Graduate College Journal – NU
Vol.15 -2020, No. (7)
ISSN: 1858-6228, http://www.neelain .edu.sd

مجلة الدراسات العليا - جامعة النيلين
المجلد 15 –2020، العدد (7)
الرقم الدولي الموحد للدوريات: 1858-6228

تحديات تقنيتي التصفية التعاونية و التصفية بناءً علي المحتوي

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

المستخلص

أصبحت أنظمة التوصية (Recommender Systems) الحل لمواجهة مشكلة التحميل الزائد للمعلومات و تخصيص الخدمات في تطبيقات الويب، فهي تساعد المستخدم في عملية البحث عن المعلومات و اتخاذ القرار، و تَستخدِم لذلك العديد من التقنيات منها تقنية التصفية التعاونية (Collaborative Filtering) و تقنية التصفية بناءً علي المحتوي (Content-Based Filtering) ، هذه الورقة تلخِّص و تستعرض التحديات التي تواجه هاتين التقنيتين و الحلول المقترحة التي تطرقت إليها بعض الدراسات لمواجهة تلك التحديات، خلُصت هذه الورقة إلي أن هذه التحديات ما زالت بيئة خصبة للبحث و التطوير و ذلك بغرض زيادة دقة نظم التوصية و تقديم مقترحات أقرب للمستخدم، إذ أن بعض الحلول أظهرت تحديات أخري، مثل دقة التوصيات في حالة التجميع و جودة البيانات في حالة ملء تقييمات مصفوفة البيانات.

الكلمات المفتاحية Keywords: نظم التوصية (Recommender Systems) ، التوصية التعاونية (Collaborative Filtering)، التوصية بناءً علي المحتوي (Content-Based Filtering).

1- المقدمة:

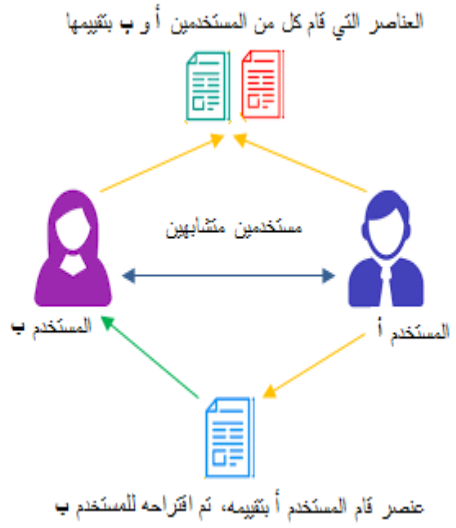
تستخدم نظم التوصية في العديد من تطبيقات الويب مثل التجارة الإلكترونية و بوابات الأخبار و الإعلانات وما إلى ذلك، فهي تعتبر أداة تساعد المستخدم علي التغلب علي التحميل الزائد للمعلومات كما تساعده في عملية البحث عن المعلومات و اتخاذ القرار مثلاً : أي منتج سيشتري؟ أو أي كتاب سيقرأ؟. حيث تقوم نظم التوصية باستخدام تقنيات تصفية البيانات لتقديم مقترحات لعناصر تتوافق مع ما يفضله المستخدم . لخّص (Jannach, D و آخرون، 2010) هذه التقنيات في ثلاث تقنيات وهي: تقنية التصفية التعاونية (Collaborative Filtering) و تقنية التصفية بناءً علي المحتوي (Content-Based Filtering) و تقنية التصفية بناءً علي المعرفة (Knowledge-Based Filtering) (لا تشملها هذه الورقة)، كما يمكن تصفية البيانات عن طريق دمج التقنيات السابقة و ذلك فيما يُعرف بتقنية التصفية المختلطة (Hybrid Filtering). تختلف نظم التوصية بإختلاف التقنية المستخدمة لتصفية البيانات، إذ أن لكل تقنية طريقة مغايرة عن التقنية الأخري، مثلاً تعتمد تقنية التصفية التعاونية علي بيانات المستخدم و بيانات المستخدمين الآخرين، أما تقنية التصفية بناءً علي المحتوي فتعتمد علي بيانات المستخدم و بيانات العنصر. في هذه الورقة سوف نستعرض التحديات التي تواجه تقنية التصفية التعاونية و تقنية التصفية بناءً علي المحتوي و نوضح الحلول المقترحة التي تناولتها بعض الدراسات ذات العلاقة.

تم تنظيم الورقة على النحو التالي: في القسم 2 سوف نوضح طريقة عمل كل تقنية ، يليه القسم 3 يوضح التقنيات الفرعية و التقنيات المختلطة، أما القسم 4 يتناول مزايا كل تقنية، في القسم 5 سنتناول التحديات و الحلول المقترحة في التقنيتين ، نختتم بالخلاصة في القسم 6.

