنفة لفئة ما وهي حقا تنتمي لتلك الفئة. و تعطى بالعلاقة الآتية:

$$\text{Precision} = \frac{tp}{tp + fp} \text{ أو } \frac{tn}{tn + fn} \quad (1.2)$$

Recall

هو نسبة الفئة المتوقعة بشكل صحيح من جميع الأمثلة المتوقعة لتلك الفئة و تعطى  
بالعبارة:

$$\text{Recall} = \frac{tp}{tp + fn} \text{ أو } \frac{tn}{tn + fp} \quad (2.2)$$

الدقة accuracy

هي نسبة جميع الفئات المتوقعة بشكل صحيح بالنسبة للعدد الكلي للبيانات وتعطى بالعبارة:

$$\text{accuracy} = \frac{tp + tn}{tp + fp + tn + fn} \quad (3.2)$$

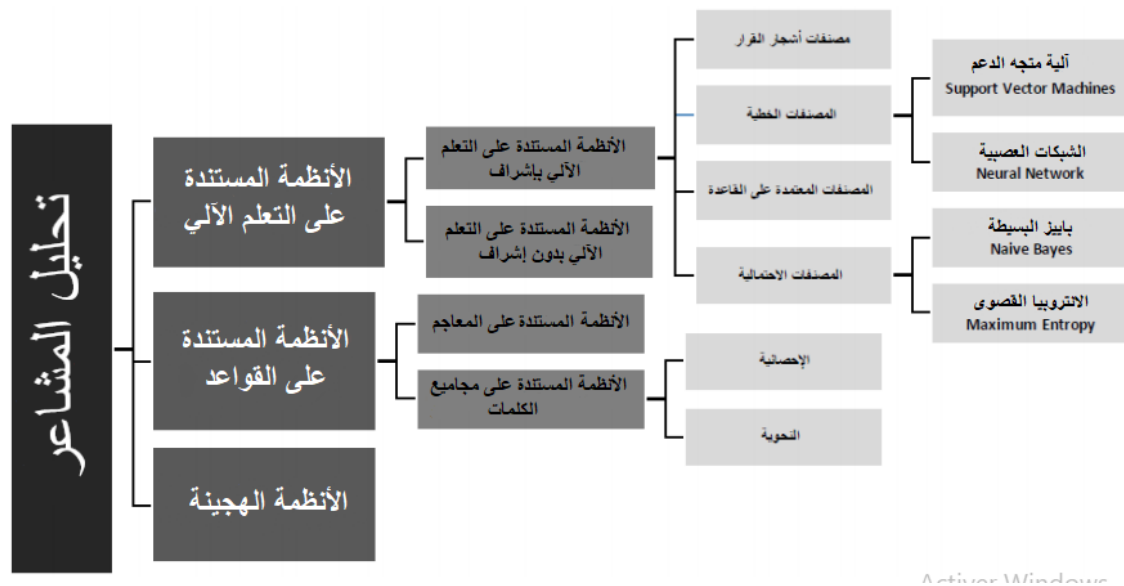
F1-score

هو مقياس لتقييم المصنفات يوحد بين Precision و Recall باستخدام المتوسط التوافقي  
ويعطى بالعبارة:

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4.2)$$

## 6.2 تحليل المشاعر

هناك الكثير من الأساليب لإنجاز نموذج تحليل المشاعر، و من بينها و التي تعتبر الأهم،  
الأنظمة المستندة على القواعد (Rule Based Systems) و الأنظمة المستندة على التعلم  
الآلي (Machine Learning Based Systems) و الأنظمة الهجينة (Hybrid Systems) [AFGG15] (الشكل 2.2).



شكل 2.2: أهم الأساليب المتبعة في عملية تحليل المشاعر

## 1.6.2 الأنظمة المستندة على القواعد

هي الأنظمة التي تستخدم مجموعة من القواعد المصممة يدويا للمساعدة في تحليل المشاعر، و تحدد هذه القواعد أشياء مثل القطبية و الذاتية، و قد تشمل تقنيات مختلفة من اللغويات الحاسوبية مثل Stemming, Tokenization, Part-Of-Speech Tagging and Parsing.

يقوم هذا النظام بتحديد قائمتين للمصطلحات الإيجابية و السلبية بناء على معجم (الأنظمة المستندة على المعاجم) أو بناء على المجاميع (الأنظمة المستندة على مجاميع الكلمات (Corpus)) (الشكل 2.2)، ثم يقوم بحساب عدد المصطلحات الإيجابية و السلبية، إذا كان عدد الإيجابية منها أكبر فإنه يصنفها على أنها إيجابية، أما إذا كان عدد المصطلحات السلبية أكبر فإنه يرجع الفئة السلبية، أما إذا كانا متساويان فإن النظام يرجع الفئة المحايدة.

كمثال على ذلك تم اقتراح محلل للمشاعر على مستوى الوثيقة لمقالات إخبارية مالية في "Rule-based Sentiment Analysis for Financial News" [ITSPCA15]. يعتمد مقترح هؤلاء الباحثين على عمل كلينر [KPF09] الذي استخدم تركيب المشاعر (و هو تحديد مشاعر وحدة لغوية متعددة الكلمات مثل عبارة أو جملة بناء على مكوناتها (فعل، إسم، صفة، حروف جر، أدوات ربط،...)) لتصنيف المشاعر على مستوى الجملة، حيث

قاموا بإجراء تعديلات على خوارزمية عمل كلينر لتلائم مستوى الوثيقة.

إستخدم مقترحو هذا العمل محلل Stanford RNN (محلل يقوم بتحليل العبارة بناء على القواعد النحوية [SBMN13]) من أجل استخراج العبارة، و المعاجم الذاتية (و هو معجم قطبية يأخذ بعين الإعتبار القطبية السابقة و القطبية السياقية للكلمة أو العبارة [WWH05]) لتحديد القطبية.

و تم تمثيل كلمات النص بمصدرها (lemma) بدلا من كلماته الأصلية مع إضافة بعض القواعد الأخرى لتركيب المشاعر.

يتكون محلل المشاعر هذا من خمسة مراحل:

#### (أ) إستخراج العبارة Phrase Extraction

في هذه المرحلة يتم تقسيم المقالة الإخبارية إلى جمل ثم يتم تحليل هذه الجمل بإستخدام محلل Stanford RNN وفق علامات تصنيف أقسام الكلام (Part Of Speech Tagging).

#### (ب) كشف القطبية Polarity Detection

في هذه المرحلة يتم تحديد قطبية كلمات النص المتحصل عليه من المرحلة السابقة بإستخدام المعجم الذاتي (ويلسون 2005)، حيث يتم تحديد قطبية كل كلمة من النص بناء على قطبية الزوج المطابق لها في المعجم.

#### (ج) تحديد القطبية للجمل Polarity Identification for Sentences

في مرحلة تحديد القطبية تطبق قواعد تركيب المشاعر، و هي مجموعة من القواعد التي يتم من خلالها تحديد قطبية الجمل.

يعتبر تسلسل القواعد مهم جدا من أجل ضمان التفسير الصحيح لمعاني الجمل.

من أجل معالجة تركيب مشاعر الجملة الإسمية و عبارة الجر قاموا بإنشاء مجموعة من القواعد بناء على ما قام به كلينر [KPF09]، كما أضافوا قواعد جديدة لمعالجة تركيب المشاعر للجملة الفعلية و الجمل الفعلية-الإسمية/الإسمية-الفعلية، النفي و أدوات الربط "لكن"، يوضح الجدول التالي بعض الأمثلة عن هاته القواعد.

جدول 2.2: أمثلة عن بعض القواعد المعمول بها لتحديد قطبية الجمل

المثال	المخرج	قطبية الجمل	التركيب	القاعدة
most difficult business decision	سلبية	سلبية-إيجابية	جملة موصوفة-جملة إسمية	الجملة الإسمية
top preferred	إيجابية	إيجابية-إيجابية	صفة-فعل	الجملة الفعلية
gains from asset sales dwindled	سلبية	إيجابية-سلبية	جملة إسمية-جملة فعلية	جملة فعلية-إسمية/إسمية-فعلية
The first half of the year was a nightmare but we have performed much better in the second half.	إيجابية	سلبية-إيجابية	ربط جملتين ب "but"	أداة الربط "but"

بعد تطبيق جميع القواعد يمكن تحديد قطبية الجمل.

(د) حساب قيمة المشاعر Sentiment Value Calculation

تحسب قيمة المشاعر للمقال بأكمله بإستعمال النسبة P/N التي تستخدم عدد الجمل الإيجابية و عدد الجمل السلبية التي تم الحصول عليها من مهمة تحديد قطبية الجملة.

يتم حساب قيمة المشاعر للمقال عن طريق حساب معدل قيم مشاعر جميع جمل المقال.

## هـ) تصنيف القطبية Polarity Classification

يصنف المقال اعتماداً إلى إجمالي قيمة المشاعر المحسوبة، وقد ركز الباحثون في عملهم على التصنيف إلى فئتين إيجابية و سلبية مع تجاهل الحيادية، و تمت الإشارة ب +1 للمقال إذا كان ذو قيمة عاطفية إيجابية أكبر و -1 إذا كان ذو قيمة عاطفية سلبية أكبر.

## النتائج التجريبية

استخدم هؤلاء الباحثون ما يقارب 200 مقال إخباري مالي من الصحف الماليزية على الإنترنت و التي تم تصنيفها يدوياً.

