

تحليل المصفوفات إلى عوامل في أنظمة الاقتراحات

أحمد حاج قدور، الياس مخايل، فيليب نخله، محمد شاكر اللحام
دمشق؛ جامعة دمشق؛ كلية الهندسة المعلوماتية – قسم الذكاء الصناعي

نبذة مختصرة

عادة ما تقوم أنظمة الاقتراحات بإعطاء اقتراحات شخصية لكل مستخدم بالاعتماد على تقييمات المستخدم للمنتج والتغذية الراجعة كما تعتمد أيضاً على بيانات مساعدة. تحليل المصفوفات إلى عوامل هو الفكرة الأساسية للتنبؤ بتقييم المستخدم الشخصي بناءً على مجموعة من المنتجات من أجل مستخدم واحد اعتماداً على التشابه بين المستخدمين والمنتجات. في هذه الورقة البحثية سنتناول نموذج جديد يعتمد على تحليل المصفوفة إلى عوامل باستخدام الشبكات العصبونية. أولاً، سنقوم ببناء مصفوفة مستخدم-منتج مع تقييمات المستخدمين للمنتجات و دون تفضيلات للتغذية الراجعة. مع هذه المصفوفة التي ستكون دخل المسألة، سنمثل بنية عميقة قابلة للتعلم على فضاء صغير الأبعاد من أجل تمثيل المستخدمين والمنتجات. ثانياً، سنقوم بتصميم تابع الخطأ اعتماداً على الإنترنت النائية المتقاطعة الذي سنستخدم خلاله على التقييمات الصريحة والتغذية الراجعة من أجل الحصول على نتائج أفضل. النتائج التجريبية لدينا تؤثر النموذج المطروح وتوابع الخطأ. وبالاعتماد على مجموعة بيانات معيارية، يمكن للنموذج المطروح أن يضاهي أحدث التقنيات المستخدمة. كما سنجري تجارب مكثفة من أجل تقييم الأداء عند استخدام أعدادات تجريبية مختلفة.

مقدمة

في عصر السرعة وازدياد تدفق المعلومات بشكل كبير، الحمل الزائد للمعلومات يعتبر من أكبر التهديدات التي تواجهنا وأقننا الحالي. تقوم أنظمة الاقتراحات بمعالجة هذه المعضلة كما تساعد في تحديد المعلومات التي تقوم بتقديمها للمستخدم كما تساعد المستخدمين على الإنترنت على الوصول بشكل أسرع إلى المعلومات المطلوبة التي تناسب احتياجاتهم [Sarwar et al. 2001]. أنظمة

الاقتراحات واسعة الانتشار في هذه الأيام في منصات التجارة الإلكترونية، على سبيل المثال اقتراح الكتب في أمازون، والأفلام في نتفلكس. إجراءات الاقتراح التي تعتمد على التصفية التشاركية تستخدم على شكل واسع في الأبحاث والصناعات. هي تعتمد على حدس بسيط أي أنه إذا قام المستخدم بتقييم المنتجات بشكل متشابه في الماضي فإنهم غالباً ما سيقومون بتقييم المنتجات بشكل متشابه في المستقبل [Sarwar et al. 2001]. ومن أشهر إجراءات التصفية التشاركية هي تحليل المصفوفة إلى عوامل الذي يتعلم على فضاء صغير الذي يمثل المستخدم والمنتج لسهولة التحليل والبساطة والمرونة. في هذا الفضاء، يقوم نظام الاقتراح يقوم باقتراح تقييمات شخصية على مجموعة من المنتجات من أجل كل مستخدم اعتماداً على التشابه بين المستخدمين وبين المنتجات.

التقييمات في فضاء مستخدم-منتج هي معرفة خارجية والتي يتم استغلالها بشكل عميق في أنظمة الاقتراح الحديثة. بسبب الاختلاف بين تقييمات المستخدمين للمنتجات فإننا نستخدم التحليل إلى عوامل بالإضافة إلى الانزياح لتحسين النتائج. من أجل تفادي ندرة التقييمات نقوم بإضافة بعض المعلومات الإضافية إلى المصفوفة على سبيل المثال عند استخدام مصفوفة اجتماعية فإننا نقوم بإضافة العلاقات بين الأشخاص [Ma et al. 2008].

