إشتقاق التقييمات الضمنية في تقنية التصفية التعاونية من معاملات الشراء و سجل السلوك

Derivation of implicit ratings in Collaborative Filtering from purchase transactions and behavior history

مروة فاروق عبدالمتعال سليمان
Marwa_2t@hotmail.com
كلية علوم الحاسوب و تقانة المعلومات - جامعة النيلين – الخرطوم – السودان

المستخلص :

تعتبر تقنية التصفية التعاونية من أشهر التقنيات المستخدمة في نظم التوصية فهي تعتمد علي تشارك الإهتمامات بين المستخدمين لتقديم توصيات ذات علاقة بالمستخدم، و تستخدم لذلك التقييمات التي يقدمها المستخدم بصورة صريحة إلا أن هذه التقييمات قد تكون مجهدة للمستخدم أو قد لا يتحري الصحة عند تقديمها مما ينتج عنه توصيات تفتقر إلي الدقة. تستهدف هذا الورقة حل مشكلة قلة تقييمات المستخدم في مصفوفة تقييمات التصفية التعاونية، تم إقتراح طريقة لإستغلال بيانات معاملات الشراء و بيانات تدفق نقر المستخدم لإشتقاق تقييمات ضمنية للمستخدم بحيث تمثل هذه التقييمات درجة تفضيل المستخدم للعنصر. يتم إضافة هذه التقييمات الضمنية إلي مصفوفة التقييمات بغرض إثراء المصفوفة و من ثم تجويد التوصيات. تم إجراء الإختبارات علي بيانات متجر تجاري إلكتروني للبيع بالتجزئة. وفقاً للبيانات المستخدمة أثبتت التجارب أن التقييمات الضمنية التي تم إشتقاقها من بيانات شراء المستخدم و من بيانات تدفق نقر المستخدم تقلل من التأثير السلبي لقلة أو تناثر تقييمات المصفوفة، كما أنها تحسّن من دقة التوصيات المقدمة للمستخدم. مستقبلياً يمكن التنقيب عن بيانات أخري لتحسين دقة التوصيات مثل الزمن المستغرق في مشاهدة عنصر محدد أو صنف محدد، كما يمكن وضع السياق الزمني للشراء في الإعتبار لتقديم توصيات ترتبط بفترة الشراء.

Abstract:
Collaborative filtering is one of the most popular techniques used in recommender systems. It relies on sharing interests among users to provide recommendations related to the user, and it uses the ratings that the user provides explicitly, but these ratings may be stressful for the user or may not investigate the correctness when presented, which results in inaccurate recommendations. This paper aims to solve the problem of lack of user ratings in the collaborative filter rating matrix; a proposed method exploits user transaction data and clickstream data to derive implicit user ratings, which represent the degree of user preference for the item. These implicit ratings are added to the rating matrix to enrich the matrix and then improve the recommendations. The experiments were conducted on the data of a retail e-commerce store, according to the dataset, experiments have shown that implicit ratings that are derived from user purchase data and clickstream data reduce the negative impact of data sparsity, and improve the accuracy of user recommendations. In the future, other data can be mined to improve the accuracy of recommendations, such as the viewing time of a specific item or category, also the purchase time context can be taken into account to provide recommendations related to the purchase period.

الكلمات المفتاحية Keywords:
نظم التوصية (Recommender Systems) ، التصفية التعاونية (Collaborative Filtering)، التقييمات الضمنية (Implicit ratings)، تناثر البيانات (Data sparsity).

1- المقدمة:

نظراً لغمر المستخدم بمجموعة من المنتجات والخيارات في بيئة التسوق عبر الإنترنت، يحتاج المستخدم إلي وسيلة تدعمه و توجهه في عملية البحث عن المنتجات و إختيار ما يلائمه من العناصر، تؤدي نظم التوصية (Recommender Systems) مهمة تصفية معلومات المستخدم بإستخدام العديد من التقنيات منها تقنية التصفية التعاونية (Collaborative Filtering) والتي تقوم علي مبدأ إذا كان المستخدمون قد تشاركوا نفس الاهتمامات في الماضي فسيكون لهم أيضًا تفضيلات متشابهة في المستقبل، إذ تعتمد هذه التقنية على التقييمات (ratings) التي يقدمها المستخدم و التي بموجبها يتم تقديم توصيات ذات علاقة بالمستخدم و ترضي تفضيلاته، إلا أن هذه التقييمات يندر تقديمها من المستخدم بصورة صريحة إذ أنها مجهدة للمستخدم أو قد لا يعبأ المستخدم بتحري الصحة في تقديمها مما ينتج عنه توصيات قد تكون غير ذات مغزي للمستخدم. تهدف هذه الورقة إلي حل مشكلة قلة

تقييمات المستخدم في مصفوفة تقييمات التصفية التعاونية، بحيث يتم إستغلال مصادر المعلومات المتاحة كمعاملات شراء المستخدم و سجل سلوك المستخدم لإستخراج معرفة عن تفضيلات المستخدم و من ثم توظيف هذه المعرفة لتقديم توصيات مخصصة للمستخدم. تم تنظيم الورقة على النحو التالي: في القسم 2 سنتناول الدراسات السابقة، و في القسم 3 سنوضح نموذج التوصية المقترح، يليه القسم 4 الذي يشرح خوارزمية التوصية المقترحة مع مثال توضيحي لها في القسم 5، أما في القسم 6 سنتناول التجربة و التقييم، يليه القسم 7 و الذي يوضح النتائج التي تم التوصل إليها، و نختتم بالخلاصة في القسم 7.

2- الدراسات السابقة:

تعمل تقنية التصفية التعاونية علي إستغلال تقييمات المستخدم للتنبؤ بما سيفضله المستخدم، وفقاً لتصنيف Ricci في [1] فإنه يمكن إستخدام تقييمات المستخدم لبناء و تعلم نموذج تنبؤي و هو ما يعرف بطرق التصفية بناءً علي النموذج (Model-Based methods)، و التي يتم فيها بناء النموذج بإستخدام خوارزميات التعلم الآلي كما في [2] أو قواعد الإرتباط و Naïve Bayes Classifier كما في [3] أو الشبكات العصبية كما في [4]. كذلك يمكن إستخدام تقييمات المستخدم للتنبؤ استناداً إلى التشابه بين المستخدمين أو التشابه بين العناصر وهو ما يعرف بطرق التصفية بناءً علي الذاكرة (Memory-Based methods) ، هذه الطرق تعتمد علي تقييمات المستخدم و التي قلّما يقدمها المستخدم بصورة صريحة، لذا لجأت بعض الدراسات إلي إشتقاق تقييمات ضمنية لتحديد تفضيلات المستخدم بغرض تجويد التوصيات كما في [5] حيث تم تحليل الأنماط المتسلسلة (sequential pattern analysis) للمستخدم مع بيانات شراء المستخدم تقديم توصيات تتعلق بالمستخدم، و في [6] تم إستخدام بيانات شراء المستخدم و بيانات تدفق نقر المستخدم (clickstream) المتسلسل لتحديد تفضيلات المستخدم، بينما تم الإعتماد علي التشابه بين نمط سلوك المستخدمين في الشراء و المقارنة و التشابه بين العناصر من خلال تضمين الكلمات (Word-Embedding) لتحديد التوصيات ذات العلاقة في [7]، و الإعتماد علي تشابه سمات العنصر لتحديد تفضيلات المستخدم مما شابهه من السمات [8]. كذلك إستفادت بعض الدراسات من بيانات المحتوي لتدعيم القصور في تقنية التصفية التعاونية من خلال إستخدام التقنية المختلطة كما في [9] و [10]. مقارنة بالطرق المذكورة أعلاه، في هذه الورقة تم إشتقاق تقييمات ضمنية للمستخدم بالتنقيب في بيانات معاملات المستخدم

بالإضافة إلي بيانات تدفق نقر المستخدم التشاركي بين المستخدمين، بحيث تعكس هذه التقييمات الضمنية درجة تفضيل المستخدم للعنصر، و التي يتم الإستناد إليها لتخصيص توصيات للمستخدم.