تم حساب Precision, Recall, F-score (القسم 5.2) من أجل تقييم أداء المحلل المقترح في الحالات الثلاث التالية: تصنيف الحالات الإيجابية، تصنيف الحالات السلبية و تصنيف جميع الحالات.

و بعد ذلك قاموا بمقارنة أداء محللهم مع أداء محلل مشاعر أساسي (Baseline Sentiment Analyser) ( يأخذ هذا المحلل بعين الاعتبار قواعد تركيب المشاعر و يستخدم مصنفات ك Naive Bayes, svm) بعد إجراء بعض التعديلات عليه ليلائم تصنيف الوثيقة. النتائج المتحصل عليها مبينة في الجدول 3.2.

جدول 3.2: مقارنة بين أداء محلل البيانات الأساسي ومحلل البيانات المقترح

محلل المشاعر المقترح			محلل المشاعر الأساسي			القياس ب %
جميع التصنيفات	التصنيفات السلبية	التصنيفات الإيجابية	جميع التصنيفات	التصنيفات السلبية	التصنيفات الإيجابية	
73.0	50.0	88.3	61.5	25.0	85.8	Recall
78.5	85.1	76.3	67.2	66.7	67.3	Precision
75.7	63.0	81.9	64.2	36.4	75.5	F-Score

أظهر محلل المشاعر المقترح أداء أفضل مقارنة بمحلل المشاعر الأساسي في جميع الحالات، حيث حقق نسبة 6.75 % لل F-score في جميع التصنيفات بينما حقق الأساسي 2.64 %.

## 2.6.2 الأنظمة المستندة على التعلم الآلي

تعتمد هذه الأنظمة على تقنيات التعلم الآلي من أجل تصنيف النص إلى سلبي أو إيجابي أو محايد، مثل خوارزمية بايز البسيطة و SVM و خوارزميات التعلم العميق.

سننتقل إلى دراسة خوارزمية بايز البسيطة و نموذج LSTM-CNN في التعلم العميق.

### (أ) بايز البسيطة Naive Bayes

فيمايلي نتطرق لما جاء في كتاب "Speech and Language Processing" [DJ19].

عرفت فكرة الإستدلال البايزي منذ عمل بايز (1763)، و تم تطبيقها لأول مرة على تصنيف النص من قبل موستيلر و والاس سنة 1964 [MW12].

بايز البسيطة هي مصنف احتمالي بسيط قائم على تطبيق نظرية بايز مع افتراضات استقلالية بسيطة بين الميزات. يقوم هذا المصنف بإرجاع الفئة  $c$  التي لها أكبر احتمال سابق من بين جميع الفئات  $c \in C$  بالنسبة لوثيقة  $d$  (المعادلة (5.2)). نستعمل رمز  $\hat{c}$  لتقدير التصنيف الأفضل.

$$\hat{c} = \underset{c \in C}{\operatorname{argmax}} P(c|d) \quad (5.2)$$

### استدلال بايز Bayesian Inference

الهدف من تصنيف بايز هو استخدام قاعدة بايز و التي تسمح بتحليل احتمال شرطي إلى 3 احتمالات أخرى كما هو موضح في المعادلة (6.2).

تستخدم نظرية بايز مفهوم الاحتمال السابق للتعبير عن الاحتمال الأصلي للحدث دون معرفة أيّ معلومات إضافية، و الاحتمال اللاحق للتعبير عن الاحتمال بعد الحصول على بعض المعلومات.

$$P(x|y) = \frac{P(y|x) \times P(x)}{P(y)} \quad (6.2)$$

حيث:

$P(x)$  هو الاحتمال السابق للحدث  $x$ .

$P(y)$  هو الاحتمال السابق للحدث  $y$ .

$P(x|y)$  هو الاحتمال اللاحق للحدث  $x$  بشرط وقوع الحدث  $y$ .

$P(y|x)$  هو الاحتمال اللاحق للحدث  $y$  بشرط وقوع الحدث  $x$ .

نعوض العبارة (6.2) في العبارة (5.2) فنجد:

$$\hat{c} = \underset{c \in C}{\operatorname{argmax}} P(c|d) = \underset{c \in C}{\operatorname{argmax}} \frac{P(d|c) \times P(c)}{P(d)} \quad (7.2)$$

و يمكن تبسيط العبارة (7.2) والتي تمثل الفئة الأكثر احتمالاً لنفس الوثيقة  $d$  بسهولة باعتبار أن القاسم  $P(d)$  ثابت في جميع الفئات و ذلك بإزالته لتصبح لدينا العبارة التالية:

$$\hat{c} = \underset{c \in C}{\operatorname{argmax}} P(c|d) = \underset{c \in C}{\operatorname{argmax}} (P(d|c) \times P(c)) \quad (8.2)$$

نقوم بحساب أكثر الفئات احتمالاً  $\hat{c}$  بالنسبة للوثيقة  $d$  عن طريق اختيار الفئة التي لديها أكبر ناتج من ضرب احتمالين: الاحتمال السابق للفئة  $c$  و احتمال الوثيقة  $d$ .

نمثل الوثيقة  $d$  بمجموعة من الميزات  $F = f_1, f_2, \dots, f_n$

$$\hat{c} = \underset{c \in C}{\operatorname{argmax}} P(f_1, f_2, \dots, f_n|c) \times P(c) \quad (9.2)$$



مع ذلك لا يزال من الصعب حساب المعادلة (9.2) مباشرة لذا تعتمد تصنيفات بايز البسيطة على فرضيتين تبسيطيتين:

الأولى هي حقبة الكلمات و التي تفترض أن موقع الكلمة غير مهم، و نفترض أيضا أن الميزات  $f_1, f_2, \dots, f_n$  هي رمز الكلمة وليس موضعها.

أما الثانية فهي فرضية الاستقلال الشرطي أو فرضية بايز البسيطة و التي بموجبها تكون الاحتمالات  $P(f_i|c)$  مستقلة مع الأخذ بالاعتبار الفئة  $c$  و بالتالي يكون (ببساطة) لدينا الجداء كالاتي:

$$P(f_1, f_2, \dots, f_n|c) = P(f_1|c) \times P(f_2|c) \times \dots \times P(f_n|c) \quad (10.2)$$

المعادلة النهائية للفئة المنتقاة من طرف مصنف بايز البسيطة  $c_{NB}$  تكون كالاتي:

$$c_{NB} = \underset{c \in C}{\operatorname{argmax}} P(c) \prod_{f \in F} P(f|c) \quad (11.2)$$

لتطبيق مصنف بايز البسيط على نص يكفي النظر للكلمات كميزات:  $F = w_1, w_2, \dots, w_n$

$$c_{NB} = \underset{c \in C}{\operatorname{argmax}} P(c) \prod_{w_i \in F} P(w_i|c) \quad (12.2)$$

يتم إجراء حسابات بايز البسيطة على غرار العمليات الحسابية لنمذجة اللغة باستعمال اللوغاريثم لتفادي مشكل حسابات القيم الصغرة و تحسين سرعة الحساب، فتصبح لدينا العبارة:

$$c_{NB} = \underset{c \in C}{\operatorname{argmax}} \log P(c) + \sum_{w_i \in F} \log P(w_i|c) \quad (13.2)$$

من أجل حساب المعادلة يجب حساب قيم  $P(c)$  و  $P(w_i|c)$ .  
 لحساب  $P(c)$  نحتاج لحساب النسبة المئوية للوثائق الموجودة في مجموعة التدريب  
 لكل فئة  $c$  حيث:

$$P(c) = \frac{N_c}{N_{doc}} \quad (14.2)$$

حيث تمثل  $N_c$  عدد الوثائق التي تنتمي للفئة  $c$ ، و  $N_{doc}$  العدد الكلي للوثائق.  
 و أما بالنسبة لحساب  $P(w_i|c)$  فهو عدد المرات التي تظهر فيها الكلمة  $w_i$  بالنسبة  
 للعدد الكلي للكلمات في وثائق الفئة  $c$ ، و تعطى عبارتها كالتالي:

$$P(w_i|c) = \frac{\text{count}(w_i, c)}{\sum_{w \in V} \text{count}(w, c)} \quad (15.2)$$

$V$ : تمثل جميع الكلمات في الوثائق.

لكن بايز البسيطة تقوم بضرب كل الإحتمالات، إحتمال الصفر في مصطلح الإحتمال  
 الشرطي يؤدي إلى إحتمال الصفر للفئة و هذا غير مقبول.

الحل الأبسط هو إضافة الواحد (Laplace Smoothing) (العبارة (16.2))، على  
 الرغم من أن هذا الأخير يتم استبداله ب Smoothing Algorithms أكثر تطوراً في  
 نمذجة اللغة إلا أنه يستخدم بشكل شائع في تصنيف نص بايز البسيطة.

$$P(w_i|c) = \frac{\text{count}(w_i, c) + 1}{\sum_{w \in V} (\text{count}(w, c) + 1)} = \frac{\text{count}(w_i, c) + 1}{(\sum_{w \in V} \text{count}(w, c) + |V|)} \quad (16.2)$$

بالنسبة للكلمات التي تكون موجودة في وثائق الإختبار و غير موجودة في وثائق  
 التدريب فإن النموذج يقوم بتجاهلها ولا يعطي لها أي احتمالية.