على كل الأحوال فإن استخدام النماذج مع التقييمات التي يقوم بها المستخدمين فقط غير كاف ليقوم باقتراح أفضل عدد من المنتجات [Hu et al. 2008]. سنستخدم أيضاً على التغذية الراجعة الضمنية والمنتجات التي تم شرائها أو البحث عنها سابقاً في نظام الاقتراح الخاص بنا.

حالياً، نظراً للإمكانيات الكبيرة، أصبح بالإمكان تطبيق الإجراءات العميقة على نطاق واسع من التجارب ومنها الرؤيا الحاسوبية ومعالجة اللغات الطبيعية.

في هذه الورقة سنقوم باستخدام كل من كل من تقييمات المستخدمين و التغذية الراجعة، سنقوم بطرح نموذج

$$Y_{i,j} = \begin{cases} 0, & \text{if } R_{i,j} = unk \\ 1 & \text{Otherwise} \end{cases} \quad (1)$$

$$Y_{i,j} = \begin{cases} 0, & \text{if } R_{i,j} = unk \\ R_{i,j} & \text{Otherwise} \end{cases} \quad (2)$$

معظم الحلول الحالية تعتمد على المعادلة الأولى في بناء المصفوفة Y . في هذا البحث سنقوم ببناء هذه المصفوفة اعتماداً على المعادلة الثانية. حيث يكون تقييم المستخدم للمنتج محجوز في المصفوفة Y . نحن هنا نقوم باعتبار قيم التقييمات مهمة لدينا لأنها تشير إلى درجة تفضيل المستخدم للمنتج. بينما نقوم بوضع 0 في حال كون التقييم غير معروف و الذي هو سيمثل التغذية الرجعية التي سنقوم بقيمتها 0 بالبداية.

في أنظمة الاقتراحات عادة ما تكون المشكلة هي إيجاد تقييم للمنتجات الغير محجوزة في المصفوفة Y , يمكن باستخدام نموذجنا توليد تقييمات جديدة كالتالي.

$$\widehat{Y}_{i,j} = F(u_i, v_j | \theta) \quad (3)$$

حيث $\widehat{Y}_{i,j}$ هو خرج النموذج الذي لدينا الذي يتنبأ بتقييم المستخدم i للمنتج j بينما θ تمثل بارامترات تابع النموذج من أجل التنبؤ بالنتائج بينما F هو التابع الذي سيقوم بإعطاء النتائج اعتماداً على البارامترات. اعتماداً على هذا التابع يمكننا اقتراح مجموعة من المنتجات لمستخدم بحيث يكون رضاه أكبر ما يمكن.

الآن يصبح السؤال كيف يمكننا تعريف التابع F . يمكننا ببساطة تعريف التابع F على أنه الجداء الداخلي للشعاعين p_i و q_j حيث يعبر هذان الشعاعان عن تمثيل المستخدم و المنتج في الفضاء الجديد.

$$\widehat{Y}_{i,j} = F^{LFM}(u_i, v_j | \theta) = p_i^T \cdot q_j \quad (4)$$

سنقوم بكتابة بعض التعريفات التي سنستخدمها في القسم التالي و هي u للتعبير عن المستخدم و v للتعبير عن المنتجة و i و j هي أدلة المستخدمين و المنتجات بالترتيب. Y هي مصفوفة تقييمات المستخدمين للمنتجات التي تم تشكيلها ف المعادلة . بينما Y^+ تمثل التقييمات التي قام بها المستخدمين بينما Y^- هي التقييمات الغير معروفة و قيمتها 0 بينما $Y_{sampled}^-$ هي الأمثلة التي نعرف أجوبتها لكننا نستخدمها من أجل التدريب. أخيراً يشكل $Y^+ \cup Y_{sampled}^-$ جميع حالات التدريب.