2- طريقة تصفية البيانات في تقنية التصفية التعاونية و تقنية التصفية بناءً علي المحتوي :

1-2 طريقة تصفية البيانات في تقنية التصفية التعاونية:

الفكرة الأساسية لتقنية التصفية التعاونية هي أنه إذا كان المستخدمون قد تشاركوا نفس الاهتمامات في الماضي فسيكون لهم أيضًا تفضيلات متشابهة في المستقبل. إذ تعتمد هذه التقنية على جمع و تحليل بيانات سلوك المستخدمين تجاه العناصر، حيث يتم التعبير عن العلاقة بين المستخدمين و العناصر من خلال التقييم (Rating)، و هو يمثل درجة تفضيل المستخدم للعنصر و التي قد يُعبر عنها المستخدم صراحةً أو ضمنياً. يعتبر المستخدمون متشابهون في حال إشتركوا في تقييم نفس العناصر، عندها يمكن التوصية بالعناصر التي قيّمها أحد المستخدمين بشكل إيجابي للمستخدم الآخر، كما في (الشكل-1).



الشكل-1: تقنية التصفية التعاونية

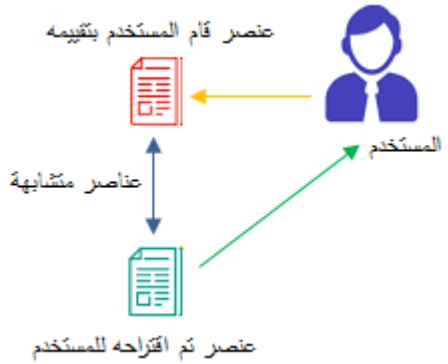
2-2 طريقة تصفية البيانات في تقنية التصفية بناءً علي المحتوي:

تعتمد تقنية التصفية بناءً علي المحتوي علي تحليل مجموعة مواصفات العناصر التي تم تقييمها مسبقًا من قبل المستخدم، حيث يتم بناء أو تمثيل ملف تعريف تفضيلات المستخدم (User Profile) استناداً إلى مواصفات العناصر (Items Features) التي قام المستخدم بتقييمها، ومن ثم استخدام هذا الملف لتقديم توصية أو مقترحات للمستخدم بعناصر تتطابق مواصفاتها مع الملف التعريفي له و تكون ذات تقييم إيجابي كما هو موضح في (الشكل-2).

أوضح (Aggarwal,C.، 2016) أن تقنية التصفية بناءً على المحتوى تعتمد على مصدرين للبيانات:

1) المصدر الأول للبيانات هو وصف للعناصر المختلفة حيث يتم تمثيل مواصفات المحتوى، مثل الوصف النصي لعنصر من قبل الشركة المصنعة.

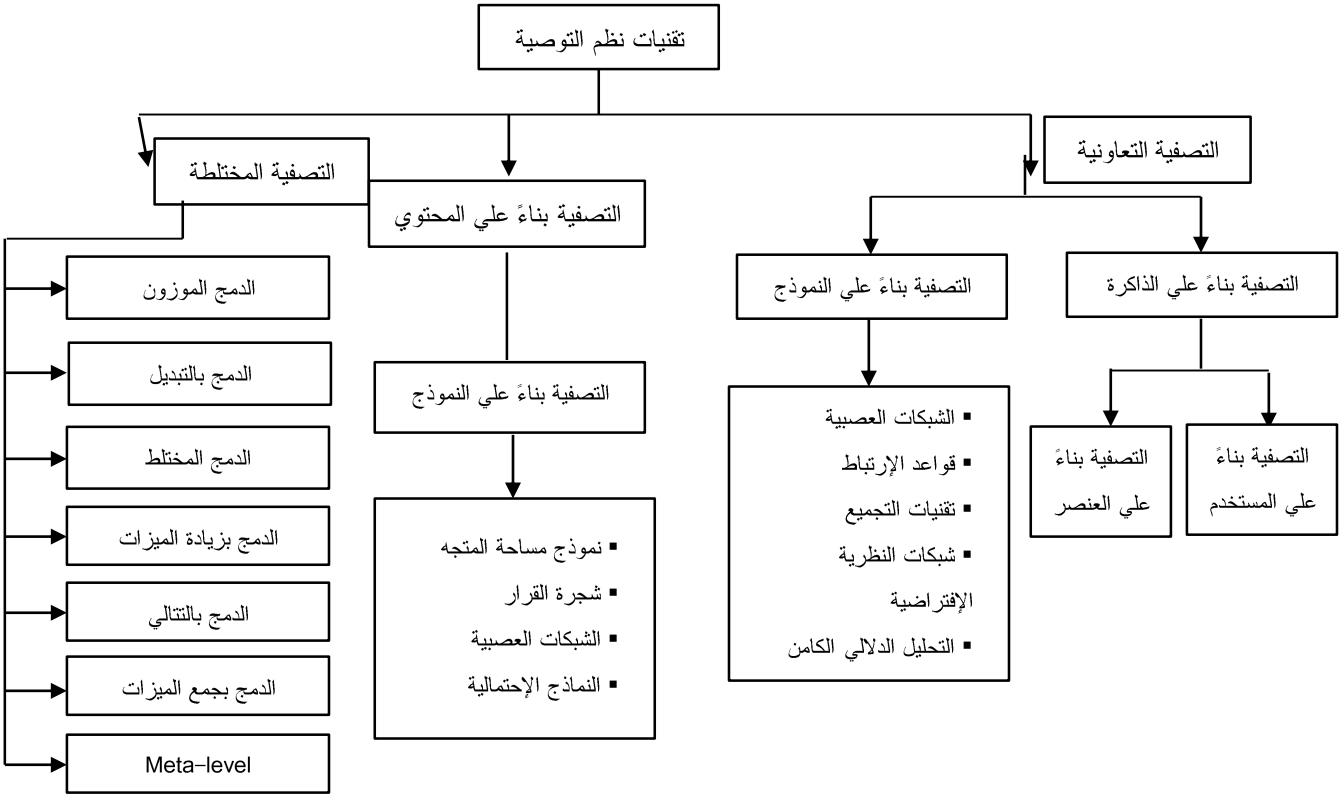
2) المصدر الثاني للبيانات هو ملف تعريف المستخدم، و يتم إنشاء ملف التعريف عن طريق جمع تقييمات المستخدم للعناصر المختلفة، يقوم المستخدم بتقييم هذه العناصر إما بصورة صريحة أو بصورة ضمنية (كما في تقنية التصفية التعاونية).