### خوارزميات التدريب و الإختبار لبايز البسيطة

#### خوارزمية 1 تدريب بايز البسيطة

TrainNaiveBayes( $D, C$ )	خوارزمية
$D$ : وثائق التدريب، $C$ : الفئات	المدخلات
$logprior$ : $logP(c)$ // لوغارثم الاحتمال السابق للفئة $c$	المخرجات
$loglikelihood$ : $logP(w c)$ // لوغارثم الاحتمال اللاحق للكلمة $w$ بالنسبة للفئة $c$	
$V$ : هو شعاع (vector) لجميع كلمات حقيبة الكلمات في جميع الوثائق	
$V \leftarrow D$ الكلمات في جميع الوثائق	1
<b>for each</b> class $c \in C$ // حساب الاحتمال $P(c)$	2
$logprior \leftarrow log \frac{N_c}{N_{doc}}$ // $N_c = c$ عدد الوثائق التي تنتمي للفئة $c$ ، العدد الكلي للوثائق	3
$bigdoc[c] \leftarrow \mathbf{append}(d)$ <b>for</b> $d \in D$ <b>with</b> class $c$ // الوثائق التي تنتمي للفئة $c$	4
<b>for each</b> word $w$ in $V$ // حساب الاحتمال $P(w/c)$	5
$count(w, c) \leftarrow$ (عدد تكرار الكلمة $w$ في $bigdoc[c]$ )	6
$loglikelihood[w, c] \leftarrow \frac{count(w, c) + 1}{\sum_{w \in V} (count(w, c) + 1)}$	7
<b>return</b> $logprior, loglikelihood, V$	8

#### خوارزمية 2 اختبار بايز البسيطة

TestNaiveBayes( $testdoc, logprior, loglikelihood, C, V$ )	خوارزمية
$testdoc$ : وثائق الاختبار، $C, V, logprior, loglikelihood$	المدخلات
الفئة التي لها أكبر احتمال $c$	المخرجات
<b>for each</b> class $c \in C$	1
$sum[c] \leftarrow logprior[c]$	2
<b>for each</b> $w$ in $testdoc$	3
<b>if</b> $w \in V$	4
$sum[c] \leftarrow sum[c] + loglikelihood[w, c]$	5
<b>return</b> $argmax_c sum[c]$	6

### مثال تطبيقي

نتطرق فيما يلي إلى مثال مصغر لتطبيق خوارزمية بايز البسيطة (الجدول 4.2)، حيث ندرس تحليل المشاعر مع الفئتين الإيجابية (+) و السلبية (-) حول مراجعات فلم حقيقية.

جدول 4.2: مثال تطبيقي حول عمل بايز البسيطة

التغريدات	الفئة	المجموعة
just plain boring	-	التدريب
entirely predictable and lacks energy	-	
no surprises and very few laughs	-	
very powerful	+	
the most fun film of the summer	+	الاختبار
predictable with no fun	?	

يتم حساب  $P(c)$  للفئتين بالمعادلة (14.2).

$$P(+) = \frac{2}{5} \quad P(-) = \frac{3}{5}$$

نلاحظ أن الكلمة "with" غير موجودة في مجموعة التدريب و كما سبق و تم الذكر فإن بايز البسيطة تقوم بتجاهلها. تحسب الإحتمالات من مجموعة التدريب للكلمات الثلاث المتبقية بالمعادلة (16.2) و تعطى النتائج كمايلي:

$$P(\text{"predictable"}|+) = \frac{0+1}{9+20} \quad P(\text{"predictable"}|-) = \frac{1+1}{14+20}$$

$$P(\text{"no"}|+) = \frac{0+1}{9+20} \quad P(\text{"no"}|-) = \frac{1+1}{14+20}$$

$$P(\text{"fun"}|+) = \frac{1+1}{9+20} \quad P(\text{"fun"}|-) = \frac{0+1}{14+20}$$

يتم حساب الفئة المختارة لجملة الإختبار "S = predictable with no fun" باستعمال المعادلة (12.2) كمايلي:

$$P(-)P(S|-) = \frac{3}{5} \times \frac{2 \times 2 \times 1}{34^3} = 6,1 \times 10^{-5}$$

$$P(+ )P(S|+) = \frac{2}{5} \times \frac{1 \times 1 \times 2}{29^3} = 3,2 \times 10^{-5}$$

و يتم اختيار القيمة الأكبر عن طريق  $argmax$  و هي التي ستمثل تصنيف العبارة المدخلة، و هي الفئة السلبية في هذه الحالة.

إستخدم Melville et al. [MGL09] خوارزمية بايز البسيطة في تحليل مشاعر لمدونات على ثلاثة مجموعات للبيانات الأولى هي مدونات Lotus و الثانية لمدونات المرشحين السياسيين و الأخيرة لمراجعات أفلام، حيث حققوا دقة قدرت ب % 88.40، % 59.24 و % 80.81 على التوالي.

كما تم استخدامها أيضا في [Haj16] من أجل تحليل الآراء في تويتر و تم استخدام مجموعة بيانات مسترجعة من موقع التواصل الإجتماعي فيسبوك حيث تحصل على دقة بلغت %97 على مجموعة الإختبار.

#### (ب) نموذج LSTM-CNN

بعدها تطرقنا لخوارزمية بايز البسيطة كواحدة من خوارزميات التعلم الآلي، في هذا الشطر سنتطرق لدراسات Katherine et al. [KSCD19] حول تحليل المشاعر للهجات العربية باستخدام الذاكرة طويلة-قصيرة المدى (LSTM) و الشبكات العصبية التلافيفية (CNN) كمثال للتعلم العميق في الأنظمة المستندة على التعلم الآلي.

مع العلم أن اللغة العربية هي إحدى أكثر اللغات تعقيدا، فاختلفت اللغة المكتوبة فيها عن المنطوقة و تعدد لهجاتها عبر مختلف المناطق العربية يجعل من التعامل معها أمرا صعبا، فهذه اللهجات لا تخضع إلى قواعد لغوية رسمية [KSCD19].

#### مجموعة البيانات

فيما يخص مجموعات البيانات قام الباحثون باستخدام ثلاث مجموعات:

LABR: مجموعة بيانات لتحليل المشاعر باللغة العربية لمراجعات كتب، تحتوي على أكثر من 63000 مراجعة، صنفت بمقياس من 1 إلى 5 نجوم حيث تعتبر المراجعات ذات التقييمات 4 و 5 إيجابية، ذات 1 و 2 سلبية و ذات التقييم 3 محايدة [AA13].

ASTD: مجموعة بيانات باللغة العربية باللهجة المصرية تم جمعها من تويتر، تحتوي على حوالي 10000 تغريدة مصنفة على أنها موضوعية، ذاتية إيجابية، ذاتية سلبية و ذاتية مختلطة [NAA15].

Shami-Senti: هي مجموعة بيانات تحليل المشاعر للهجات بلاد الشام (لبنان، سوريا، فلسطين، الأردن)، مأخوذة من الإنترنت مصنفة على أنها إيجابية، سلبية و محايدة [KSCD18].

### معالجة البيانات

تمت معالجة البيانات بإزالة الأحرف الخاصة، علامات الترقيم، علامات التشكيل، الأرقام، الأحرف غير العربية و الأحرف المكررة مع الإحتفاظ بحرفين مكررين فقط.

استبدلت كل جملة بشعاع ترميز الكلمة المقابل لها في نموذج Aravec المدرب مسبقا (مشروع مفتوح المصدر مجاني لتمثيل الكلمة على اللغة العربية تم تنفيذه بواسطة word2vec [SEEB17]). في حالة ما إذا كانت الكلمات موجودة في النص و لا تحتوي على ترميز في النموذج يتم البحث عن الكلمات الأكثر شبها لها لإستخدامها في الحصول على شعاع ترميز الكلمة المقابل لها في نموذج Aravec حيث لا يتجاوز الإختلاف بين الكلمتين حرفين سواء في بداية الكلمة أو نهايتها.

تم تحديد طول كل جملة ب 70 كلمة كحد أقصى، إذا كانت الجملة لا تستوفي الطول المطلوب، تتم زيادة أصفار فيما تبقى من الشعاع (Post Padding) لضمان أن جميع الجمل لها نفس الطول.

يوضح الجدول 5.2 توزيع الفئات في مجموعات البيانات .

جدول 5.2: توزيع الفئات في مجموعات البيانات

الحيادية	الإيجابية	السلبية	مجموعة البيانات
243	1,064	935	Shami-Senti
6,580	6,578	6,580	LABR 3 Balanced
	6,580	6,578	LABR 2 Balanced
	42,832	8,222	LABR 2 Un-Balanced
738	665	1,496	ASTD

قاموا بتجربة نموذجين أساسيين، الأول هو نموذج LSTM بسيط و الثاني نموذج Kaggle الفائز بمسابقة Kaggle لتحليل المشاعر [Kagb].

### نموذج LSTM

تتكون شبكة LSTM التي اقترحها هؤلاء الباحثون من طبقة ترميز (Embedding Layer) مع ترميز كلمة مدرب مسبقا متبوعة بطبقتي LSTM، الأولى ب 128 مخرج و الثانية ب 64، تليهما طبقة تنشيط Relu متصلة بالكامل مع 100 مخرج و 0.5 تسرب، آخر طبقة تستخدم دالة أسية كثيفة (Sigmoid) للتنبؤ بالفئات.