لتحليل المصفوفة من أجل الحصول على أفضل N منتج مقترح. أولاً سنقوم ببناء مصفوفة مستخدم-منتج باستخدام تقييمات المستخدمين فقط دون وجود تغذية رجعية، والتي تختلف عن عدة اجرائيات أخرى تستخدم إما معطيات المستخدم أو التغذية الرجعية فقط. مع هذه المصفوفة التي ستكون دخل المسألة، سنقوم ببناء شبكة عصبونية سوف تتعلم على فضاء المعطيات لدينا. هذه البنية مستوحاة من نموذج دلالي عميق تم برهانه أنه سيكون مفيد في البحث على الويب [Huang et al., 2013], حيث يمكن ربط الاستعلامات مع البيانات في فضاء كامن عن طريق عدة طبقات من التنبؤات الغير خطية. بالإضافة إلى ذلك سنقوم بتصميم تابع خطأ اعتماداً على الإنتروبيا المتقاطعة والتي تضمن استخدام كل من البيانات الخارجة و التغذية الرجعية. في المجمل يمكن تلخيص ما سبق من خلال النقاط التالية.

- نهدف إلى تصميم نموذج جديد يعتمد على تحليل المصفوفات إلى عوامل بالإضافة إلى شبكة عصبونية التي تقوم بربط المستخدمين بالمنتجات عن طريق فضاء صغير الأبعاد دون إسقاطات خطية. نستخدم مصفوفة ثنائية تأخذ التقييمات دون تغذية رجعية أولاً كدخل للنموذج.
- سنقوم بتصميم تابع حساب خطأ الذي يأخذ بعين الاعتبار التقييمات و التغذية الرجعية من أجل نتائج أفضل.
- النتائج التجريبية ترينا أداء نموذجنا الذي يمكن أن يضاهي أحدث التقنيات في هذا المجال.

تقسيم هذه البحث سيكون كالتالي. في القسم الثاني سنقوم بمناقشة المسألة. كمان سنقوم بتعريف بنية النموذج في القسم الثالث. في القسم الرابع سنقوم بتجربة البيانات التي لدينا على النموذج و سنقوم بعرض النتائج. في النهاية بالقسم الخامس سنقوم بتلخيص ما سبق.

عرض المشكلة

يوجد لدينا M مستخدم نرمز لهم ب U حيث $U = \{u_1, u_2, \dots\}$ و N منتج نرمز لهم ب V حيث $V = \{v_1, v_2, \dots\}$. بينما ستكون المصفوفة $R \in \mathbb{R}^{N \times M}$ تمثل مصفوفة التقييمات حيث $R_{i,j}$ يمثل تقييم المستخدم i للمنتج j و سنضعه unk اذا كان التقييم غير معروف. هناك طريقتين لبناء مصفوفة تفاعل المستخدم مع المنتجات والتي سنعتبر عنها ب $Y \in \mathbb{R}^{N \times M}$ المأخوذة من R كالتالي،

النموذج المطروح

في هذا القسم سنتكلم أولاً عن النموذج الدلالي الذي استوحينا منه نموذجنا حيث سنقوم بتعريف بنية النموذج و سنقوم أخيراً بالتحدث عن تابع حساب الخطأ الذي لدينا و خوارزمية التدريب.

النموذج الدلالي العميق

تم عرض هذا النموذج في [Haung et al., 2013] من أجل البحث على الويب. يستخدم هذا النموذج شبكة عصبونية عميقة لتصنيف مجموعة من البيانات من أجل الاستعلام المطلوب. في البداية يقوم هذا النموذج بربط الاستعلام و البيانات في فضاء عام صغير الأبعاد دون استخدام إسقاطات خطية. بعدها من أجل تقييم البحث العلاقة بين الاستعلام و كل البيانات يتم حسابها عن طريق تجنب التشابه بين التشابه بين الأشعة في الفضاء الذي لدينا و البيانات. تقوم الشبكة العصبونية بالتدرب بشكل متزايد حتى يصبح الخطأ أقل ما يمكن و تصبح الاستعلامات قريبة جداً من القيم الحقيقية.