الشكل-2 : تقنية التصفية بناءً علي المحتوي.

3- التقنيات الفرعية و المختلطة في تقنية االتصفية التعاونية و تقنية التصفية بناءً علي المحتوي:

تحتوي كل من تقنية التصفية التعاونية و تقنية التصفية بناءً علي المحتوي علي تقنيات فرعية و نماذج تنبؤية يتم استخدامها لتصفية البيانات في نظم التوصية، كما يمكن تصفية البيانات من خلال دمج هذه التقنيات و ذلك فيما يعرف بالتقنية المختلطة. (الشكل-3) يوضح التقنيات الفرعية و التقنيات المختلطة لكلا التقنيتين و التي سنتناولها بالشرح في هذا القسم.



الشكل- 3: التقنيات الفرعية و التقنيات المختلطة في تقنية االتصفية التعاونية و تقنية التصفية بناءً علي المحتوي – تمت ترجمته من (Sinha,B., Dhanalakshmi, R.، 2019)

1-3 التقنيات الفرعية من تقنية التصفية التعاونية:

تعمل تقنية التصفية التعاونية علي استغلال تقييمات المستخدم بصورة مباشرة أو غير مباشرة للتنبؤ بما سيفضله المستخدم، ففي حالة إستخدام التقييمات بصورة مباشرة و هو النوع الأول من طرق التصفية التعاونية يكون التنبؤ استناداً إلى تشابه أو تجاور المستخدم أو العنصر وهي ما عرّفه (Francesco و آخرون، 2010) بطرق التصفية بناءً علي التجاور (neighborhood-Based) أو طرق التصفية بناءً علي الذاكرة (Memory-Based methods) ، و تستخدم هذه الطرق العديد من الخوارزميات لقياس التشابه منها خوارزمية الجار الأقرب (k-Nearest Neighbors) و التي يمكنها قياس التشابه بين المستخدم و المستخدمين الآخرين و هو ما يسمي بالتصفية بناءً علي المستخدم (User-Based Filtering) أو التشابه بين العنصر و العناصر الأخري و هو ما يسمي بالتصفية بناءً علي العنصر (Item-Based Filtering). أما في النوع الثاني من طرق التصفية التعاونية يتم فيه إستغلال تقييمات المستخدم للعناصر بصورة غير مباشرة و ذلك عن طريق إستخدام تقييمات المستخدم لتعلم نموذج تنبؤي و تسمي بطرق التصفية بناءً علي النموذج (Model-Based methods)، وفي هذه الطريقة يتم بناء نموذج لتفاعلات المستخدم مع العنصر بالإضافة إلي العوامل التي تمثل الخصائص الكامنة للمستخدمين أو للعناصر، مثل فئات المستخدمين حسب التفضيل أو فئات العناصر حسب الأصناف، يلي ذلك تدريب هذا النموذج على استخدام البيانات المتاحة كي يستخدم لاحقًا للتنبؤ بتقييمات المستخدمين للعناصر الجديدة. إستخدمت العديد من الدراسات أساليب مختلفة لبناء نظم توصية تعاونية تعتمد علي النموذج مثلاً إستخدم (Covington, P. و آخرون، 2016) الشبكات العصبية (Neural Networks) حيث تم بناء شبكة عصبية لتوليد العناصر المرشحة للاقتراح و شبكة عصبية أخري لترتيب تلك المقترحات، أما (Gao, C. و آخرون، 2018) فقد إستخدموا الشبكات العصبية لتعلم تفاعلات و سلوك المستخدمين تجاه العناصر. كما يمكن بناء نموذج توصية بإستخدام قواعد الإرتباط (Association Rules) مثلاً (Cho,Y. و آخرون ، 2002) إستخدموا قواعد الإرتباط لتحديد الفئة التي ينتمي إليها العنصر، أما (Davidson,J. و آخرون، 2010) فقد إستخدموها لتعريف العناصر المتشابهة. كذلك من الممكن إستخدام تقنيات التجميع (Clustering Techniques) مثلاً قام (Li, Q. , Kim, B. ، 2003) بتجميع العناصر و تقديم معلومات عن المحتوي للعناصر التي تم تجميعها و (Das, A. و آخرون، 2007) قاموا بتجميع المستخدمين بناءً علي تفاعلهم مع العناصر. أيضاً من نماذج التصفية التعاونية إستخدام شبكات النظرية الافتراضية (Bayesian Networks) فلقد إستخدمها (Decampos, L. و آخرون، 2010) لتمثيل كيفية ارتباط المستخدمين و العناصر و قياس قوة هذا الإرتباط. كما يمكن إستخدام نموذج التحليل الدلالي الكامن (Latent Semantic Analysis) لبناء نظم توصية، حيث فرض (Hofmann, T.،2004) إحتمالية تواجد مشترك بين متغيرات مخفية و مجموعة الأزواج المتكونة من (مستخدم، عنصر). أما (Funk, S. ، 2006) فقد إستخدم طريقة معالجة عوامل مصفوفة البيانات لتصفية البيانات و ذلك فما يعرف باسم تحليل القيمة الفردية (Singular Value Decomposition).