بعد عدة تجارب قام بها الباحثون قرروا استخدام محسن Adam مع دالة الخطأ categorical cross entropy من أجل التصنيف المتعدد و RMSprop للتصنيف الثنائي. يوضح الجدول 6.2 المعطيات المستخدمة في النموذج و قد مثلت تلك التي تؤدي إلى نتائج أفضل بخط عريض.

جدول 6.2: المعطيات العامة لنماذج التعلم العميق

المعطيات	القيم
تقسيم مجموعة البيانات	80% تدريب، 10% تحقق، 10% اختبار
العدد الأقصى للميزات	[7K, 10K, <b>15K</b> , 25K, 40K]
طول الترميز	[ <b>300</b> , 100]
نموذج الترميز	CBOW, Skip-gram
قابلية التدريب للترميز	True, false
أقصى طول للجملة	[100, <b>70</b> , 50]
المرشح	[128, 64, 32]
حجم الأنوية Kernel size	[6, 5, 4, 3, 2]
حجم التجميع	[5, 4, 3, 2, 1]
Batch size	[256, 128, 100, <b>50</b> , 32]
أقصى فترة	1000, <b>100</b> , 50, 10
التسرب	0.7, <b>0.5</b> , 0.2
المحسن	Adam, RMSprop, SGD
دالة التنشيط	Softmax, Sigmoid, Relu
تقسيم LABR (تدريب، تحقق، اختبار)	[20, 10, 70]
تقسيم ASTD و Shami-Senti (تدريب، تحقق، اختبار)	[10, 10, 80]

قاموا بتطبيق هذا النموذج على مجموعتي البيانات ASTD و Shami-Senti محققا دقة (Accuracy) بلغت 53% لكليهما. و في مجموعة البيانات LABR 3 قاموا بتطبيق النموذج في ثلاث تجارب، التجربة الأولى بالنموذج الأساسي، التجربة الثانية بنموذج BiLSTM-LSTM، حيث غيروا الطبقة الثانية بـ BiLSTM، أما الثالثة بنموذج BiLSTM-BiLSTM.

لاحظوا أن النموذج BiLSTM-BiLSTM أعطى نتائج أفضل من حيث الدقة مقارنة بالنماذج الأخرى على مجموعة البيانات LABR 3 فاعتمده في مهمة التصنيف الثنائي (إيجابية، سلبية) في مجموعات البيانات الثلاث (LABR 2، ASTD، Shami-Senti). النتائج موضحة في الجدول التالي:



جدول 7.2: دقة نموذج BiLSTM-LSTM للتصنيف الثنائي في مجموعات البيانات

الدقة	مجموعة البيانات
55.34 %	LABR 2 المتوازنة
81%	LABR 2 غير المتوازنة
68.8 %	ASTD
54.5 %	Shami-Senti

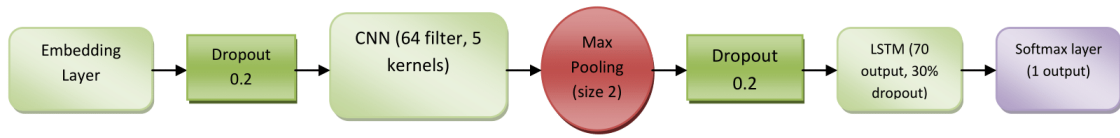
جدول 8.2: مصفوفة الارتباك لنموذج BiLSTM-LSTM للتصنيف الثنائي في مجموعات البيانات

مجموعة بيانات LABR2 غير المتوازنة				مجموعة بيانات ASTD			
المتوقع				المتوقع			
سلبية	إيجابية			سلبية	إيجابية		
505	8036	إيجابية	الحقيقي	45	11	إيجابية	الحقيقي
114	1555	سلبية		136	23	سلبية	

حقق هذا النموذج نتائج بدقة 81 % للمجموعة LABR 2 غير المتوازنة و 68.5 % ل ASTD، و بالنظر لمصفوفة الارتباك (الجدول 8.2) لوحظ أن النموذج لا يتعلم جيدا و يرجع ذلك لتحيزه لفئة الأغلبية كون مجموعتي البيانات غير متوازنتين حيث أن المجموعة ASTD تحتوي على بيانات سلبية أكثر، و المجموعة LABR 2 على بيانات إيجابية أكثر.

نموذج Kaggle الفائز بمسابقة Kaggle لتحليل المشاعر

فاز هذا النموذج في مسابقة Kaggle لتحليل المشاعر الإنجليزية بدقة وصلت إلى 96 %، حيث استخدموا مجموعة بيانات Amazon Fine Food Reviews المكونة من 586.454 مراجعة بتقييم من 1 إلى 5 لكل مراجعة. يستخدم هذا النموذج مزيجا من طبقات LSTM و CNN، حيث يتكون من طبقة تلافيفية مع تجميع أقصى ذو بعد 2 يليها 0.2 تسرب، بعده طبقة LSTM واحدة، و أخيرا طبقة متصلة بالكامل للتنبؤ بفئة المشاعر. النموذج موضح في الشكل 3.2.



شكل 3.2: نموذج Kaggle

قامت Katherine et al. بتدريب هذا النموذج باستخدام البيانات LABR ،ASTD ، Shami-Senti و يوضح الجدول 9.2 النتائج المتحصل عليها:

جدول 9.2: دقة نموذج Kaggle للتصنيف الثنائي و الثلاثي في مجموعات البيانات

التصنيف الثنائي	التصنيف الثلاثي	مجموعة البيانات
52.3 %	49 %	Shami-Senti
80.6 %		LABR 2 غير المتوازنة
53.1 %		LABR 2 المتوازنة
	60 %	LABR 3
70.7 %	59.3 %	ASTD

جدول 10.2: مصفوفة الارتباك لنموذج Kaggle للتصنيف الثنائي و الثلاثي في مجموعات البيانات

مجموعة بيانات LABR2 غير المتوازنة				مجموعة بيانات ASTD			
المتوقع				المتوقع			
سلبية	إيجابية			سلبية	إيجابية		
387	8153	إيجابية	الحقيقي	51	5	إيجابية	الحقيقي
78	1591	سلبية		147	12	سلبية	

كانت نتائج الدقة عالية بالنسبة ل LABR2 غير المتوازنة و ASTD حيث كانت على التوالي % 80.6 و % 70.7 و بالعودة إلى مصفوفة الارتباك (الجدول 10.2) فإن

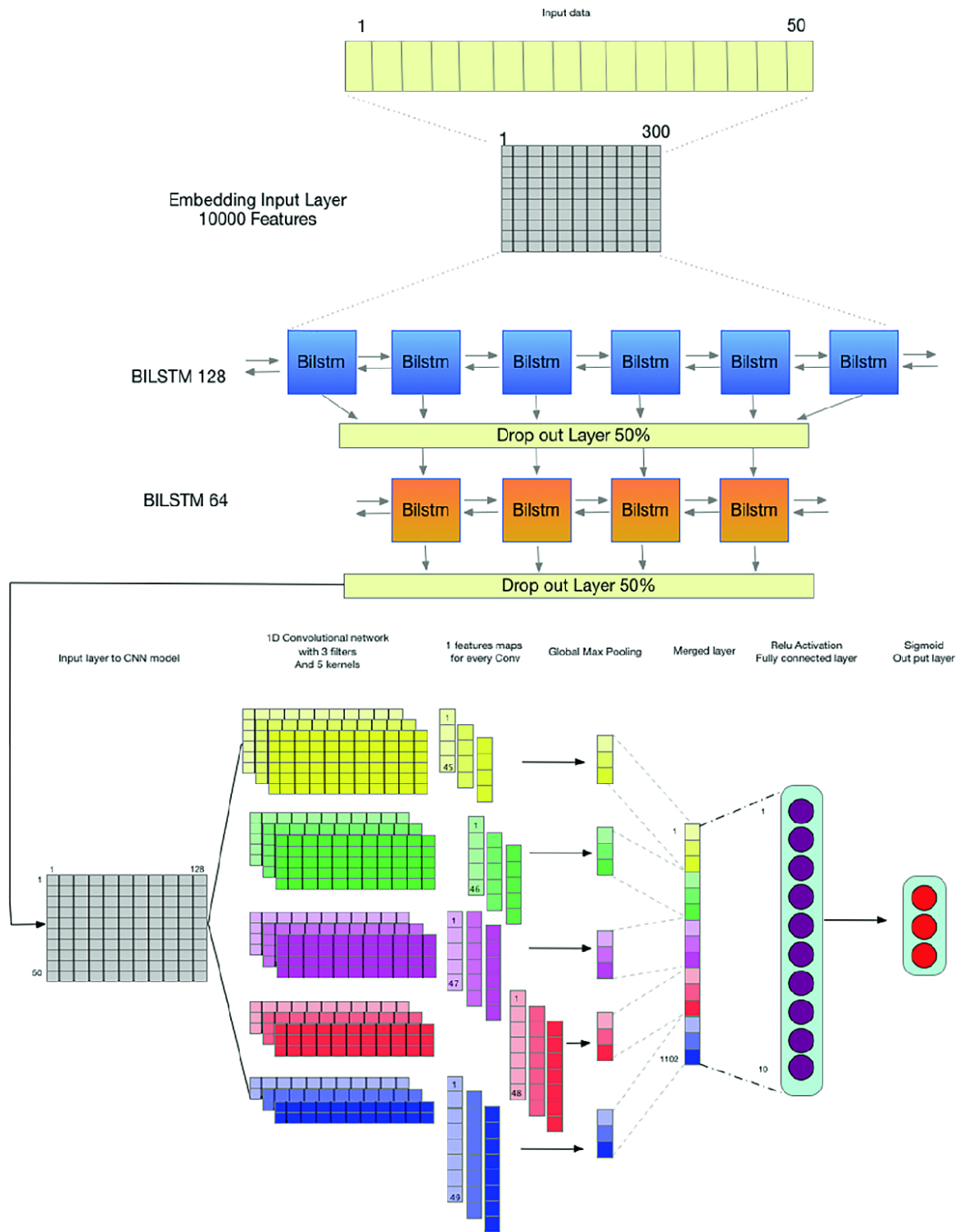
النموذج يتحيز في كل مرة إلى فئة الأغلبية مما أوقعه في مشكلة الإفراط في التعلم.