3- نموذج التوصية المقترح:

يعمل النموذج المقترح علي حل مشكلة قلة بيانات التقييم في مصفوفة التقييمات في تقنية التصفية التعاونية، تم إقتراح طريقة لإشتقاق تقييمات ضمنية من سجل سلوك المستخدم بحيث تمثل هذه التقييمات درجة تفضيل المستخدم للعنصر حسب تفاعلاته، تعمل هذه التقييمات الضمنية و بخلاف التقييمات الصريحة علي إثراء مصفوفة التقييمات بتقييمات يسهل الحصول عليها، كما أنها متاحة بصورة كافية من كل المستخدمين. بإضافة تقييمات المستخدم الضمنية إلي مصفوفة تقييمات التصفية التعاونية تصبح المصفوفة غنية بالتقييمات مما ينعكس إيجاباً علي دقة التوصيات المقدمة للمستخدم.

يهدف النموذج المقترح إلي إستغلال بيانات معاملات شراء المستخدم و سجل سلوك المستخدم لإشتقاق تقييمات للمستخدم بصورة ضمنية، تعمل هذه التقييمات الضمنية علي إثراء مصفوفة التقييمات التعاونية بتقييمات تعبر عن درجة تفضيل المستخدم للعناصر التي تفاعل معها أو التي قام بشرائها مسبقاً، و ذلك بغرض مساعدة و توجيه المستخدم عن طريق تقديم توصيات بعناصر تقع ضمن إهتمامات المستخدم.

4- خوارزمية التوصية المقترحة:

تستقبل الخوارزمية جدول الشراء و النقر و مصفوفة الشراء كمدخلات، يتم إستخدام مصفوفة الشراء لتوليد التقييمات الضمنية للمستخدم بناءً علي بيانات معاملات الشراء، بينما يتم إستخدام بيانات جدول الشراء و النقر لتوليد التقييمات الضمنية بناءً علي تدفق نقر المستخدم، و من ثم يتم إمداد مصفوفة الشراء بالتقييمات الضمنية المشتقة و ذلك للتنبؤ بتقييمات المستخدم للعناصر التي لم يقييمها و التي تعتبر مخرج للخوارزمية. فيما يلي سنوضح بالتفصيل مدخلات و مخرجات هذه الخوارزمية و خطوات تنفيذها.

4-1 المدخلات:

1) جدول الشراء:

يحتوي جدول الشراء علي بيانات معاملات الشراء و التي توضح العناصر المشتراه لكل مستخدم في كل جلسة كما يوضح الجدول التالي.

جدول–1: جدول الشراء

session_id	user_id	item_id	Time	quantity
1	1	b	2014-04-03T10:26:07.632Z	1
1	1	a	2014-04-03T10:26:07.753Z	1
2	1	a	2014-04-07T09:22:28.132Z	1
3	2	b	2014-04-07T09:22:28.176Z	2
3	2	c	2014-04-03T11:04:11.417Z	1
4	3	c	2014-04-02T10:42:17.227Z	1
5	3	d	2014-04-02T13:17:46.940Z	1

2) جدول النقر:

يحتوي جدول النقر علي بيانات تدفق نقر المستخدم والتي توضح أحداث النقر علي العناصر خلال كل جلسة للمستخدم كما يوضح الجدول التالي.

جدول–2: جدول النقر

session_id	user_id	item_id	Category	Time
1	1	a	2	2014-04-07T10:51:09.277Z
1	1	b	3	2014-04-07T10:54:09.868Z
2	1	c	1	2014-04-07T10:54:46.998Z
2	1	d	4	2014-04-07T10:57:00.306Z
3	2	c	1	2014-04-07T13:56:37.614Z
3	2	a	2	2014-04-07T13:57:19.373Z
4	2	d	4	2014-04-07T13:58:37.446Z
5	3	d	4	2014-04-02T13:17:46.940Z
6	3	b	3	2014-04-02T13:26:02.515Z

3) جدول الشراء و النقر:

يحتوي جدول الشراء و النقر علي بيانات الجلسة كاملة فهو تكامل لأحداث النقر و الشراء لكل مستخدم وفقاً للتسلسل الزمني كما يوضح الجدول التالي.

جدول–3: جدول الشراء و النقر

session_id	user_id	clicks	purchases

1	1	a,b,c,a,c	b,a,a
2	1	c,d,a	
3	2	b,c,d,b	b,b,c
4	2	b,c,e,e	c,c
5	3	d,b,c	c,d
6	3	b,e,e	
7	4	c,d,a,a	

4) مصفوفة الشراء:

وهي عبارة عن مصفوفة تقييمات المستخدم التي تعتمد علي بيانات الشراء، تمثل الصفوف المستخدمين، بينما تمثل الأعمدة العناصر المشتراه من قبل المستخدمين، يتم تمثل تقييمات المستخدم في هذه المصفوفة بالمتغير المتضمن داخل بيانات معاملة الشراء وهو متغير الكمية المشتراه من العنصر المحدد خلال جلسات المستخدم المختلفة كما يوضح الجدول التالي، في الجدول يشار لتقييم العناصر غير المشتراه و التي ليست لها قيمة لكمية المشتروات بالرمز (؟).

جدول-4 مصفوفة الشراء

	a	b	c	d
1	2	1	؟	؟
2	؟	2	3	؟
3	؟	؟	1	1

4-2 المخرجات:

تتبع الخوارزمية عدة خطوات يتم تنفيذها علي المدخلات المحددة لتحديد تقييمات المستخدم الضمنية حسب تفاعلات المستخدم، ومن ثم يتم إستخدام هذه التقييمات المشتقة للتنبؤ بدرجة تفضيل المستخدمين للعناصر غير معروفة التقييم، تنتج الخوارزمية مصفوفة بالتقييمات المتوقعة للعناصر التي لم يقييمها المستخدمون كمخرج.

4-3 خطوات تنفيذ خوارزمية التوصية المقترحة:

المدخلات: جدول الشراء، جدول النقر، جدول الشراء و النقر، مصفوفة الشراء.

المخرجات: مصفوفة التقييمات المتوقعة للعناصر التي لم يقييمها المستخدمون بناءً علي التقييمات الضمنية للمستخدمين.

1: تطبيع تقييمات مصفوفة الشراء في مصفوفة جديدة هي مصفوفة التقييمات الضمنية، ينتج (عنصر، التقييم الضمني للشراء).