2-3 التقنيات الفرعية في تقنية التصفية بناءً علي المحتوي:

لا تحتوي تقنية التصفية بناءً علي المحتوي علي تقنيات فرعية غير أنه كما بيّن (Aggarwal, C.، 2016) من الممكن إنشاء الملف التعريفي للمستخدم عبر بناء نموذج للتنبؤ باهتمامات المستخدم، إذ يتم إستخدام تقييمات المستخدم بالإضافة إلي مواصفات العناصر لبناء النموذج و من ثم تدريبه. تستخدم تقنية التصفية بناءً علي المحتوي العديد من الأساليب لبناء نموذج التصفية، مثل إستخدام نموذج مساحة المتجه (Vector Space Model) كما في (Lops, P. و آخرون، 2011) حيث تم بناء ملفات تعريف المستخدمين عن طريق تحويل تمثيل الكلمات المفتاحية إلي تمثيل معاني هذه الكلمات و ذلك بإستخدام المعرفة المعجمية المخزنة في قاعدة البيانات المعجمية MultiWordNet. كذلك من الممكن بناء النموذج بإستخدام شجرة القرار Decision Tree كما في (Pazzani, M. و آخرون، 1996) و (Cantador, I. و آخرون، 2009). كذلك يمكن بناء النموذج بإستخدام الشبكات العصبية مثل (Musto,C. و آخرون، 2016) فقد قاموا ببناء شبكة عصبية عميقة قادرة علي تقديم مقترحات تم تعلمها من المعلومات النصية عن طريق السؤال و الجواب، و ذلك بإستغلال التشابه بين الأسئلة و ملفات تعريف المستخدمين مع الإجابات و العناصر. ايضاً من أساليب بناء النموذج إستخدم النماذج الإحتمالية كما في (Pazzani, M. و آخرون، 1996) و (Jayech, K. , Mahjoub, M.، 2012) و (Nguyen, S ، 2019)

3-3 التقنيات المختلطة في تقنية التصفية التعاونية و تقنية التصفية بناءً علي المحتوي:

يقصد بتقنية التصفية المختلطة (Hybrid Filtering Technique) الجمع بين تقنيتين أو أكثر بطرق مختلفة و ذلك لتقوية نقاط الضعف التي تنطوي علي استخدام هذه التقنيات بصورة مفردة، أوضح (Burke, R.، 2002) أنه يمكن دمج هذه التقنيات بعدة طرق وهي:.

1) **الدمج الموزون (Weighted)** : يتم فيه جمع نتائج العديد من تقنيات التوصية بصورة عددية.
2) **الدمج بالتبديل (Switching)** : يتم فيه التبديل بين تقنيات التوصية وفقًا للوضع الحالي.
3) **الدمج المختلط (Mixed)** : يتم فيه تقديم توصيات من عدة تقنيات مختلفة في نفس الوقت.
4) **الدمج بجمع الميزات (Feature combination)** : يتم فيه تجميع الميزات من مصادر بيانات التوصية المختلفة معًا في خوارزمية توصية واحدة.
5) **الدمج بزيادة الميزات (Feature augmentation)** : يتم استخدام تقنية توصية واحدة لحساب ميزة أو مجموعة من الميزات ، والتي تعد جزءًا من الإدخال للتقنية التالية.
6) **الدمج بالتتالي (Cascade)** : يتم فيه التوصية بالتتالي بحيث يتم استغلال التوصيات المسبقة كأساس لإشتقاق التوصيات الحالية.

7) **الدمج عن طريق Meta-level** : يتم تطبيق تقنية توصية واحدة وتنتج نوعًا من النموذج ، وهو ما يمثل عندئذ المدخلات المستخدمة في التقنية التالية.

بيّن (Çano, E. , Morisio, M.، 2017) أن الجمع بين تقنيتي التصفية التعاونية و التصفية بناءً علي المحتوي من التقنيات الناجحة في كثير من الأنظمة، وفقاً لما ذكره (Burke, R. ، 2002) يعتبر الدمج الموزون و الدمج بالتبديل و الدمج بجمع الميزات و الدمج بزيادة الميزات من أكثر طرق الدمج المتكررة بين هاتين التقنيتين.