تم تطبيق النموذج السابق تبعاً لوضع المستخدمين [Elkahky et al., 2015]. بشكل مختلف عن عملنا بحيث أن هذا النموذج يركز على وضع المستخدمين مع الكثير من الخصائص ك تاريخ البحث و الاستعلامات السابقة. بينما نحن سنستخدم فقط التقييمات المحجوزة و التغذية الرجعية بما أننا فقط سنركز على عدد محدد من الاقتراحات.

النموذج العميق الذي يعتمد على تحليل المصفوفات

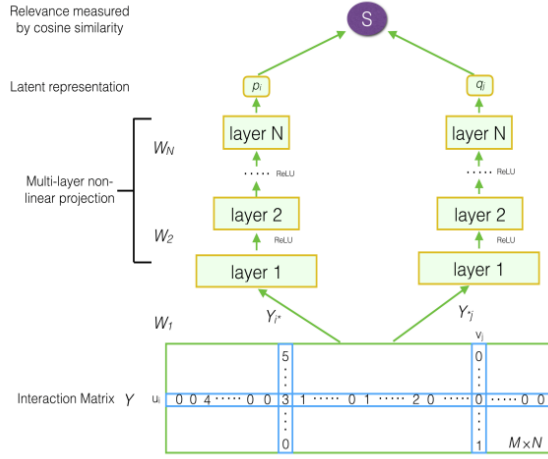
كما ذكرنا في القسم السابق قمنا بتشكيل مصفوفة Y اعتماداً على المعادلة 2. مع هذه المصفوفة ك دخل سنعرض بنية مكونة من شبكة عصبونية عميقة للمستخدمين و شبكة عميقة للمنتجات و التي سيشرحها لنا الشكل 1.

من المصفوفة Y كل مستخدم يتم تمثيله بشعاع و الذي هو سطر من المصفوفة حيث Y_{i*} يمثل المستخدم i بينما كي منتج يتم تمثيله بشعاع و الذي هو عمود في هذه المصفوفة حيث أن Y_{*j} يمثل المنتج j , نرسم لدخل ب x و للخروج ب y بينما نرسم للطبقات المتوسطة ب l_i حيث $i=1, \dots, N$ بينما أوزان الطبقة i ب W_i و مقدار الازاحة للطبقة i ب b_i و الخرج النهائي h . تصبح العلاقات في النهاية.

$$l_1 = W_1 \cdot x$$

$$l_i = f(W_i \cdot l_{i-1} + b_i), i = 2, 3, \dots, N - 1$$

$$h = f(W_N \cdot l_{N-1} + b_N) \quad (5)$$



الشكل 1

سنقوم باستخدام التابع $ReLU$ في جميع الطبقات أي أنه

$$f(x) = \max(0, x) \quad (6)$$

في بنيتنا لدينا شبكتان عصبونيتان متعددة الطبقات تأخذ إما سطر أو عمود كدخول، حتى تتمكن من تحويل المستخدم i أو المنتج j إلى الفضاء يمكننا الاستفادة من المعادلتين.

$$p_i = f_{\theta_N^U}(\dots f_{\theta_3^U}(W_{U2} f_{\theta_2^U}(Y_{i*} \cdot W_{U1})) \dots)$$

$$q_j = f_{\theta_N^V}(\dots f_{\theta_3^V}(W_{V2} f_{\theta_2^V}(Y_{*j}^T \cdot W_{V1})) \dots)$$

(7)

حيث تمثل W_U الأوزان لشبكة المستخدمين و W_V أوزان شبكة المنتجات بينما يمكننا حساب قيمة التقييم المتوقعة من المستهلك للمنتج كالتالي.

$$\begin{aligned} \hat{Y}_{i,j} &= F^{DMF}(u_i, v_j | \theta) = \cos(\widehat{p_i}, \widehat{q_j}) \\ &= \frac{p_i^T \cdot q_j}{\|p_i\| * \|q_j\|} \end{aligned} \quad (8)$$