### نموذج BiLSTM-CNN المقترح من طرف Katherine et al.

تم اقتراح هذا النموذج بناء على نتائج نموذج Kaggle الذي أظهر أن المزج بين LSTM و CNN يعزز من دقة النموذج.

على عكس نموذج Kaggle فإن طبقة LSTM في هذا النموذج تسبق طبقة CNN، و بدل LSTM استخدم الباحثون BiLSTM من أجل استخراج الميزات من كلا الإتجاهين. و قد بينت النتائج أن هذا الترتيب أعطى نتائجاً أفضل بالنسبة لجميع مجموعات البيانات (الجدول 11.2).

هذا النموذج يحتوي على طبقة ترميز تم إنشاؤها باستخدام أشعة ترميز الكلمة المدربة مسبقاً بطول 300 و طبقتي BiLSTM ب 128 و 64 مخرج على التوالي و 0.5 تسرب. يتم إدخال مخرجات طبقة BiLSTM الثانية إلى طبقات CNN متوازية مع خمس أنوية [2,3,4,5,6] و 3 مرشحات [32,64,128]، كل طبقة CNN تتبع بطبقة تجميع أقصى شامل، يتم دمج جميع المخرجات في شعاع ذو بعد واحد في طبقة متسلسلة في نهاية شبكة CNN، و يدخل هذا الشعاع في طبقة Relu متصلة بالكامل مع 10 مخرجات و في الأخير استعملوا طبقة سينية مع 3 مخرجات في التصنيف الثلاثي و مخرج واحد في التصنيف الثنائي، الشكل 4.2.



شكل 4.2: النموذج المقترح [KSCD19].

حقق هذا النموذج دقة تتراوح ما بين 81% و 93% للتصنيفات الشائبة و 66% إلى 76% للتصنيفات الثلاثية. بالنسبة لبيانات LABR 2 غير المتوازنة قد حققت نسبة عالية تبلغ 80.2%، و هذا راجع بحسب وصفهم إلى أنها لا تتعلم جيدا بسبب

مشكلة عدم توازن البيانات مع أنها تحتوي على حجم معقول من البيانات. النتائج موضحة في الجدول 11.2.

جدول 11.2: دقة النموذج المقترح و النتائج المقارنة مع النموذجين الأساسيين لتصنيفات الشائبة و الثلاثية

التصنيف الشائبي			التصنيف الثلاثي			مجموعة البيانات
LSTM	Kaggle	النموذج المقترح	LSTM	Kaggle	النموذج المقترح	
54.5 %	25.3 %	93.5 %	53 %	49 %	76.4 %	Shami-Senti
55.34 %	80.6 %	80.2 %				LABR 2 غير المتوازنة
81 %	53.1 %	81.14 %				LABR 2 المتوازنة
			41.9 %	60 %	66.42 %	LABR 3
68.5 %	70.7 %	85.58 %	53 %	59.3 %	68.62 %	ASTD

كما تم استعمال فكرة الدمج بين الشبكات العصبية التلافيفية و الشبكات العصبية المتكررة أيضا في [APEI18]، حيث قام الباحثون في هذا العمل بتجربة نموذجهم المقترح (CNN-LSTM Arabic Sentiment Analysis Model) على أربع مجموعات بيانات و المتمثلة في (ASTD, Ar-Twitter, Main-AHS and Sub-AHS). بالنسبة للمعالجة المسبقة للبيانات إعتمدوا على ثلاث مستويات للترميز، الأول هو مستوى الحرف حيث كل حرف من كل كلمة من البيانات يمثل بشعاع، أما المستوى الثاني فهو مستوى (Ch5gram-level) حيث تمثل كل كلمة مكونة من خمسة أحرف بشعاع، و أخيرا مستوى الكلمة حيث تمثل كل كلمة من مجموعة البيانات بشعاع. النتائج المتحصل عليها بعد تجربة النموذج على البيانات الثلاث في المستويات الثلاث موضحة في الجدول (12.2).

جدول 12.2: دقة نموذج CNN-LSTM لمجموعات البيانات الأربع في مستويات الترميز الثلاث

ASTD	Ar-Twitter	Sub-AHS	Main-AHS	مستويات الترميز
0.7419	0.8131	0.9164	0.8941	مستوى الحرف
0.7762	0.8283	0.9568	0.9163	مستوى Ch5-gram
0.7641	0.8810	0.9510	0.9424	مستوى الكلمة

### 3.6.2 الأنظمة الهجينة

هي أنظمة تعتمد على الدمج بين الأنظمة المستندة على القواعد و الأنظمة المستندة على التعلم الآلي.

تم اعتماد هذا النهج في [SR15] لتحليل المشاعر لهجة المصرية على مستوى الجملة، حيث تم دمج تقنية التعلم الآلي SVM و نهج التوجيه الدلالي (semantic orientation)، و أعطى هذا التهجين نتائج واعدة.

أيضا اعتمد التهجين في تحليل المشاعر لمراجعات أفلام [ACCF16]، اشتمل هذا التهجين على القواعد الدلالية (semantic rules) ، fuzzy sets و تقنية التعلم الآلي غير الخاضع للإشراف. بلغت دقة هذا النموذج 76%

## 7.2 تحديات تحليل المشاعر

كان هناك تطور كبير في الأبحاث على تحليل المشاعر في العقد الماضي بهدف الوصول إلى مصنف مشاعر دقيق قدر الإمكان. ظهرت الكثير من المشاكل التي أصبحت تشكل تحديات يطمح الباحثون إلى حلها. فيمايلي نذكر بعضا من هذه التحديات :

### المشاعر الضمنية و السخرية Implicit Sentiment and Sarcasm

"كيف يمكن لأي شخص شراء هذا المنتج " و "يجب على المرء أن يشكك في قوة عقل الكاتب الذي كتب هذا الكتاب"، هذه الجمل لا تحتوي على أي كلمات سلبية لكنها تحمل مشاعر سلبية ضمنية تفهم من سياق الجملة، بالنسبة للجملة الثانية فهي تحمل معاني سخرية بمشاعر سلبية تفهم أيضا من سياق الكلام.

### إحباط التوقعات Thwarted Expectations

"كان يجب أن يكون الفلم رائعا، فالممثلون كلهم من الدرجة الأولى، و طاقم الإخراج و التصوير جيد"، هناك الكثير من الكلمات الإيجابية في هذه الجمل، لكن المشاعر العامة حول الفلم سلبية بسبب الجملة الأولى التي تحتوي على مشاعر سلبية تؤدي إلى إحباط ما هو متوقع حول الفلم.

## الثقافة العامة World knowledge

في بعض الأحيان يجب أن يكون للمرء ثقافة عامة واسعة لكي يستطيع فهم المشاعر المعبر عنها في بعض الجمل مثل "صرح دونالد ترامب بقرار جديد يشبه قراراته المعتادة" هنا يجب على الشخص أن يكون على اطلاع على قرارات دونالد ترامب ليكون قادراً على فهم المشاعر في الجملة.

## تحديد الكيان Entity Detection

قد يكون هناك كيانات متعددة في الجملة ما قد يصعب تحديد مشاعر الجملة، لذا من المهم تحديد الكيان المقصود بالمشاعر. "هاتف كوندور أفضل من هاتف دوجي"، هذه الجملة تحمل مشاعر إيجابية بالنسبة لكيان كوندور و سلبية بالنسبة لكيان دوجي.

## النفى Negation

تعتبر معالجة النفي من أصعب المشاكل في تحليل المشاعر لأنه قد يعكس قطبية الجمل في بعض الأحيان، كما هو الحال في الجملة "أنا لا أحب الذهاب للتسوق"، لكن على سبيل في الجملة "لا أحب التسوق فحسب بل أعشقه"، كلمة "فحسب" حالت دون عكس قطبية الجملة رغم وجود أداة النفي لذا يجب مراعاة هذا النوع من التركيبات لأدوات النفي مع كلمات مثل فقط، فحسب...، و في المثال "أنا لا أحب التسوق لكن أحب التنزه في الخارج" نلاحظ أنه تم عكس قطبية الجملة حتى كلمة "لكن" فقط، أي يجب الأخذ بعين الاعتبار أن النفي قد يؤثر على جزء من الجملة فقط.

## 8.2 خاتمة

رأينا في هذا الفصل بعض المفاهيم الأولية حول تحليل المشاعر و معالجة اللغة الطبيعية، و ألقينا نظرة على بعض الأمثلة لتطبيق تحليل المشاعر في الأنظمة المستندة على القواعد و الأنظمة المستندة على التعلم الآلي، كما أشرنا إلى بعض تحديات هذا المجال.