4- مزايا تقنية التصفية التعاونية و تقنية التصفية بناءً علي المحتوي:

لدي كل من تقنية التصفية التعاونية و تقنية التصفية بناءً علي المحتوي العديد من المزايا التي تجعل كل تقنية تختلف عن التقنية الأخري، أوضح (Ricci, F. و آخرون، 2010) أن تقنية التصفية التعاونية تتغلب على بعض القيود المفروضة على تقنية التصفية بناءً علي المحتوى، فالعناصر التي قد لا يتوفر لها محتوي أو يصعب الحصول عليه، يمكن لتقنية التصفية التعاونية تقديم توصية أو مقترحات بهذه العناصر من خلال تقييمات المستخدمين الآخرين. علاوة على ذلك تعتمد التوصيات التعاونية على جودة العناصر التي يتم تقييمها من قِبل المستخدمون بدلاً من الاعتماد على المحتوى الذي قد يكون مؤشراً سيئاً لجودة العنصر، كذلك و بخلاف الأنظمة بناءً على المحتوى يمكن لأنظمة التصفية التعاونية أن توصي بعناصر ذات مواصفات مختلفة تماماً طالما أبدى المستخدمون الآخرون إهتماماً بهذه العناصر، مما يضفي بعض التنوع في العناصر المقترحة. و بالمقابل تتغلب تقنية التصفية بناءً علي المحتوي علي تقنية التصفية التعاونية في أنها تضمن إستقلالية المستخدم إذ أنها تقدم توصية أو مقترحات بالإعتماد علي الملف التعريفي للمستخدم و الذي يستند علي تقييمات المستخدم الحالي دون الحاجة لتقييمات مستخدمين أُخر (المستخدمين المشابهين/المجاورين). علاوة علي ذلك تمتاز هذه التقنية بالشفافية إذ أنها يمكن أن تقدم توضيحات للمستخدم عن سبب اختيار العناصر التي تم التوصية بها أو اقتراحها و ذلك من خلال توضيح مواصفات تلك العناصر والتي تطابق مواصفات العناصر الموجودة بالملف التعريفي للمستخدم أي العناصر التي تم تقييمها مسبقاً من قبل المستخدم نفسه و ليس من قِبل مستخدمين أُخر (المستخدمين المشابهين/المجاورين). كذلك يمكن لهذه التقنية تقديم توصية أو مقترحات لعناصر لم يتم تقييمها من قِبل أي مستخدم، أو تقديم مقترحات لعناصر غير شائعة أو غير رائجة، إذ أنها تعتمد علي مواصفات العنصر و الملف التعريفي للمستخدم دون التأثر بتقييمات مستخدمين أُخر (المستخدمين المشابهين/المجاورين)، مما يجعل من السهل التوسع إلي عدد كبير من المستخدمين.

يمكن تلخيص ما سبق في (الجدول-1) حيث نقارن بين تقنية التصفية التعاونية و تقنية التصفية بناءً علي المحتوي وفقاً للمزايا المذكورة لتوضيح إمكانية توفر كل ميزة أو عدم توفرها في كل تقنية.

الجدول-1: مزايا تقنية التصفية التعاونية و تقنية التصفية بناءً علي المحتوي:

الرقم	الميزة	تقنية التصفية التعاونية	تقنية التصفية بناءً علي المحتوي
1	**عدد المستخدمين**	تعتمد علي عدد من المستخدمين المتشابهين	تعتمد علي مستخدم واحد
2	**إضافة عنصر جديد**	لا تقدم مقترحات للعنصر الجديد لأنها تعتمد علي وجود تقييمات مسبقة للعنصر	تقدم مقترحات للعناصر الجديد لأنها تعتمد علي مواصفات العنصر
3	**مواصفات العنصر**	تقدم مقترحات لعناصر ليست لديها مواصفات لأنها تعتمد علي تقييمات هذه العناصر	لا تقدم مقترحات لعناصر ليست لديها مواصفات لأنها تعتمد عليها في التوصية
4	**جودة العناصر المقترحة**	تعتمد جودة العناصر المقترحة علي تقييمات المستخدمين المشابهين (مؤشر قوي)	تعتمد جودة العناصر المقترحة علي مواصفات المحتوي (مؤشر ضعيف)
5	**إستقلالية المستخدم**	تقدم مقترحات مرتبطة بإختيارات المستخدمين المشابهين	تقدم مقترحات مرتبطة بإختيارات المستخدم لأنها تعتمد علي الملف التعريفي للمستخدم
6	**الشفافية**	لا تقدم تفسير للمستخدم عن سبب إختيار المقترحات لأنها تعتمد علي إختيارات المستخدمين المشابهين	تقدم تفسير للمستخدم عن سبب إختيار المقترحات لأنها تطابقها مع مواصفات عناصر المستخدم
7	**تنوع المقترحات**	تقدم مقترحات قد تتنوع لأنها تعتمد علي إختيارات المستخدمين المشابهين	تقدم مقترحات تطابق مواصفات العناصر التي قام المستخدم بتقييمها مسبقاً
8	**تغير سلوك المستخدم**	تدرك التغيير في سلوك المستخدم على مدى الزمن (Sinha,B., Dhanalakshmi, R. ، 2019)	لا تدرك التغيير في سلوك المستخدم على مدى الزمن

5- التحديات و الحلول المقترحة في تقنية التصفية التعاونية و تقنية التصفية بناءً علي المحتوي:

تواجه كلا التقنيتين العديد من التحديات التي تتعلق بجمع البيانات و تحديد المقترحات و غيرها، سوف نتناول هذه التحديات و نستعرض الحلول التي تطرقت إليها بعض الدراسات.