في نموذجنا بجانب التعلم باستخدام الشبكات العصبونية نريد التأكيد مرة أخرى على أن المعرفة التي لدينا تتكون في المرة الأولى تكون من تقييمات المستخدمين فقط كما أننا يمكننا أن نعرف توجه المستخدم العام من نوعه

$$L = \sum_{y \in Y^+ \cup Y^-} \frac{Y_{ij}}{MaxR} \log(\hat{Y}_{ij}) + \frac{(1 - Y_{ij})}{MaxR} \log(1 - \hat{Y}_{ij}) \quad (11)$$

حيث MaxR هو أعلى تقييم لدينا و هو 5 في نظام ال 5 نجوم سنقوم بتسمية هذا التابع بـ Normalized Binary Cross Entropy اختصاراً NCE .

الآن سنقوم بعرض خوارزمية التدريب المتبعة و سنشرحها.

Input: Iter // Number of iterations

Neg_ratio //number if negative sampling ratio

R //original rating matix

Initialization:

Randomly initialize W_u & W_v

Set Y depending on equation 2

Set Y^+ as all non-zero ratings in Y

Set Y^- as all zero ratings in Y

Set $Y_{sampled}^-$ as sample interaction from Y^+ and its number is equal to $neg_ratio * ||Y^+||$

Set Train_set as $Y^+ \cup Y_{sampled}^-$

//start training

For It from 1 to Iter do

For each rating in T of user i and item j in T do:

Set p_i & q_j from equation 7

Set \hat{Y}_{ij}^o from equation 8 & 12

Set L from equation 11

Finally we use back propagation to optimize model parameters

End for

End for

الأفلام التي يشاهدها او المنتجات التي يستخدمها كما يمكننا أن نعرف تصنيف الفيلم و نوعيته. يمكننا القول بأننا باستخدام هذه التمثيلات ستكون مفيدة جدا في المستقبل.

تابع حساب الخطأ

من أهم المكونات التي لدينا في نموذجنا هو تعريف تابع يقوم بتحسين البنية الخاصة بنا بحيث يكون هذا التابع يعتمد على التقييمات المحجوزة و التغذية الرجعية كما ذكرنا سابقاً. يعطى الشكل العام لتابع التحسين كالتالي.

$$L = \sum_{y \in Y^+ \cup Y^-} l(y, \hat{y}) + \lambda \Omega(\theta) \quad (9)$$

حيث أن $l()$ هو تابع حساب الخطأ و $\Omega(\theta)$ هو المنظم.

من أجل أنظمة الاقتراحات يوجد لدينا بشكل عام نوعين من توابع التحسين و هما.

- Point-wise
- Pair-wise

للسهولة سوف نستخدم Point-wise كتابع تحسين و الذي يقوم بإهمال المنظم و يعتمد فقط على تابع حساب الخطأ.

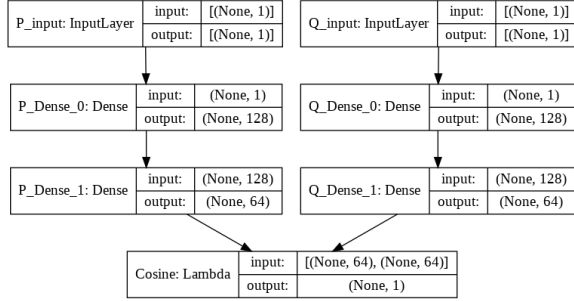
كما أنه يوجد العديد من توابع الخطأ لكننا سنستخدم على تابع Binary Cross Entropy لأنه الأفضل عند التعامل مع المسائل ذات التصنيف المتعدد و ذات تسميات متعددة. حيث أن حساب الخطأ التربيعة سوف يعطي أخطاء بقيم كبيرة و يكون قانون تابع الخطأ الذي سنستخدمه كالتالي.

$$L = \sum_{y \in Y^+ \cup Y^-} Y_{ij} \log(\hat{Y}_{ij}) + (1 - Y_{ij}) \log(1 - \hat{Y}_{ij}) \quad (10)$$