2: **لكل** جلسة نقر تنتمي إلي جدول الشراء و النقر:

3: أحسب معدل النقر لكل عنصر، ينتج (مستخدم، عنصر، معدل النقر)

4: **لكل** جلسة شراء تنتمي إلي جدول الشراء قارن مع جلسة النقر التي تنتمي إلي جدول النقر:

5: **إذا كان** العنصر مشترك بين جلسة النقر و جلسة الشراء:

6: أحسب معدل الإشتراك للعناصر المفردة المشتركة بين جلسة النقر و جلسة الشراء، ينتج (عنصر، معدل الإشتراك)

7: **وإلا:**

8: أحسب معدل عدم الإشتراك للعناصر المفردة غير المشتركة بين جلسة النقر و جلسة الشراء، ينتج (عنصر، معدل عدم الإشتراك)

9: **نهاية الشرط.**

10: **نهاية تكرار** جلسة الشراء.

11: التقييم الضمني للنقر= **معدل النقر** + **معدل الإشتراك** + **معدل عدم الإشتراك**، ينتج (مستخدم، عنصر، تقييم ضمني للنقر)

12: **نهاية تكرار** جلسات النقر.

13: **لكل** تقييم للعنصر ينتمي **إلي التقييم الضمني للنقر** إختبر:

14: **إذا لم** يوجد تقييم مسبق من المستخدم للعنصر في **مصفوفة التقييمات الضمنية**، (عنصر، تقييم ضمني للشراء):

15: أضف **التقييم الضمني للنقر** في **مصفوفة التقييمات الضمنية**، (مستخدم، عنصر، تقييم ضمني للنقر).

16: **نهاية الشرط.**

17: **نهاية تكرار** التقييم الضمني للنقر.

18: طبّق خوارزمية التصفية التعاونية علي **مصفوفة التقييمات الضمنية**.

19: **مصفوفة التقييمات المتوقعة للعناصر التي لم يقييمها المستخدمون** هي المصفوفة الناتجة عن التصفية التعاونية وفقاً للتقييمات الضمنية.

20: أرجع: **مصفوفة التقييمات المتوقعة للعناصر التي لم يقييمها المستخدمون بناءً علي التقييمات الضمنية للمستخدمين.**

الشكل–1: خطوات تنفيذ خوارزمية التوصية المقترحة

الخطوات التالية تشرح الخوارزمية الموضحة في الشكل–1:

(1) تستقبل الخوارزمية مصفوفة الشراء كمدخل و التي يتم فيها تمثيل الكمية المشتراه من كل عنصر خلال جلسات المستخدم كتقييم للمستخدم، يتم تطبيع هذه التقييمات بتحويلها لتقييمات في شكل مقياس رقمي لقيمة بين 0 و1 بإستخدام معادلة متجه الوحدة Unit Vector المعادلة–1، تمثل هذه القيمة درجة تفضيل المستخدم للعنصر من ضمن العناصر المشتراه. تحتوي مصفوفة التقييمات الضمنية وهي المصفوفة الجديدة الناتجة عن عملية التطبيع علي التقييمات الضمنية للشراء المشتقة من كمية المشتروات (سطر رقم 1).

$$\hat{u} = \frac{u}{|u|}$$

معادلة–1: معادلة متجة الوحدة

$\hat{u}$ تشير إلي متجه الوحدة، u تشير إلي المتجهه، $|u|$ تشير إلي حجم المتجه.

(2) لكل جلسة نقر للمستخدم يتم حساب معدل النقر لكل عنصر من عناصر جلسة النقر (سطر رقم 3) وفقاً للمعادلة–2، تم إستبعاد العناصر ذات معدلات النقر التي تقل عن الحد الأدني لمعدل النقر (0.3):

معدل النقر = عدد مرات تكرار العنصر في الجلسة/ عدد عناصر جلسة النقر

معادلة–2: معادلة معدل النقر

(3) لحساب معدل الإشتراك (سطر رقم 6) يتم مقارنة جلسة نقر المستخدم مع جلسات الشراء لتحديد العناصر المشتركة بين جلسة النقر و بين جلسات الشراء، حيث يتم حساب معدل الإشتراك للعنصر المشترك بين جلسة النقر و بين أي جلسة شراء وفقاً للمعادلة–3، ثم يتم إيجاد مجموع معدلات الإشتراك للعنصر مع جميع جلسات الشراء، و من ثم يتم تطبيع معدلات إشتراك العناصر المشتركة بتحويل مجموع معدل إشتراك كل عنصر مشترك إلي قيمة بين 0 و1 بإستخدام معادلة min-max scaler وفق الصيغة الموضحة في المعادلة–4:

معدل إشتراك العنصر = عدد العناصر المفردة المشتركة في جلسة النقر/عدد العناصر المفردة في جلسة النقر و جلسة الشراء

معادلة–3: : معادلة معدل الإشتراك

$$\acute{x} = \frac{x - min}{max - min}$$

معادلة–4: معادلة min-max scaler

$\acute{x}$ تشير إلي القيمة بعد التطبيع، x تشير إلي القيمة الفعلية،

min تشير إلي أقل قيمة، max تشير إلي أعلي قيمة.

(4) لحساب معدل عدم الإشتراك (سطر رقم 8) يتم مقارنة جلسة نقر المستخدم مع جلسات الشراء لتحديد العناصر غير المشتركة بين جلسة النقر و بين جلسات الشراء، حيث يتم حساب معدل عدم الإشتراك للعنصر غير المشترك بين جلسة النقر و بين أي جلسة شراء وفقاً للمعادلة (5–13)، ثم يتم إيجاد مجموع معدلات عدم الإشتراك للعنصر مع جميع جلسات الشراء، و من ثم يتم تطبيع معدلات عدم إشتراك العناصر غير المشتركة بتحويل مجموع معدل عدم إشتراك كل عنصر غير مشترك إلي قيمة بين 0 و1 بإستخدام المعادلة–4:

معدل عدم إشتراك العنصر = عدد العناصر المفردة غير المشتركة في جلسة النقر/عدد العناصر المفردة في جلسة النقر و جلسة الشراء

معادلة–5: معادلة معدل عدم الإشتراك

(5) لإشتقاق التقييمات الضمنية المشتقة من بيانات تدفق نقر المستخدم يتم جمع معدل النقر و معدل الإشتراك و معدل عدم الإشتراك لكل عنصر (سطر رقم 11) وفقاً للمعادلة–6، بحيث يمثل المجموع الكلي للمعدلات الثلاث هو تقييم المستخدم الضمني لجلسة النقر أو التقييم الضمني للنقر، تم جمع معدل النقر و معدل الإشتراك و معدل عدم الإشتراك بإستخدام الأوزان التالية (0.2، 0.5، 0.3) علي التوالي، تم تجربة الأوزان (0.5، 0.3، 0.2) و (0.3، 0.35، 0.35) إلا أن الأوزان المستخدمة (0.2، 0.5، 0.3) أعطت نتائج أفضل. تم إستبعاد العناصر ذات التقييمات الضمنية للنقر التي تقل عن الحد الأدني للتقييم الضمني للنقر (0.3).