## الفصل الثالث

### دراسة تجريبية Implementation

#### 1.3 مقدمة

يمثل هذا الفصل تجربتنا لنموذج تحليل المشاعر للغة العربية القائم على الشبكات العصبية التلافيفية و الشبكات العصبية المتكررة BiLSTM-CNN-LSTM.

نموذجنا مستوحى من النموذج الفائز بمسابقة Kaggle لتحليل المشاعر (CNN-LSTM) [?] و النموذج المقترح في مقال [KSCD19] و هو نموذج (BiLSTM-CNN) حيث قررنا دمج هاتين الفكرتين.

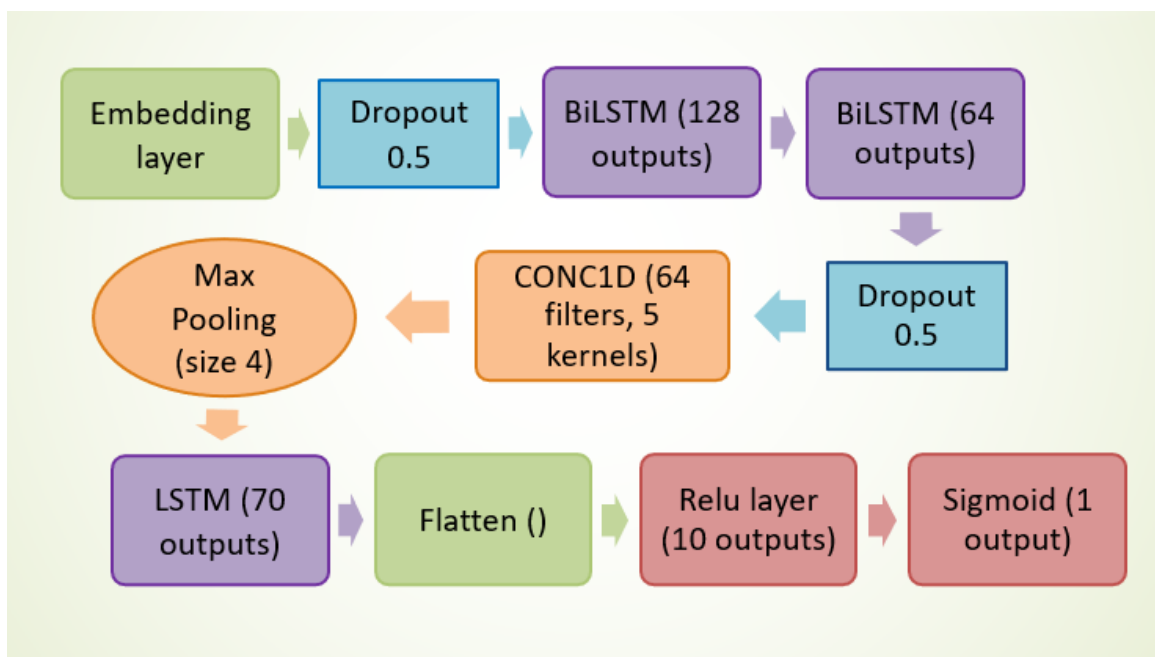
في هذا الفصل سنتعرف على مجموعة البيانات المختارة و هندسة الشبكة المعتمدة و النتائج المتحصل عليها.

#### 2.3 هندسة الشبكة

مما رأينا سابقا في نتائج النموذج المقترح في [KSCD19] LSTM-CNN ونتائج نموذج Kaggle الفائز في مسابقة Kaggle لتحليل المشاعر فإنهما أعطيا نتائج جيدة لذا إفترضنا أن الدمج بين الفكرتين قد يعطي نتائج جيدة أيضا.

تتكون شبكة نموذجنا من طبقة ترميز Embedding layer، طبقتين من الشبكة المتكررة بذاكرة طويلة-قصيرة المدى ثنائية الإتجاه (BiLSTM)، طبقة تلافيفية (CNN) و أخيرا طبقة متكررة بذاكرة طويلة-قصيرة المدى (LSTM). الشكل (1.3)





شكل 1.3: هندسة الشبكة

بعد العديد من التجارب التي غيرنا خلالها المعطيات (parameters) قررنا اختيار المعطيات المبينة في الجدول (1.3) و التي أعطت أفضل أداء لنموذجنا.

جدول 1.3: المعطيات المعتمدة في النموذج

128	حجم الترميز
100	الطول المحدد للسلسلة
128	حجم شبكة BiLSTM الأولى
64	حجم شبكة BiLSTM الثانية
70	حجم شبكة LSTM
0.5	التسرب
64	المرشح
5	حجم النواة
4	حجم التجميع الأقصى
30	Batch size
2	الفترات

من أجل بناء النموذج استخدمنا نموذج متسلسل (Sequential) الذي يسمح بإنشاء نماذج طبقة تلو طبقة.

الطبقة الأولى هي عبارة عن طبقة ترميز تستقبل الإدخال المتمثل في سلاسل بطول 100 و 30000 كلمة التي تعبر عن الميزات، متبوعة بطبقة تسرب بمعدل 0.5 يعني أن 50% من الخلايا العصبية سيتم إيقافها بشكل عشوائي و يساعد هذا في تجنب الوقوع في الإفراط في التعلم.

تليها طبقتي BiLSTM ب 128 و 64 مخرج على التوالي متبوعتين أيضا بطبقة تسرب بمعدل 0.5. بعد ذلك نضيف طبقة تلافيفية أحادية (CONV1D) بنية شبكة تلافيفية معروفة جيدا للتعامل مع التسلسلات) بمرشح 64 و حجم نواة 5 و وظيفة تنشيط Relu، تليها طبقة تجميع أقصى ذو بعد 4.

يلي ذلك طبقة LSTM ب 70 مخرج متبوعة بطبقة تسطح لإدراج جميع المخرجات في شعاع ذو بعد 1. في الأخير لدينا طبقتي إتصال بالكامل، الأولى بوظيفة تنشيط Relu مع 10 مخرجات و الثانية بوظيفة تنشيط سينية (Sigmoid) كون التصنيف ثنائي. استخدمنا Adam كمحسن و binary\_crossentropy كدالة خطأ و الدقة (Accuracy) كمقياس نهائي لتقييم النموذج.

### 3.3 مجموعة البيانات

اخترنا مجموعة البيانات Arabic ULMPiT Model [sit] المعدة من طرف عابد الخولي و هي متوفرة على منصة Kaggle لمسابقات علوم البيانات. مجموعة البيانات هذه مأخوذة من [EKE18]، و هي عبارة عن مراجعات لفنادق مقيمة من 1 إلى 5، في عملنا قمنا بتصنيف المراجعات ذات التقييمات 1 و 2 إلى سلبية، و ذات التقييمات 3، 4، 5 إلى إيجابية. تحتوي هذه المجموعة على 105697 مراجعة، اعتمدنا على 80% منها للتدريب و 20% للاختبار.

### 4.3 معالجة البيانات

من أجل معالجة البيانات قمنا ب:

إزالة جميع الأحرف غير العربية، علامات الترقيم، الرموز.

ثم قمنا بتحويل النص إلى سلسلة من الأعداد الصحيحة :

```
Tokenizer(nb_words, split=' ')
```

`nb_words`: العدد الأقصى للكلمات التي يجب الاحتفاظ بها بناء على تكرارها.

`split`: نوع الفاصل بين كلمة و كلمة.

و من خلال ذلك قمنا بإنشاء قاموس المفردات بناء على تكرار الكلمات:

```
Tokenizer.fit_on_texts()
```

استبدلنا كل كلمة في السلسلة النصية بالعدد الذي يقابلها في قاموس المفردات:

```
Tokenizer.texts_to_sequences()
```

و للتأكد من أن جميع التسلسلات في القائمة لها نفس الطول استخدمنا مايلي:

```
pad_sequences(X, maxlen)
```

`maxlen`: الطول المحدد للسلسلة.

بعد استبدال التقييمات (1-5) لفئتين إيجابية و سلبية قمنا بتحويل هاتين الفئتين إلى

متغيرات وهمية عن طريق الدالة:

```
pd.get_dummies().values
```

مثال:

ليكن لدينا مجموعة الجمل الآتية:

```
docs = ['هذا احسن عمل قمت به',
        'عمل جيد جدا',
        'جهد عظيم لكن نتيجة سيئة',
        'عمل جميل جدا',
        'هذا العمل الذي قمت به ممتاز جدا']
```

بعد تطبيق `D = Tokenizer.fit_on_texts(docs)` نتحصل على قاموس المفردات الآتي:

```
{'عمل': 1, 'جدا': 2, 'هذا': 3, 'قمت': 4, 'به': 5, 'احسن': 6, 'جيد': 7, 'جهد': 8, 'عظيم': 9, 'لكن': 10, 'نتيجة': 11, 'سيئة': 12, 'جميل': 13, 'العمل': 14, 'الذي': 15, 'ممتاز': 16}
```

و باستخدام `X = Tokenizer.texts_to_sequences(D)` تستبدل كل كلمة بما يقابلها في القاموس:

```
[[3, 6, 1, 4, 5], [1, 7, 2], [8, 9, 10, 11, 12], [1, 13, 2], [3, 14, 15, 4, 5, 16, 2]]
```

ثم نطبق `pad_sequences(X, maxlen)` بطول سلسلة (`maxlen`) يساوي 10 فنحصل في النهاية على التسلسلات الآتية:

```
[[ 0  0  0  0  0  3  6  1  4  5]
 [ 0  0  0  0  0  0  0  1  7  2]
 [ 0  0  0  0  0  8  9 10 11 12]
 [ 0  0  0  0  0  0  0  1 13  2]
 [ 0  0  0  3 14 15  4  5 16  2]]
```

### 5.3 بيئة العمل Environment

من أجل تنفيذ نموذجنا استعنا بمنصة Kaggle لمسابقات علوم الكمبيوتر [Kaga]، وهي منصة مفتوحة لعلماء البيانات و ممارسي التعلم الآلي، توفر مجموعات بيانات إضافة إلى بيئة عمل سحابية (Cloud based environment) تدعم لغات برمجة مثل Python, R.