باختصار تابع حساب الخطأ التربيعة يعتمد فقط على التقييمات السابقة بينما تابع الخطأ الذي لدينا يأخذ بعين الاعتبار التغذية الرجعية. في هذا البحث سنقوم بتعديل التابع الذي لدينا حتى نستطيع الدمج بين التقييمات السابقة و خرج النموذج الذي لدينا, حتى نقوم بهذا الدمج سوف نقسم التقييمات السابقة على 5 و التي هي أعلى قيمة في تقييم الخمس نجوم و ذلك لأن خرج النموذج بين 0 و 1 و لكن التقييمات بين 1 إلى 5 فنقوم بالقسمة على 5 من أجل توحيد مجال القيم فيصبح التابع على الشكل التالي.

على الترتيب و سنستخدم تابع $ReLU$ في العصبونات كما ذكرنا سابقاً. و أيضاً سنستخدم $Adam$ خوارزمية للتحسين و معدل التعلم سيكون 0.001 و في النهاية سنقوم باستخدام تابع الخطأ الذي قمنا بتعريفه سابقاً.

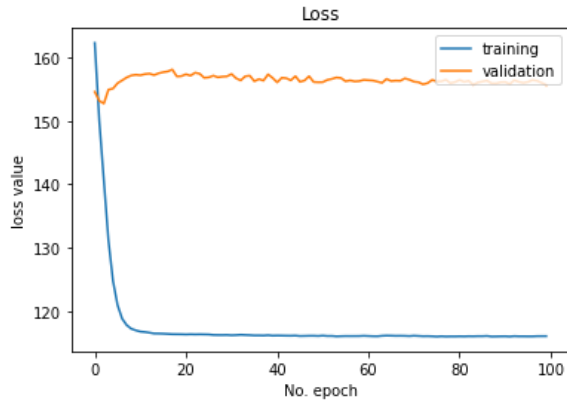
بنية النموذج.



عرض النتائج

بعد أن قمنا بتدريب النموذج الذي لدينا على 100 عصر و حجم كل دفعة 256 كانت النتائج كالتالي:

الخط البياني لتابع الخطأ.



بينما كانت قيمة $RMSE$ كالتالي.

RMSE on test data

```

In [97]: y_pred = dmf.model.predict({"P_input": X1_test, "Q_input": X2_test})
m = keras.metrics.RootMeanSquaredError()
m.update_state(y_test, y_pred)
print(f"RMSE = {m.result()}")

RMSE = 2.9300060272216797

```

من أجل تابع حساب الخطأ و لأنه من الممكن ان يكون الخرج سالب فإننا سنحتاج لاستخدام المعادلة 12 من أجل تقادي الحصول على القيمة 0 أو على قيم سالبة. سنضع μ قيمة صغير جدا أكبر من 0 و لتكن 10^{-6} في تجربتنا و تكون المعادلة على الشكل التالي.

$$\hat{Y}_{ij}^o = \max(\hat{Y}_{ij}, \mu) \quad (12)$$

التجريب

في هذا القسم سنقوم بتعريف البيانات التي لدينا و حيث سنقوم بشرح طريقة تمثيلها و بعد ذلك سنقوم بتجربتها على النموذج المطروح و رسم منحني التعلم و سنقارنها مع قيمة $RMSE$ لأحدث تقنية حيث أن $RMSE$ هي الجذر التربيعي لمتوسط مجموع مربعات الأخطاء و هي اختصار لـ $Root Mean Square Error$.

تعريف البيانات التي لدينا

سنستخدم في هذا البحث مجموعة البيانات التي تم استخدامها في جائزة نتفلكس. و هي موجودة على الرابط التالي.

<https://www.kaggle.com/netflix-inc/netflix-prize-data>

تكون البيانات موجودة على الشكل التالي.

"رقم الفيلم:

رقم المستخدم, التقييم, تاريخ التقييم"

يمكن أن يكون تحت رقم الفيلم عدد من المستخدمين و ليس مستخدم واحد.