التقييم الضمني للنقر = α * معدل النقر + β * معدل الإشتراك + γ * معدل عدم الإشتراك

حيث $0 < \gamma$ ،β ،$\alpha < 1$

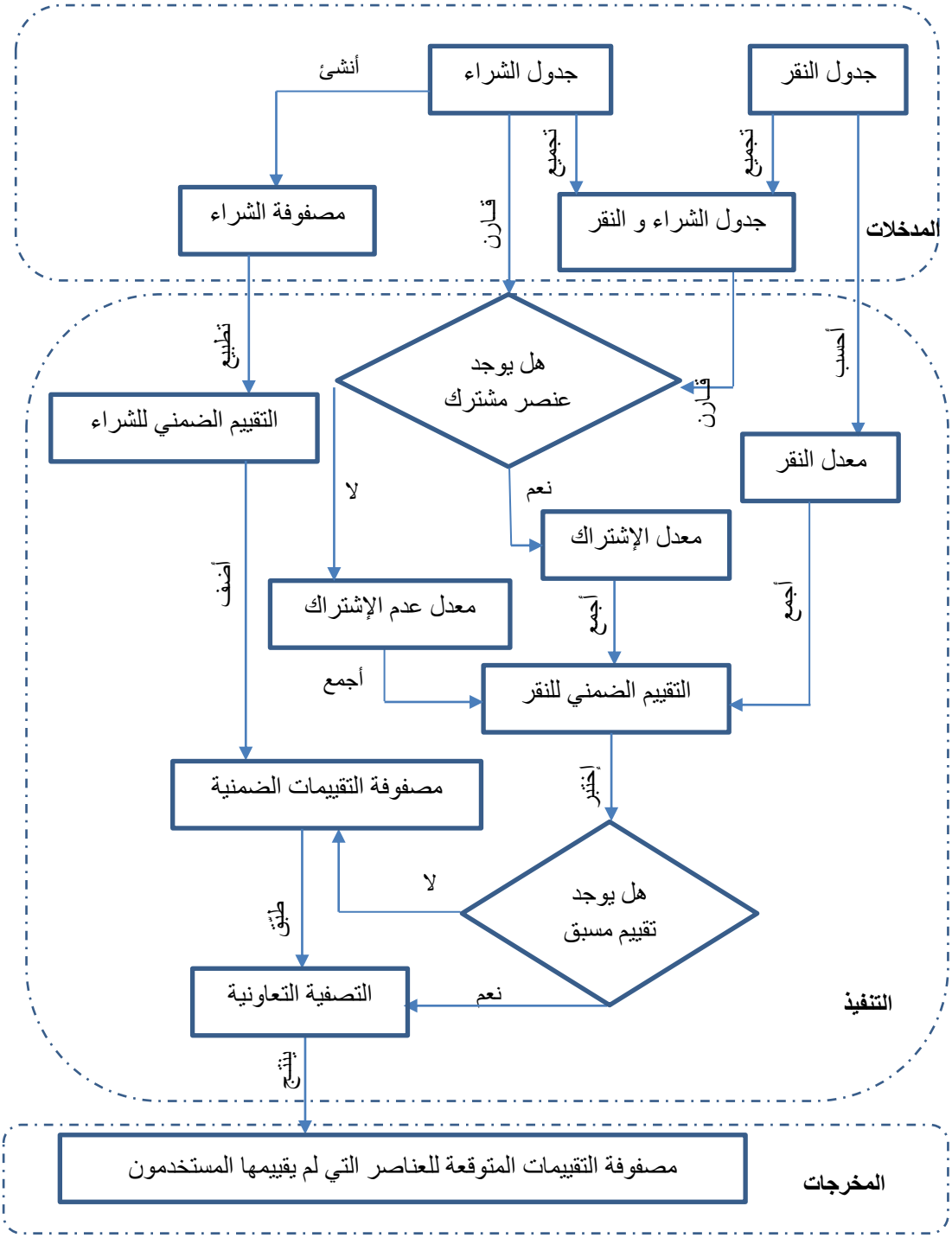
معادلة–6: معادلة التقييم الضمني للنقر

(6) أضف التقييمات الضمنية للنقر إلي مصفوفة التقييمات الضمنية في سطر رقم 1، بحيث يضاف التقييم الضمني للعناصر التي لم يسبق للمستخدم تقييمها، أو بمعني آخر في حال لم يشتري المستخدم تلك العناصر من قبل (سطر رقم 14، 15).

(7) طّبق خوارزمية التصفية التعاونية علي مصفوفة التقييمات الضمنية الناتجة عن تكامل تقييمات المستخدم الضمنية للشراء و تقييمات المستخدم الضمنية للنقر للتنبؤ بالتقييمات غير المعروفة للمستخدمين (سطر رقم 18). في مصفوفة تقييمات المستخدم المدخلة إلي خوارزمية التصفية التعاونية تم إستبدال كل القيم غير الصفرية بالقيمة 1 لأغراض تجريبية.

(8) تنتج مصفوفة التقييمات المتوقعة للعناصر التي لم يقييمها المستخدمون بناءً علي التقييمات الضمنية للمستخدمين كمخرج للخوارزمية (سطر رقم 20).

الشكل (5–3) التالي يوضح مخطط خوارزمية التوصية المقترحة و الذي يشتمل علي مدخلات و مخرجات و خطوات تنفيذ الخوارزمية المقترحة.



شكل–2: مخطط خوارزمية التوصية المقترحة

5– مثال توضيحي لخوارزمية التوصية المقترحة:

سنوضح خوارزمية التوصية المقترحة لإشتقاق تقييمات ضمنية للمستخدمين بإستخدام جدول الشراء (الجدول–1) و جدول النقر (الجدول–2) و جدول الشراء و النقر (الجدول–3) و مصفوفة الشراء (الجدول–4) في القسم 4–1.

- لتطبيع مصفوفة الشراء (الجدول–5) يتم إستخدام المعادلة–1، مثلاً لحساب التقييم الضمني للشراء للمستخدم (1) يتم بحساب متجة الشراء للعنصرين (a, b) كالتالي:

$$\hat{a} = \frac{2}{\sqrt{2^2 + 1^2}}, \qquad \hat{b} = \frac{1}{\sqrt{2^2 + 1^2}}$$

الجدول–5 التالي يوضح التقييمات الضمنية للشراء، نجد أن نسبة تناثر التقييمات في مصفوفة التقييمات الضمنية تمثل 50%. في الجدول يشير الرمز (؟) للتقييمات المفقودة للعناصر التي ليس لها تقييمات ضمنية للشراء.

جدول–5: التقييمات الضمنية للشراء

User/item	a	b	c	d
1	0.894	0.447	؟	؟
2	؟	0.554	0.832	؟
3	؟	؟	0.707	0.707

- من الشراء و النقر (الجدول–3) نجد أن هنالك جلسات نقر لا تنتهي بالشراء تخص المستخدمين (1، 3، 4)، لإشتقاق التقييمات الضمنية لهذه الجلسات سنقارن هذه الجلسات مع كل جلسات الشراء لحساب معدل النقر و معدل الإشتراك و معدل عدم الإشتراك لكل جلسة، فيما يلي سنحسب المعدلات الثلاث التي تخص كل مستخدم علي حدا، و من ثم نحدد العناصر ذات التقييم الضمني للنقر لكل مستخدم ليتم إضافتها لاحقاً إلي مصفوفة التقييمات:
- الجدول–5 و الجدول–6 و الجدول–7 توضح علي التوالي معدلات النقر و معدلات الإشتراك و معدلات عدم الإشتراك للمستخدم رقم (1) من جلسة النقر رقم (2) و جلسات الشراء (1، 3، 4، 5)، سنشير لمعدل العنصر غير المضمن ضمن عناصر الجلسة بالرمز (–)، حسابياً سنحدد المعدل لهذه العناصر بالقيمة 0.