1-5 التحديات و الحلول المقترحة في تقنية التصفية التعاونية:

تواجه دقة نظم التوصية التي تستخدم طرق التصفية بناءً علي الذاكرة (Memory-Based methods) عدد من التحديات نوضحها فيما يلي:

1) **قلة البيانات (Data sparsity)**: و هو ما يعرف بمشكلة قلة أو تناثر البيانات وهي مشكلة شائعة في معظم أنظمة التوصية من هذا النوع و ذلك بسبب أن المستخدمين يقيّمون عادة نسبة صغيرة فقط من العناصر المتاحة و بالتالي يكون من الصعب تقديم مقترحات دقيقة، و ذلك لقلة المعلومات المتوفرة حول تقييمات المستخدمين الآخرين.

2) **البداية الباردة (Cold-Start)**: تظهر هذه المشكلة بسبب أن المستخدمين أو العناصر المضافة حديثاً إلى النظام قد لا تحتوي على تقييمات على الإطلاق مما يفاقم مشكلة قلة البيانات، لذا فمن المرجح أن لا تتوفر تقييمات مشتركة للمستخدمين أو للعناصر المضافة حديثاً لمقارنتها مع الموجودة مسبقاً.

3) **التحيز للعناصر الشائعة (Popularity Bias)** : و هي الحالة التي يتم تقديم المقترحات بالعناصر الشائعة بشكل متكرر، تظهر هذه المشكلة عندما لا تتوفر تقييمات مشتركة للمستخدمين أو للعناصر، و بالتالي يكون التنبؤ لعدد محدود من العناصر وهي العناصر الأكثر شيوعاً أو رواجاً بين المستخدمين المشابهين/المجاورين.

4) **قابلية التوسع (Scalability)**: ظهرت هذه المشكلة مع نمو أعداد المستخدمين و العناصر بصورة كبيرة، إذ تحتاج العديد من الأنظمة إلى الاستجابة الفورية للمتطلبات عبر الإنترنت وتقديم توصيات لجميع المستخدمين بغض النظر عن سجلهم التاريخي و تقييماتهم، الأمر الذي يتطلب زيادة إمكانية التوسع لدي هذه الأنظمة.

5) **Gray Sheep**: يقصد بها حالة عدم وجود مجموعة مشابهة للمستخدم، ففي حالة عدم الإتفاق أو عدم الإشتراك في التقييم مع أي مجموعة من مجموعات المستخدمين تكون المقترحات غير دقيقة و ذلك لعدم أو لقلة تواجد المستخدمين المشابهين/المجاورين.

تناولت العديد من الدراسات هذه التحديات و طرحت بعض الحلول تم توضيحها في الجدول (2)، ففي حالة قلة البيانات عمدت بعض الدراسات إلي استخدام ملء التقييمات المفقودة بقيم إفتراضية كما في (Konstan, J. و آخرون، 1997)، أما (Kim, B. , Li, Q. ،2003) و (Xue, G. و آخرون، 2005) فقد إستخدموا تقنيات التجميع لملء التقييمات المفقودة ببيانات المجموعات. في بعض الدراسات تم ملء التقييمات المفقودة بإستخدام معلومات المحتوى باعتبارها أكثر موثوقية كما في (Degemmis, M. و آخرون، 2007) و (Kula, M. ، 2015) البعض الآخر إستخدم تقييمات ناتجة من تقنية أخري لملء التقييمات المفقودة مثل (Geetha,G. و آخرون، 2018). ركزت بعض الدراسات علي إستغلال مصادر المعلومات الضمنية لجمع بيانات المستخدم إذ يحتوي هذا النوع من المصادر علي قدر كبير من المعلومات، و من ثم تحويل تلك البيانات التي تم جمعها إلي تقييمات لتغذية مصفوفة البيانات، كما في (Wang, B. و آخرون، 2018). و لمواجهة مشكلة التحيز للعناصر الشائعة فقد وضعوا (Gautam, A. و آخرون، 2014) معيار لقياس درجة حداثة العنصر لتفادي اقتراح العناصر الشائعة، أما (Abdollahpouri, H. و آخرون، 2019) فقد أعادوا ترتيب المقترحات و ذلك للترويج للعناصر الأقل شيوعاً. و لمواجهة تحدي قابلية التوسع تم استخدام تقنيات التجميع (Clustering) حيث تم تجميع المستخدمين باستخدام خوارزمية K-means كما في (Xue, G. و آخرون، 2005)، أما (Das, A. و آخرون ،، 2007) فقد قاموا بتجميع المستخدمين أو العناصر ثم قياس التشابه بين المستخدم و مجموعات التجميع ليتم اختيار المستخدمين المشابهين/المجاورين من المجموعة الأقرب للمستخدم. أما بالنسبة لمشكلة Gray Sheep لجأت بعض النظم إلي تقديم مقترحات بالعناصر الأكثر شيوعاً بين هذه المجموعة كما في (Kim, M. , Im, I.، 2014)، البعض الآخر عمد إلي تحويل المستخدمين الذين تختلف تفضيلاتهم مع المستخدم المستهدف إلى جيران متشابهين جدد و ذلك بفرض التشابه في الاختلاف مع المستخدم المستهدف كما في (Fazziki, A. و آخرون، 2019).