تتيح هذه المنصة وصولاً مجانياً ب ( GPU (NVIDIA TESLA P100 و 16 GB ذاكرة وصول عشوائية (RAM)، مساحة تخزين لكل مشروع (4.9 GB) و تقتصر كل جلسة على 9 ساعات.

إستعملنا Python كلغة برمجة لتوفرها على مجموعة واسعة من المكتبات (Libraries) تسهل عمل المبرمج في مجال التعلم العميق:

Tensorflow: مكتبة أساسية يمكن استخدامها لإنشاء نماذج التعلم العميق مباشرة أو عن طريق مكتبات مبسطة مبنية عليها.

Keras: هي مكتبة قوية و سهلة الإستخدام تسمح بتطوير و تدريب نماذج الشبكات العصبية، مطورة من أجل إجراء تجارب سريعة، قابلة للتفاعل مع Tensorflow.

Pandas: مكتبة Python لمعالجة و تحليل البيانات.

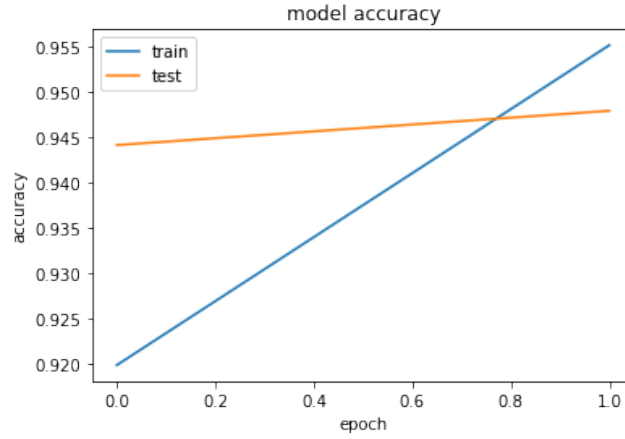
Sickit-learn: مكتبة رئيسية في Python تستخدم عادة في مشاريع التعلم الآلي، تتميز بخوارزميات تصنيف، انحدار (regression) و تجميع (clustering). و هي مكتبة مصممة للتعامل مع المكتبات الرقمية و العلمية Numpy، Scipy.

Numpy: هي مكتبة حسابية علمية متخصصة في معالجة المصفوفات، كما أن لديها وظائف في مجال الجبر الخطي (Linear Algebra)، مشروع مفتوح المصدر يمكن استخدامه بحرية من Python.

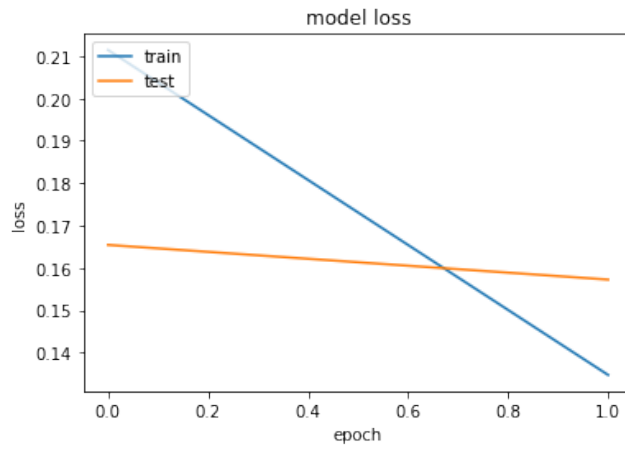
### 6.3 مناقشة النتائج

تشرح المنحنيات أدناه ( الشكل 2.3) قيمتي الدقة و الخطأ خلال فترات التدريب في مجموعة التدريب و مجموعة الإختبار و هي على التوالي 0.96, 0.12 و 0.95, 0.15.

نلاحظ أنه لا يوجد فرق كبير بين الدقة في مجموعة التدريب و مجموعة الإختبار و نسبة الخطأ في مجموعة الإختبار تعتبر مقبولة و هذا يدل على أن النموذج يعمل بشكل جيد.



(أ) منحنى الدقة



(ب) منحنى الخطأ

شكل 2.3: الدقة التي حصل عليها نموذجنا في الأعلى وفي الأسفل نسبة خطأ النموذج لكل من مجموعات الإختبار و التدريب أثناء تدريب النموذج.

```
705/705 - 40s
          precision    recall  f1-score   support

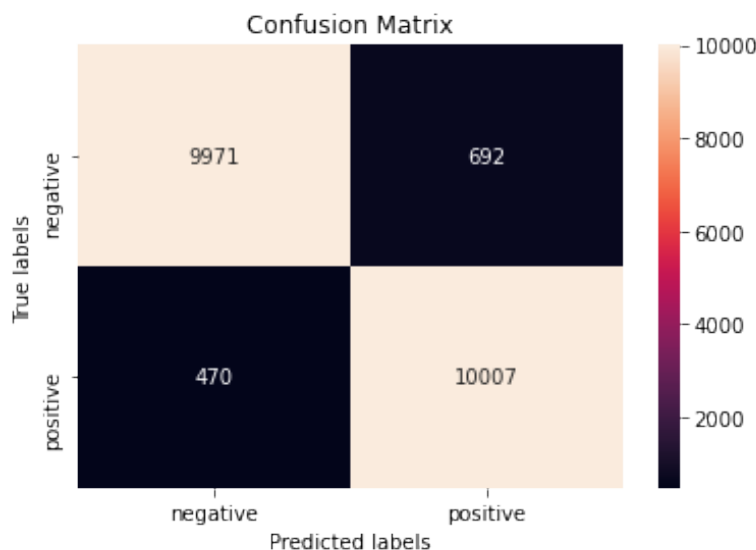
negative   0.95      0.94      0.94     10663
positive   0.94      0.96      0.95     10477

accuracy               0.95     21140
```

شكل 3.3: قيم مقاييس التقييم المتحصل عليها في الإختبار

أعطى النموذج نتائج جيدة في جميع قيم مقاييس التقييم لجميع الفئات مقارنة لدقة الإختبار (الشكل 3.3)، مما يعني أن النموذج مستقر (لا توجد علامة على الإفراط في التعلم) لأنه أعطى نتائج جيدة أثناء الإختبار.

يظهر الشكل 4.3 مصفوفة الارتباك Confusion matrix و التي تمكننا من تقييم مدى دقة أداء النموذج، توضح المصفوفة تصنيف الآراء لكل فئة حيث تم تصنيف 10007 من 10663 مراجعة بشكل صحيح بالنسبة للفئة الإيجابية و 9971 من أصل 10477 بالنسبة للفئة السلبية.



شكل 4.3: مصفوفة الارتباك

أعطى نموذجنا نتائج جيدة بالنسبة للغة العربية فقد بلغت نسبة الدقة 95% و نسبة الخطأ 15% على الرغم من عدم تجذير الكلمات (stemming) و الإبقاء على الكلمات المستبعدة (stop words) و عدم الأخذ بخصائص اللغة العربية و التي نرى أنها عوامل يمكن أن تحسن من أداء النموذج.

### 7.3 خاتمة

تم في هذا الفصل إنشاء نموذج لتحليل المشاعر على مجموعة بيانات لمراجعات فنادق باللغة العربية، حيث قمنا بشرح الخطوات التي اتبعناها في الدراسة التجريبية.

حقق نموذجنا نتائج جيدة بدقة وصلت إلى 95% و نسبة خطأ قدرت ب 15% على مجموعة بيانات Arabic ULMFiT Model.

كانت النتائج جيدة و لكن نرى أنه بالإمكان تحسينها إذا نظرنا في خصائص اللغة العربية  
و إستعمال التجدير و نزع الكلمات المستبعدة...



## خاتمة

تناولت مذكرتنا موضوع تحليل المشاعر باستعمال نهج التعلم العميق، درسنا هذا الموضوع حول اللغة العربية لأنه على الرغم من أن مجال تحليل المشاعر مجال بحثي واسع إلا أنه لم يتطرق إليه كثيرا بالنسبة للغة العربية.

لدراسة هذا الموضوع بدأنا بتغطية الشبكات العصبية الإصطناعية و أساليب التعلم العميق خاصة الشبكات العصبية التلافيفية و الشبكات العصبية المتكررة، ثم بعد ذلك انتقلنا لتحليل المشاعر حيث عالجنا بعض المفاهيم الأساسية المتعلقة به بالإضافة إلى بعض الدراسات السابقة حوله في الأنظمة المستندة على القواعد و الأنظمة المستندة على التعلم الآلي.

و كمساهمة متواضعة منا اقترحنا نموذجا لتحليل المشاعر للغة العربية باستعمال الشبكات العصبية المتكررة و الشبكات العصبية التلافيفية على مجموعة بيانات Arabic ULMFiT Model.