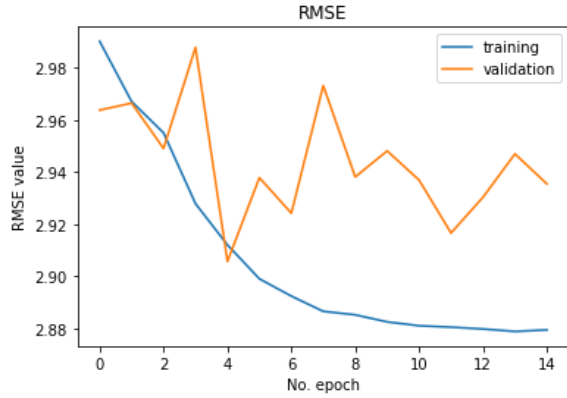
معالجة البيانات

سنقوم بمعالجة البيانات عن طريق وضع مصفوفة ثنائية حيث تعبر أدلة هذه المصفوفة عن رقم المستخدم (رقم السطر) و رقم الفيلم (رقم العمود) و ستكون قيمة الخانة هي تقييم المستخدم لهذا المنتج و ستكون 0 في حالة عدم وجود تقييم كم ذكرنا في القسم الثالث.

تدريب النموذج

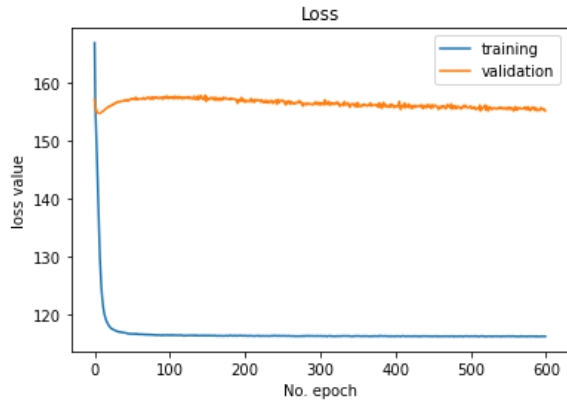
بعد أن قمنا بتحويل البيانات إلى مصفوفة ثنائية سنقوم بتدريب النموذج حيث سيكون عدد طبقات كل شبكة 3 , كما سيكون عدد العصبونات في كل شبكة $[1, 128, 64]$

قيمة RMSE

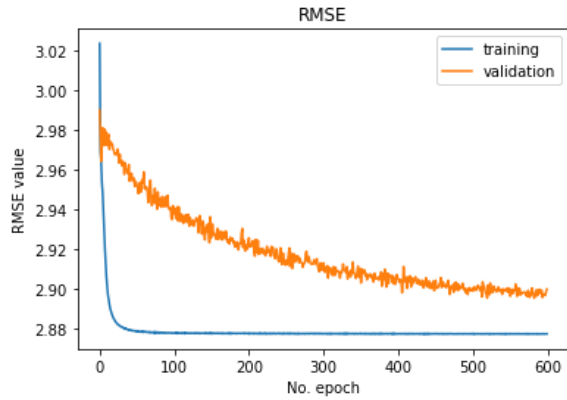


نلاحظ أن النتائج أصبحت أسوأ عند استخدام عدد عصور أقل و معدل تعلم أعلى.

بينما عندما حافظنا على القيم السابقة لتصبح معدل التعلم 0.0001 و عدد العصور 600 حصلنا على النتائج التالية. تابع الخطأ.

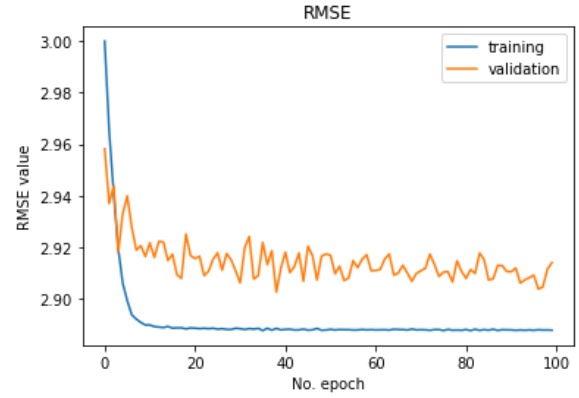


قيمة RMSE



نلاحظ تحسن النتائج قليلاً عند استخدام المعطيات السابقة.