جدول–5: معدل النقر للمستخدم رقم (1)

User/ item	a	b	c	d
1	0.33	-	0.33	0.33

جدول–6: معدل الإشتراك للمستخدم رقم (1)

Session_id/ item	a	b	c	d
1	1/4	0/4	0/4	0/4
3	0/4	0/4	1/4	0/4
4	0/3	0/4	1/3	0/3
5	0/3	0/4	2/3	2/3
sum	0.25	0.0	1.25	0.6
Min-max scaler	0.2	0.0	1.0	0.48

- من الجدول–6 نجد أن معدل إشتراك العناصر لا يتأثر بتكرار شراء العنصر كما في حالة العنصر(a) في جلسة الشراء رقم (1)، إذ أن تكرار العنصر تم أخذه في الإعتبار من خلال تضمين الكمية المشتراه من كل عنصر في التقييم الضمني للشراء.

جدول–7: معدل عدم الإشتراك للمستخدم رقم (1)

Session_id/ item	a	b	c	d
1	0/4	1/4	2/4	2/4
3	1/4	1/4	0/4	1/4
4	1/3	-	0/3	1/3
5	1/3	-	0/3	0/3
sum	0.917	0.5	0.5	1.08
Min-max scaler	0.719	0.0	0.0	1.0

لحساب التقييمات الضمنية للمستخدم (1) نجمع المعدلات السابقة بإستخدام الأوزان المحددة لكل معدل كما يلي:

a = (0.33*0.2) + (0.2*0.5) + (0.719*0.3) = 0.381

b = (– *0.2) + (0.0*0.5) + (0.0*0.3) = 0.0

c = (0.33*0.2) + (1.0*0.5) + (0.0*0.3) = 0.566

d = (0.33*0.2) + (0.48*0.5) + (1.0*0.3) = 0.606

- الجدول–8 و الجدول–9 و الجدول–10 التالية توضح علي التوالي معدلات النقر و معدلات الإشتراك و معدلات عدم الإشتراك للمستخدم رقم (3) من جلسة النقر رقم (6) و جلسات

الشراء (1، 3، 4، 5)، سنشير لمعدل العنصر غير المضمن ضمن عناصر الجلسة بالرمز (-)، حسابياً سنحدد المعدل لهذه العناصر بالقيمة 0.

جدول-8: معدل النقر للمستخدم رقم (3)

User/ item	a	b	c	d	e
3	-	0.33	-	-	0.67

جدول-9: معدل الإشتراك للمستخدم رقم (3)

Session_id/ item	a	b	c	d	e
1	0/3	1/3	-	-	0/3
3	-	1/3	0/3	-	0/3
4	-	0/3	0/3	-	0/3
5	-	0/4	0/4	0/4	0/4
sum	0.0	0.6	0.0	0.0	0.0
Min-max scaler	0.0	1	0.0	0.0	0.0

- من الجدول-9 نجد أن معدل إشتراك العناصر لا يتأثر بتكرار النقر علي العنصر كما في حالة العنصر(e) في جلسة النقر رقم (6)، إذ أن تكرار النقر علي العنصر تم أخذه في الإعتبار من خلال حساب معدل النقر لكل عنصر في التقييم الضمني للنقر.

جدول-10: معدل عدم الإشتراك للمستخدم رقم (3)

Session_id/ item	a	b	c	d	e
1	1/3	0/3	-	-	1/3
3	-	0/3	1/3	-	1/3
4	-	1/3	1/3	-	1/3
5	-	1/4	1/4	1/4	1/4
sum	0.33	0.583	0.917	0.25	0.917
Min-max scaler	0.119	0.449	1.0	0.0	1.0

لحساب التقييمات الضمنية للمستخدم (3) نجمع المعدلات السابقة بإستخدام الأوزان المحددة لكل معدل كما يلي:

a = (- *0.2) + (0.0*0.5) + (0.119*0.3) = 0.036

b = (0.33*0.2) + (1.0*0.5) + (0.449*0.3) = 0.716

c = (- *0.2) + (0.0*0.5) + (1.0*0.3) = 0.3

d = (- *0.2) + (0.0*0.5) + (0.0*0.3) = 0.0

e = (0.67*0.2) + (0.0*0.5) + (1.0*0.3) = 0.434

- الجدول–11 و الجدول–12 و الجدول–13 التالية توضح علي التوالي معدلات النقر و معدلات الإشتراك و معدلات عدم الإشتراك للمستخدم رقم (4) من جلسة النقر رقم (7) و جلسات الشراء (1، 3، 4، 5)، سنشير لمعدل العنصر غير المضمن ضمن عناصر الجلسة بالرمز (–)، حسابياً سنحدد المعدل لهذا العنصر بالقيمة 0.

جدول–1: معدل النقر للمستخدم رقم (4)

User/ item	a	b	c	d
3	0.5	–	0.25	0.25

جدول–12: معدل الإشتراك للمستخدم رقم (4)

Session_id/ item	a	b	c	d
1	1/4	0/4	0/4	0/4
3	0/4	0/4	1/4	0/4
4	0/3	–	1/3	0/3
5	0/3	–	2/3	2/3
sum	0.25	0.0	1.25	0.6
Min-max scaler	0.2	0.0	1.0	0.48

جدول–13: معدل عدم الإشتراك للمستخدم رقم (4)

Session_id/ item	a	b	c	d
1	0/4	1/4	2/4	2/4

3	1/4	1/4	0/4	1/4
4	1/3	–	0/3	1/3
5	1/3	–	0/3	0/3
sum	0.917	0.5	0.5	1.08
Min-max scaler	0.719	0	0	1

لحساب التقييمات الضمنية للمستخدم (4) نجمع المعدلات السابقة بإستخدام الأوزان المحددة لكل معدل كما يلي:

a = (0.5 *0.2) + (0.2*0.5) + (0.7719*0.3) = 0.416

b = (– *0.2) + (0.0*0.5) + (0.0*0.3) = 0.0

c = (0.25*0.2) + (1.0*0.5) + (0.0*0.3) = 0.55

d = (0.25*0.2) + (0.48*0.5) + (1.0*0.3) = 0.59

بعد جمع المعدلات يتم تحديد التقييمات الضمنية للنقر لكل مستخدم لمجموعة العناصر، يتم تحديد المستخدمين و العناصر في مصفوفة التقييمات الضمنية للنقر حسب تفاعل المستخدم مع العنصر، الجدول–14 التالي يوضح التقييمات الضمنية للنقر المشتقة لكل من المستخدم (1، 3، 4) للعناصر الموضحة في الجدول التالي، في الجدول يشير الرمز (؟) للتقييمات المفقودة للعناصر التي ليس لها تقييمات ضمنية للنقر.