حقق نموذجنا نتائجاً بدقة 95% و نسبة خطأ قدرت ب 15%.

على الرغم من أن النتائج تبدو مرضية إلا أننا نرى أن الأخذ بخصائص اللغة العربية كالتشكيل و استعمال تجذير الكلمات (Stemming) و التخلص من الكلمات المستبعدة أيضا يمكن أن يساهم في تحسين أداء النموذج و لو كان لدينا وقت أكثر لتوسعنا في البحث أكثر.

كان بودنا أن نعمل على مجموعة بيانات أكبر و أكثر أهمية أيضا، لكن قلة البيانات المنظمة على اللغة العربية حالت دون ذلك.

## المراجع العلمية

- [AA13] Mohamed Aly and Amir Atiya. Labr: A large scale arabic book reviews dataset. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages ,498–494 2013.
- [ACCF16] Orestes Appel, Francisco Chiclana, Jenny Carter, and Hamido Fujita. A hybrid approach to sentiment analysis. In *2016 IEEE Congress on Evolutionary Computation (CEC)*, pages .4957–4950 IEEE, 2016.
- [AFGG15] D Alessia, Fernando Ferri, Patrizia Grifoni, and Tiziana Guzzo. Approaches, tools and applications for sentiment analysis implementation. *International Journal of Computer Applications*, 125(3), 2015.
- [APEI18] Abdulaziz M Alayba, Vasile Palade, Matthew England, and Rahat Iqbal. A combined cnn and lstm model for arabic sentiment analysis. In *International Cross-Domain Conference for Machine Learning and Knowledge Extraction*, pages .191–179 Springer, 2018.
- [AS98] Raj Aggarwal and Yonghua Song. Artificial neural networks in power systems. ii. types of artificial neural networks. *Power Engineering Journal*, ,47–41:(1)12 1998.
- [Bro18] Jason Brownlee. *Better Deep Learning: Train Faster, Reduce*

- 
- Overfitting, and Make Better Predictions*. Machine Learning Mastery, 2018.
- [BSF94] Yoshua Bengio, Patrice Simard, and Paolo Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, ,166–157:(2)5 1994.
- [cou] understanding-dropout-yagbr. <https://www.coursera.org/lecture/deep-neural-network/understanding-dropout-YaGbR>.
- [DJ19] Jurafsky Daniel and H. Martin James. *Speech and Language Processing*. Cambridge University Press, 10 2019.
- [EKE18] Ashraf Elnagar, Yasmin S Khalifa, and Anas Einea. Hotel arabic-reviews dataset construction for sentiment analysis applications. In *Intelligent Natural Language Processing: Trends and Applications*, pages .52–35 Springer, 2018.
- [GBC16] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.
- [Haj16] Dr.Mohammed Moustafa Hajouz. Sentiment analysis in twitter. 2016.
- [HGK18] Alex Hernández-García and Peter König. Data augmentation instead of explicit regularization. *arXiv preprint arXiv:1806.03852*, 2018.
- [HOT06] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural computation*, ,1554–1527:(7)18 2006.

- 
- [HS97] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, ,1780–1735:(8)9 1997.
- [HS13] Michiel Hermans and Benjamin Schrauwen. Training and analysing deep recurrent neural networks. In *Advances in neural information processing systems*, pages ,198–190 2013.
- [HW59] David H Hubel and Torsten N Wiesel. Receptive fields of single neurones in the cat’s striate cortex. *The Journal of physiology*, ,591–574:(3)148 1959.
- [HZRS16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778 2016.
- [Img] Imagenet larage scale visuel recognition challenge[ilsvrc]. <http://www.image-net.org/challenges/LSVRC/>.
- [ITSPCA15] Li Im Tan, Wai San Phang, Kim On Chin, and Patricia Anthony. Rule-based sentiment analysis for financial news. In *2015 IEEE International Conference on Systems, Man, and Cybernetics*, pages .1606–1601 IEEE, 2015.
- [JL06] Nitin Jindal and Bing Liu. Mining comparative sentences and relations. In *Aaai*, volume 22, page 9, 2006.
- [Kaga] Kaggle. <https://www.kaggle.com>.
- [Kagb] Kaggle model. <https://www.kaggle.com/monsterspy/conv-lstm-sentiment-analysis-keras-acc-96-0>.

- 
- [KB14] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [KGC17] Jan Kukačka, Vladimir Golkov, and Daniel Cremers. Regularization for deep learning: A taxonomy. *arXiv preprint arXiv:1710.10686*, 2017.
- [KPF09] Manfred Klenner, Stefanos Petrakis, and Angela Fahrni. Robust compositional polarity classification. 2009.
- [Kri07] David Kriesel. A brief introduction on neural networks. 2007.
- [KSCD18] Kathrein Abu Kwaik, Motaz Saad, Stergios Chatzikyriakidis, and Simon Dobnik. Shami: A corpus of levantine arabic dialects. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.
- [KSCD19] Kathrein Abu Kwaik, Motaz Saad, Stergios Chatzikyriakidis, and Simon Dobnik. Lstm-cnn deep learning model for sentiment analysis of dialectal arabic. In *International Conference on Arabic Language Processing*, pages .121–108 Springer, 2019.
- [KSH12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages .1105–1097 Curran Associates, Inc., 2012.
- [LBH15] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, ,444–436:(7553)521 2015.
- [Lid01] Elizabeth D Liddy. Natural language processing. 2001.

- 
- [Liu15] Bing Liu. *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press, 2015.
- [MCCD13] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
- [Mer] Merriam webster dictionary. <https://www.merriam-webster.com/>.
- [MG<sup>+</sup>06] Gilad Mishne, Natalie S Glance, et al. Predicting movie sales from blogger sentiment. In *AAAI spring symposium: computational approaches to analyzing weblogs*, pages ,158–155 2006.
- [MGL09] Prem Melville, Wojciech Gryc, and Richard D Lawrence. Sentiment analysis of blogs by combining lexical knowledge with text classification. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages ,1284–1275 2009.
- [MP43] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, ,133–115:(4)5 1943.
- [MW12] Frederick Mosteller and David L Wallace. *Applied Bayesian and classical inference: the case of the Federalist papers*. Springer Science & Business Media, 2012.
- [NAA15] Mahmoud Nabil, Mohamed Aly, and Amir Atiya. Astd: Arabic sentiment tweets dataset. In *Proceedings of the 2015 conference on empirical methods in natural language processing*, pages 2515–,2519 2015.

- 
- [NH04] Kamal Nigam and Matthew Hurst. Towards a robust metric of opinion. In *AAAI spring symposium on exploring attitude and affect in text*, volume 598603, 2004.
- [PGCB13] Razvan Pascanu, Caglar Gulcehre, Kyunghyun Cho, and Yoshua Bengio. How to construct deep recurrent neural networks. *arXiv preprint arXiv:1312.6026*, 2013.
- [PLV02] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up? sentiment classification using machine learning techniques. *arXiv preprint cs/0205070*, 2002.
- [PMB12] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. Understanding the exploding gradient problem. *CoRR*, *abs/1211.5063*, 2:417, 2012.
- [PP14] Sharat Prasad and Piyush Prasad. Deep recurrent neural networks for time series prediction. 07 2014.
- [SAWN14] Martin Sundermeyer, Tamer Alkhouli, Joern Wuebker, and Hermann Ney. Translation modeling with bidirectional recurrent neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages ,25–14 2014.
- [SBMN13] Richard Socher, John Bauer, Christopher D Manning, and Andrew Y Ng. Parsing with compositional vector grammars. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 455–,465 2013.

- [SEEB17] Abu Bakr Soliman, Kareem Eissa, and Samhaa R El-Beltagy. Aravec: A set of arabic word embedding models for use in arabic nlp. *Procedia Computer Science*, ,265–117:256 2017.
- [sit] Arabic ulmfit model. <https://www.kaggle.com/abedkhooli/arabic-ulmfit-model/discussion>.
- [SP97] Mike Schuster and Kuldeep K Paliwal. Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing*, ,2681–2673:(11)45 1997.
- [SR15] Amira Shoukry and Ahmed Rafea. A hybrid approach for sentiment classification of egyptian dialect tweets. In *2015 First International Conference on Arabic Computational Linguistics (ACLing)*, pages .85–78 IEEE, 2015.
- [Tou92] Claude Touzet. *les réseaux de neurones artificiels, introduction au connexionnisme*. 1992.
- [TSSW10] Andranik Tumasjan, Timm O Sprenger, Philipp G Sandner, and Isabell M Welpe. Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Fourth international AAAI conference on weblogs and social media*, 2010.
- [Tur02] Peter D Turney. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. *arXiv preprint cs/0212032*, 2002.
- [VD18] Subir Varma and Sanjiv Das. *Introduction to Deep Learning*. 04 2018.
- [WWH05] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceed-*



---

*ings of human language technology conference and conference on empirical methods in natural language processing*, pages ,354–347 2005.

[ZL11] Lei Zhang and Bing Liu. Identifying noun product features that imply opinions. In *Proceedings of the 49th annual meeting of the Association for Computational Linguistics: human language technologies*, pages ,580–575 2011.

[ZSSR17] Valentino Zocca, Gianmario Spacagna, Daniel Slater, and Peter Roelants. *Python Deep Learning*. Packt Publishing Ltd, 2017.

[ 13] and . , .2013