بينما الخط البياني لـ RMSE.



ملاحظة: لم نتمكن من تدريب النموذج على جميع البيانات لأنه يحتاج إلى كميات كبيرة من الذاكرة 1 تيرابايت تقريباً لذلك النتائج المعروضة هي على قسم من البيانات.

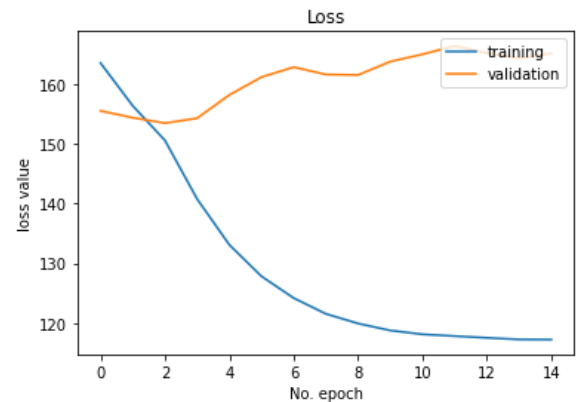
```
print(X2)
print(X2.shape)
print(np.max(y))
print(y.shape)

MemoryError                                Traceback (most recent call last)
<ipython-input-79-8eb86d180381> in <module>
      6 y.append(r)
      7
----> 8 X1 = np.array(X1)
      9 X2 = np.array(X2)
     10 y = np.array(y)

MemoryError: Unable to allocate 1.62 TiB for an array with shape (100480587, 17770)
```

سنقوم بمقارنة النتيجة السابقة مع نفس النموذج لكن عندما نستخدم معدل تعلم أكبر و هو 0.01 و عدد عصور أقل و هو 10 مع الحفاظ على باقي القيم فتكون النتائج كالتالي.

تابع الخطأ.



Loss Function

ملخص

في هذا البحث قمنا بتعريف نموذج يعتمد على الشبكات العصبونية. عن طريق الشبكات العصبونية و اسقاط المستخدمين و المنتجات في فضاء صغير الأبعاد. في نموذجنا قمنا باستخدام البيانات السابقة الخاصة بالمستخدمين و التغذية الرجعية حيث كان الدخل هو التقييم للمنتجات دون وجود تفضيل مبدئياً للتغذية الرجعية. من جهة أخرى قمنا بتصميم تابع خطأ من أجل تدريب النموذج بحيث يأخذ كل من التقييمات و التغذية الرجعية. و رأينا أخيراً النتائج التي حصلنا عليها و قمنا بمقارنتها.

المراجع

[Sarwar et al., 2001] Badrul Sarwar, George Karypis, et al. Item-based collaborative filtering recommendation algorithms. In WWW, 2001.

[Ma et al., 2008] Hao Ma, Haixuan Yang, Michael R Lyu, and Irwin King. Sorec: Social recommendation using probabilistic matrix factorization. In CIKM, 2008.

[Hu et al., 2008] Yifan Hu, Yehuda Koren, and Chris Volinsky. Collaborative filtering for implicit feedback datasets. In Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on, pages 263–272. Ieee, 2008.

[Huang et al., 2013] Po-Sen Huang, Xiaodong He, Jianfeng Gao, et al. Learning deep structured semantic models for web search using clickthrough data. In Proceedings of the 22nd ACM international conference on Conference on information & knowledge management, pages 2333–2338. ACM, 2013.

[Elkahky et al., 2015] Ali Mamdouh Elkahky, Yang Song, and Xiaodong He. A multi-view deep learning approach for cross domain user modeling in recommendation systems. In Proceedings of the 24th International Conference on World Wide Web, pages 278–288. ACM, 2015.