جدول–14: مصفوفة التقييمات الضمنية للنقر

User/ item	a	b	c	d	e
1	0.381	0.0	0.566	0.606	؟
3	0.036	0.716	0.3	0.0	0.434
4	0.416	0.0	0.55	0.59	؟

سيتم تغذية مصفوفة التقييمات الضمنية بالتقييمات الضمنية التي تم إشتقاقها من جلسات نقر المستخدم، بحيث يراعي عند إضافة التقييمات إضافة تقييمات العناصر التي لم يسبق للمستخدم تقييمها (بمعني آخر ليس لديها تقييمات في مصفوفة الشراء)، كذلك إضافة التقييمات التي لا تقل عن الحد الأدني للتقييم الضمني (0.3)، في الجدول (5–19) تم تظليل العناصر التي لا تحقق شروط الإضافة.

الجدول–15 التالي يوضح مصفوفة التقييمات الضمنية الناتجة بعد إضافة التقييمات الضمنية للنقر. في الجدول يشير الرمز (؟) إلي التقييمات المفقودة للعناصر التي ليس لها تقييمات ضمنية للشراء أو للنقر.

جدول–15: مصفوفة التقييمات الضمنية

User/item	a	b	c	d	e
1	0.894	0.447	0.566	0.606	؟
2	؟	0.554	0.832	؟	؟
3	؟	0.716	0.707	0.707	0.434
4	0.416	؟	0.55	0.59	؟

- تمثل مصفوفة التقييمات الضمنية المدخل لخوارزمية التصفية التعاونية و التي تنتج مصفوفة التقييمات المتوقعة بناءً علي تلك التقييمات الضمنية للعناصر التي لم يقييمها المستخدمون. نلاحظ في الجدول–15 لمصفوفة التقييمات الضمنية أنه تمت إضافة تقييمات لعناصر جديدة للمستخدمين مثل العنصر (c) للمستخدم رقم (1)، كذلك تم إضافة عناصر جديدة للمصفوفة كالعنصر(e) للمستخدم رقم (3)، كما تم إضافة مستخدمين جدد للمصفوفة كالمستخدم رقم (4). تمثل نسبة تناثر التقييمات في مصفوفة التقييمات الضمنية نسبة 35% و هي نسبة أقل من نسبة مصفوفة الشراء مما يعني أن التقييمات الضمنية تقلل من تأثير مشكلة قلة بيانات مصفوفة التقييمات.

6- التجربة و التقييم:

تم إجراء التجارب لحساب نسبة تناثر التقييمات في مصفوفة التقييمات و لإختبار مدي كفاءة النموذج المقترح و الذي يعتمد علي التقييمات الضمنية مقارنة مع أسلوب التوصية الأساسي و الذي يعتمد علي مشتروات المستخدم كتقييمات صريحة للمستخدم. تم إستخدام مجموعة البيانات التجارية التي تخص بائع تجزئة عبر الإنترنت في أوروبا و التي أُنشئت بواسطة شركة YOOCHOOSE GmbH لدعم المشاركين في تحدي RecSys 2015 [11]، تتضمن مجموعة البيانات علي (87,318) جلسة شراء و (2,152,741) جلسة نقر، ليصبح العدد الكلي للجلسات بعد تجميع جلسات النقر و جلسات الشراء هو (2,235,663) جلسة، نظراً لنقص بيانات المستخدم في مجموعة البيانات المستخدمة تم إضافة

بيانات المستخدم و توزيعها بشكل عشوائي. تم تقسيم مجموعة البيانات بإستخدام طريقة K-Fold Cross Validation، حيث تم تقسيم البيانات إلي 10 مجموعات متساوية الحجم تقريباً، يتم إختيار إحدي المجموعات لتمثل بيانات الإختبار، بينما تمثل بقية المجموعات و عددها 9 مجموعات بيانات التدريب. تم تقييم النموذج المقترح بإستخدام مكتبة CaseRecommender [12]، تم مراعاة طرق التوصية التعاونية المختلفة (التوصية بناءً علي المستخدم، التوصية بناءً علي العنصر)، كما تم مراعاة طرق قياس التشابه المختلفة (معامل إرتباط بيرسون، مقياس جيب التمام).

7- النتائج:

بحساب نسبة تناثر البيانات كما في الجدول-16 و الجدول-17 نجد أن إضافة التقييمات الضمنية المشتقة من بيانات شراء المستخدم و من تفاعلاته إلي مصفوفة التقييمات قد قلل من نسبة تناثر التقييمات في المصفوفة من نسبة (99.78%) إلي نسبة (99.71%) في مرحلة التدريب و ذلك بإضافة (37,553) تقييم ضمني، و من نسبة (99.92%) إلي نسبة (99.91%) في مرحلة الإختبار و ذلك بإضافة (4,173) تقييم ضمني كما يوضح الشكل-3. هذه النسب و إن كانت ضئيلة إلا أنها تقلل من تأثير فجوة تناثر تقييمات المستخدم في المصفوفة مقارنة مع تقييمات مشتروات المستخدم، كما أنها تساهم في زيادة المعرفة المكتسبة عن المستخدم.

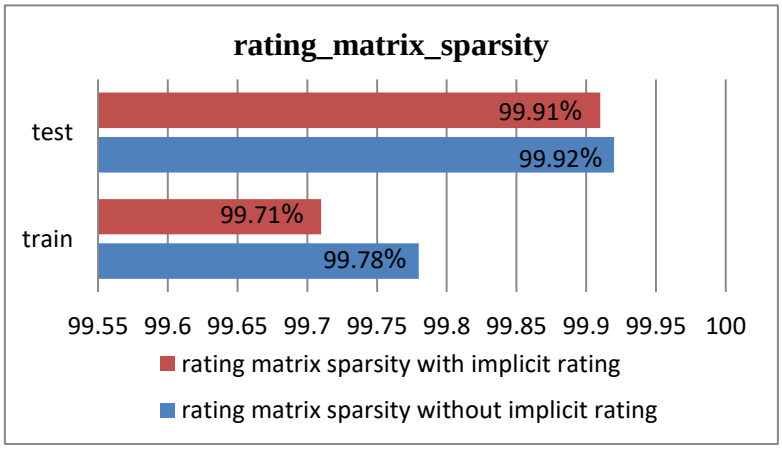
جدول-16: نسبة تناثر التقييمات لنموذج التوصية المقترح

	users	items	interactions	sparsity
QV_user_cosine				
train data	6457	9381	173190	99.71%
test data	6070	3489	19244	99.91%
QV_user_correlation				
train data	6457	9381	173190	99.71%
test data	6070	3489	19244	99.91%
QV_item_cosine				
train data	6457	9381	173190	99.71%
test data	6070	3489	19244	99.91%
QV_item_ correlation				