الجدول-2: تحديات تقنية التصفية التعاونية و بعض الحلول المقترحة و مصادرها

التحدي	الحلول المقترحة	المصدر
قلة البيانات		
	• ملء التقييمات المفقودة بقيمة متوسط تقييمات المستخدم.	(Konstan, J. و آخرون، 1997)
	• تجميع العناصر و ادخال تقييمات مجموعة العناصر التي تم تجميعها إلي تصفية تعاونية.	(Li, Q. , Kim, B. ،2003)
	• ملء التقييمات المفقودة من خلال تجميع المستخدمين ثم إستخدام تقنيات تجانس المجموعة (Cluster-Based Smoothing Technique) و معالجة اللغات الطبيعية (Nature Language Processing) لتقدير تلك التقييمات بناء علي المواضيع ذات العلاقة بالمجموعة.	(Xue, G. و آخرون، 2005)
	• إستخدم بيانات المحتوي لمعرفة التشابه بين المستخدمين و ذلك من خلال دمج المعرفة اللغوية (Linguistic Knowledge) في عملية تعلم ملفات تعريف المستخدمين.	(Degemmis, M. و آخرون، 2007)
	• إستخدام بيانات المحتوي لملء مصفوفة العوامل بالخصائص الكامنة (Latent Factors) المشتقة من بيانات تعريف العنصر و المستخدم.	(Kula, M. ، 2015)
	• تحويل سلوك المستخدم و تسلسل العناصر إلي تقييمات لتغذية مصفوفة البيانات.	(Wang, B. و آخرون، 2018)
البداية الباردة		
	• ملء التقييمات المفقودة بقيمة متوسط تقييمات المستخدم و ذلك للتمكن من تقديم مقترحات للمستخدم الجديد.	(Konstan, J. و آخرون، 1997)
	• تجميع العناصر و ادخال تقييمات مجموعة العناصر التي تم تجميعها إلي تصفية تعاونية.	(Li, Q. , Kim, B. ،2003)
	• إستخدام بيانات المحتوي لملء مصفوفة العوامل بالخصائص الكامنة (Latent Factors) المشتقة من بيانات تعريف العنصر و المستخدم.	(Kula, M. ، 2015)
	• ملء التقييمات المفقودة بالتقييمات المتوقعة الناتجة عن تطبيق تقنية التصفية بناءً علي المحتوي بالإضافة إلي تقنيات التجميع و التصنيف و من ثم إجراء تصفية تعاونية علي مصفوفة البيانات الناتجة.	(Geetha,G. و آخرون، 2018)
	• تحويل سلوك المستخدم و تسلسل العناصر إلي تقييمات لتغذية مصفوفة البيانات.	(Wang, B. و آخرون، 2018)

التحيز للعناصر الشائعة	• وضع مقياس لدرجة حداثة العنصر (Novelty Score) و الذي يزداد مع ازدياد شيوع العنصر بين المستخدمين و يتناقص مع ازدياد نسبة المستخدمين الذين قاموا بتقييم العنصر، يتم اقتراح العناصر المرتبطة بالمستخدم و التي لم يرها أو لم يقييمها من قبل.	(Gautam, A. و آخرون، 2014)
	• تخصيص المقترحات للمستخدم عن طريق إعادة ترتيب المقترحات بالترويج للعناصر ذات الذيل الطويل (Long-Tail Items) و ذلك بزيادة تمثيل العناصر الأقل شيوعاً و التي تقع ضمن اهتمامات المستخدم.	(Abdollahpouri, H. و آخرون، 2019)
قابلية التوسع	• تجميع المستخدمين باستخدام خوارزمية K-means ثم استخدام تقنيات تجانس بيانات المجموعة لملء التقييمات المفقودة.	(Xue, G. و آخرون، 2005)
	• تجميع المستخدمين بناءً علي تداخل تقييمات المستخدمين لنفس مجموعة العناصر.	(Das, A. و آخرون ،، 2007)
Gray Sheep	• تحديد المستخدمين ذوي التفضيلات الفريدة من خلال درجة مركزية المستخدم والتي تقاس بعدد الإرتباطات المباشرة من و إلي العقدة ثم تقديم مقترحات بالعناصر الشائعة بين هؤلاء المستخدمين.	(Kim, M. , Im, I.، 2014)
	• اختيار مشابهين جدد بجامع التشابه في عدم الاتفاق و بالتالي يتم ايجاد مستخدمين مشابهين.	(Fazziki, A. و آخرون، 2019)

2-5 التحديات و الحلول المقترحة في تقنية التصفية بناءً علي المحتوي:

أما تقنية التصفية بناءً علي المحتوي فتواجه العديد من التحديات التي تم توضيحها مع بعض الحلول المقترحة في الجدول (3)، و تتمثل هذه التحديات في:

1) **محدودية المحتوي (Limited Content Analysis)**: و ذلك عندما يتم تحليل عدد و نوع محدد من المواصفات المرتبطة بالعناصر المقترحة، و التي يتم اكتسابها إما بصورة تلقائية أو تقديمها من قِبل المستخدم، إذ أنه لا يمكن لهذه التقنية تقديم توصية أو مقترحات مناسبة إذا كان المحتوى الذي تم تحليله يحتوي على معلومات غير كافية، مثل استخراج الصفات الجمالية من النص أو استخراج معلومات عن الوسائط المتعددة كالصوت و الصورة، ففي مثل هذه الحالات لا تتوفر معلومات كافية لتمييز العناصر التي قد يفضلها أو لا يفضلها المستخدم.
2) **التخصيص الزائد (Over-Specialization)**: تقدم هذه التقنية مقترحات بالعناصر التي تكون درجة مطابقتها بالملف التعريفي للمستخدم عالية و بالتالي تكون المقترحات ذات درجة محدودة من الحداثة لان العناصر المقترحة تكون مماثلة لتلك التي تم تقييمها بالفعل، أي بصيغة أخري لا توجد توصيات تتجاوز ملف تعريف المستخدم.
3) **المستخدم الجديد (New User)**: و هي حالة من مشكلة البداية الباردة و ذلك حين يضاف مستخدم جديد للنظام ففي هذه الحالة لا تتوفر للنظام تقييمات كافية للمستخدم مما يجعل من الصعب معرفة تفضيلات المستخدم.