train data	6457	9397	173190	99.71%
test data	6057	3453	19244	99.91%

جدول–17: نسبة تناثر تقييمات أسلوب التوصية الأساسي

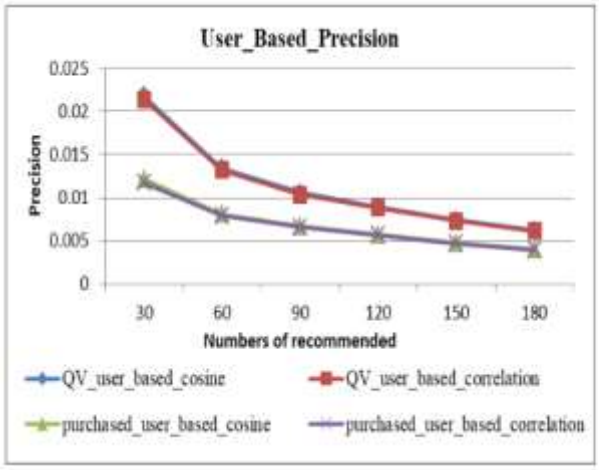
	users	items	interactions	sparsity
purchase_user_based_cosine				
train data	6457	9358	135637	99.78%
test data	5701	3427	15071	99.92%
purchase_user_based_correlation				
train data	6457	9377	135637	99.78%
test data	5749	3449	15071	99.92%
purchase_item_based_cosine				
train data	6457	9342	135637	99.78%
test data	5733	3425	15071	99.92%
purchase_item_based_correlation				
train data	6457	9399	135637	99.78%
test data	5700	3414	15071	99.92%

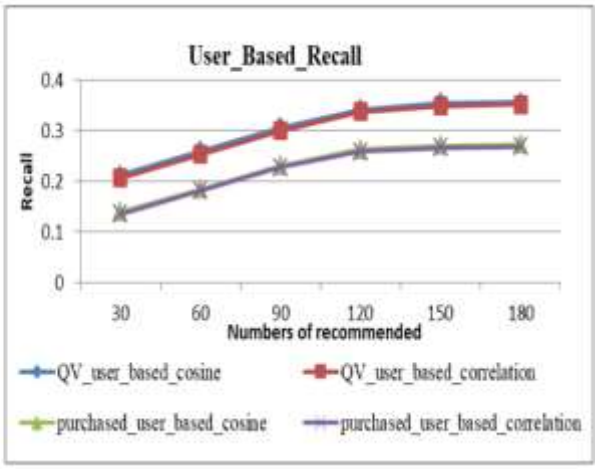


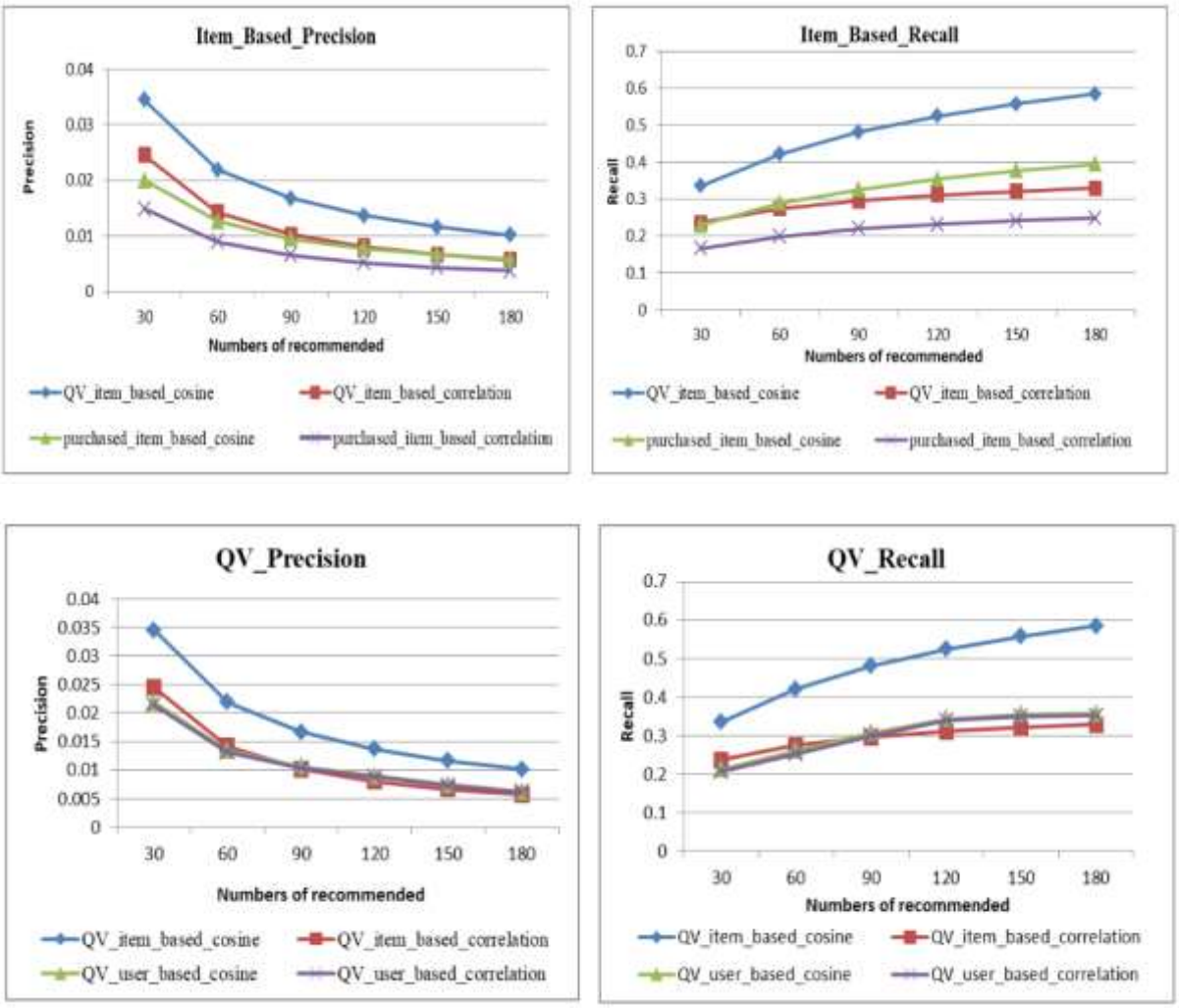
شكل–3: نسبة تناثر البيانات في مصفوفة التقييمات قبل و بعد إضافة التقييمات الضمنية

وفقاً لمقاييس الدقة و الإستدعاء أثبتت التجارب التي تقارن أداء نموذج التوصية المقترح مع أسلوب التوصية الأساسي تفوّق نموذج التوصية المقترح (QV_based) علي نموذج التوصية الأساسي

(purchased_based) في حالتي التوصية بناءً علي المستخدم و التوصية بناءً علي العنصر و بإختلاف طرق التشابه. كذلك أظهرت النتائج أفضلية نموذج التوصية المقترح بناءً علي العنصر بإستخدام مقياس جيب التمام (QV_item_based_cosine) في تقديم توصيات ذات دقة عالية مقارنة مع طرق التوصية الأخري و التي تتضمن نموذج التوصية المقترح بناءً علي العنصر بإستخدام معامل إرتباط بيرسون (QV_item_based_correlation) و نموذج التوصية المقترح بناءً علي المستخدم بإستخدام معامل إرتباط بيرسون (QV_user_based_correlation) و نموذج التوصية المقترح بناءً علي المستخدم بإستخدام مقياس جيب التمام (QV_user_based_cosine)، إذ أن نموذج التوصية المقترح بناءً علي العنصر بإستخدام مقياس جيب التمام و بإختلاف عدد العناصر الموصي بها يعطي القيم الأعلي للإستدعاء و هو نسبة العناصر ذات الصلة بالمستخدم و التي تم التوصية بها من نسبة جميع العناصر التي تم التوصية بها، حيث تصل نسبة العناصر الموصي بها و التي تقع ضمن تفضيلات المستخدم إلي ما يقارب (60%) كحد أقصي من مجموع العناصر الموصي بها، و هي نسبة جيدة مقارنة مع نسبة الإستدعاء لأسلوب التوصية الأساسي و التي تقارب (40%) كحد أقصي من مجموع العناصر الموصي. وفقاً لهذه التجارب تشير نتائج التقييم إلي أن المستخدم يفضل العناصر التي تشابه العناصر التي تفاعل معها أو قام بشرائها مسبقاً أكثر من العناصر التي تفاعل معها أو قام بشرائها المستخدمون المشابهون له.







شكل–4: الدقة و الإستدعاء لنموذج التوصية المقترح و أسلوب التوصية الأساسي

مما سبق يمكن تلخيص النتائج التالية:

1) المعرفة المكتسبة عن المستخدم و المتمثلة في التقييمات الضمنية التي تم إشتقاقها من بيانات شراء المستخدم و من بيانات تدفق نقر المستخدم تقلل من تناثر تقييمات المستخدم في مصفوفة التقييمات.
2) يمكن تخصيص توصيات للمستخدم بإستخدام تقنية التصفية التعاونية بناءً علي التقييمات الضمنية المشتقة من بيانات شراء المستخدم و من بيانات تدفق نقر المستخدم.
3) التقييمات الضمنية التي تم إشتقاقها من بيانات شراء المستخدم و من بيانات تدفق نقر المستخدم تحسن من دقة التوصيات من نسبة 40% إلي 60% كحد أقصي للتوصيات.

8- الخلاصة:

تستخدم نظم التوصية العديد من طرق تصفية المعلومات منها طريقة التصفية التعاونية، تستند التصفية التعاونية علي التقييمات التي يقدمها المستخدم و التي بموجبها يتم تقديم توصيات ذات علاقة بالمستخدم و ترضي تفضيلاته، إلا أن المستخدم قد يضن بتقديم هذه التقييمات، إذ أنها مجهدة للمستخدم أو قد لا يعبأ المستخدم بتحري الصحة في تقديمها، مما ينتج عنه توصيات تفتقر إلي الدقة. ينصب التركيز الرئيسي لهذه الورقة علي تسخير مصادر المعلومات المتاحة كمعاملات شراء المستخدم و سجل سلوك المستخدم للتعرف علي تفضيلات المستخدم، و من ثم إستخدام هذه المعرفة لتقديم توصيات ذات مغزي للمستخدم. تم الإستفادة من بيانات شراء المستخدم متمثلة في كمية المشتروات، و من بيانات تدفق نقر المستخدم متمثلة في عدد مرات النقر لإشتقاق تقييمات ضمنية للمستخدم تعكس درجة تفضيل المستخدم للعنصر، ومن ثم تم إثراء مصفوفة التقييمات التعاونية بهذه التقييمات الضمنية بغرض التقليل من التأثير السلبي لتناثر بيانات التقييم. تم إستخدام بيانات متجر تجاري للبيع بالتجزئة لإشتقاق التقييمات الضمنية للمستخدم و من ثم تم تطبيق تقنية التصفية التعاونية علي مصفوفات تقييمات المستخدم الضمنية المشتقة، أظهرت النتائج وفقاً للبيانات المستخدمة أن التقييمات الضمنية المشتقة لعبت دوراً إيجابياً في تقليل نسبة تناثر تقييمات المستخدم في مصفوفة التقييمات من حيث الكم، كذلك أثبتت النتائج كفاءة التقييمات الضمنية المشتقة في تحسين دقة التوصيات المقدمة للمستخدم، و ذلك عن طريق تقديم توصيات بعناصر تشابه العناصر التي تفاعل معها المستخدم أو قام بشرائها مسبقاً.

المصادر و المراجع:

[1] F. Ricci, L. Rokach, B. Shapira and P. Kantor, “Recommender Systems Handbook”, 1st. ed., Springer-Verlag, Berlin, Heidelberg, 2011.

[2] A. Martí nez, C. Schmuck, S. Pereverzyev, C. Pirker and M. Haltmeier, “A machine learning framework for customer purchase prediction in the non-contractual setting”. 2020, Eur. J. Oper. Res., 281, 588-596.

[3] K. Kumari, H. Khajgiwale, S. Umare, A. Kumar, G. Phadtare and V. Chaudhary, “Comportement Analyse And Recommendation Using ClickStream Data”. Int. J. of Advance Research and Innovative Ideas in Education, 2017, 3, 1256-1260.

[4] D. Köhn, S. Lessmann and M. Schaal, “Predicting Online Shopping Behaviour from Clickstream Data using Deep Learning”, 2020, Expert Systems with Applications. 150. 113342. 10.1016/j.eswa.2020.113342.

[5] K. Choi, D. Yoo, G. Kim and Y. Suh, “A hybrid online-product recommendation system: Combining implicit rating-based collaborative filtering and sequential pattern analysis”. Electron. Commer. Rec. Appl. 11, 4 (July, 2012), 309–317, 2012, DOI:https://doi.org/10.1016/j.elerap.2012.02.004

[6] Y. Xiao and C. Ezeife, "E-Commerce Product Recommendation Using Historical Purchases and Clickstream Data", DaWaK, 2018.

[7] B. Wang, F. Ye and J. Xu, "A Personalized Recommendation Algorithm Based on the User's Implicit Feedback in E-Commerce", Future Internet. 10. 117. 10.3390/fi10120117, 2018.

[8] H. Hwangbo and Y. Kim, "Session-Based Recommender System for Sustainable Digital Marketing", Sustainability, 11, 3336., 2019, https://doi.org/10.3390/su11123336

[9] P. Chavan, B. Thoms and J. Isaacs, "A Recommender System for Healthy Food Choices: Building a Hybrid Model for Recipe Recommendations using Big Data Sets", 2021, 10.24251/HICSS .2021.458.

[10] M. Farashah, A. Etebarian, R. Azmi and R. Dastjerdi, "A hybrid recommender system based-on link prediction for movie baskets analysis". J Big Data 8, 32 (2021), 2021, https://doi.org/10.1186/s40537-021-00422-0

[11] D. Ben-Shimon, A. Tsikinovsky, M. Friedman, B. Shapira, L. Rokach and J. Hoerle, Recsys challenge 2015 and the yoochoose dataset, In Proc. of the 9th ACM Conf. on Recommender systems. ACM, September 2015.

[12] A. Costa, E. Fressato, F. Neto, M. Manzato and R. Campello, “Case recommender: a flexible and extensible python framework for recommender systems”. In Proc. of the 12th ACM Conf. on Recommender Systems (RecSys '18). ACM, NY, USA, 2018 494-495. DOI: https://doi.org/10.1145/3240323.3241611.