تطرقت بعض الدراسات إلي هذه التحديات منها (Degemmis, M. و آخرون، 2007) و (Degemmis, M. و آخرون، 2015) الذين عكفوا علي زيادة حدود المحتوي و ذلك عن طريق تفسير و إدارك معاني المفردات في ملف المستخدم التعريفي باستخدام مصادر معرفية مختلفة، أما (Singh, V. و آخرون، 2011) فقد استخدموا تقنيات التحليل العاطفي (Sentiment Analysis) لفهم و تفسير آراء المستخدم.

و للتغلب علي مشكلة التخصيص الزائد إستخدم (Adamopoulos, P., Tuzhilin, A. ، 2014) تقنية إختيار الجار الإحتمالي Probabilistic Neighborhood Selection Technique لإختيار المستخدمين المشابهين بدلاً عن إستخدام تقنية الجار الأقرب K-Nearest Neighbors، كذلك إستخدم (Son, J., Kim, S.، 2017) شبكة متعددة المسارات لربط العناصر ثم قياس أوجه الشبه بينها.

لجأت بعض الدراسات إلي إستخدام التقنية المختلطة لمعالجة مشكلة المستخدم الجديد فقد اقترح (Sun, D. و آخرون، 2011) إستخدام تقييمات

المستخدمين بالإضافة إلي معلومات المستخدم الديموغرافية، أما (Geetha,G. و آخرون، 2018) فقد إستخدموا معلومات مجموعات المستخدمين لإجراء تصفية تعاونية.

الجدول-3: تحديات تقنية التصفية بناءً علي المحتوي و بعض الحلول المقترحة و مصادرها

التحدي	الحلول المقترحة	المصدر
محدودية المحتوي	• دمج المعرفة اللغوية في عملية تعلم ملفات تعريف المستخدمين و ذلك من خلال تفسير و إدارك معاني المفردات في ملف المستخدم التعريفي باستخدام المعرفة المعجمية المخزنة في قاعدة البيانات المعجمية WordNet.	(Degemmis, M. و آخرون، 2007)
	• تصنيف معاني المفردات، إما من أعلي لأسفل و فيه يعتمد المعنى علي مصادر معرفة خارجية، أو من أسفل لأعلي و فيه يعتمد معني المفردات على استخدامها في مجموعة كبيرة من المستندات النصية.	(Degemmis, M. و آخرون، 2015)
	إستخدام تقنيات التحليل العاطفي (Sentiment Analysis) لفهم و تفسير آراء المستخدم.	(Singh, V. و آخرون، 2011)
التخصيص الزائد	• إستخدام تقنية إختيار الجار الإحتمالي (Probabilistic Neighborhood Selection Technique) لإختيار المستخدمين المشابهين بدلاً عن إستخدام تقنية الجار الأقرب (K-Nearest Neighbors) أو الأكثر تشابهاً ، بحيث يتم أخذ عينات مرجحة من المستخدمين المشابهين مع الأخذ في الإعتبار مستويات التشابه بين المستخدم المستهدف و جميع المستخدمين المشابهين المرشحين، و بالتالي تقديم مقترحات عن طريق إختيار مجموعات متنوعة من المستخدمين المتشابهين.	(Adamopoulos, P., Tuzhilin, A. ، 2014)
	• إستخدام شبكة متعددة المواصفات (multi attribute network) لتمثيل مواصفات العناصر تستخدم هذه الشبكة تقنيات المركزية (centrality) والتجميع (Clustering) لتمثيل العلاقات المشتركة بين العناصر، في تحليل هذه الشبكة يتم قياس أوجه التشابه بين العناصر المرتبطة بشكل مباشر و غير مباشر مما يضمن تنوع العناصر المقترحة.	(Son, J., Kim, S. ، 2017)
المستخدم الجديد	• إستخدام تقنية مختلطة حيث يتم تجميع المستخدمين استنادًا إلى اهتماماتهم ثم استخدام نتائج المجموعات و معلومات المستخدمين الديموغرافية (الجنس، العمر، المهنة) لإنشاء شجرة قرار لربط المستخدمين الجدد بالمستخدمين الحاليين ثم إجراء تصفية تعاونية.	(Sun, D. و آخرون، 2011)

• إستخدام تقنية مختلطة بالإضافة إلي تقنيات التجميع و التصنيف بحيث يتم تغذية مصفوفة بيانات التصفية التعاونية بالتقييمات الناتجة عن التقنية بناءً علي المحتوي.	(Geetha,G. و آخرون، 2018)

6- الخلاصة:

تساعد أنظمة التوصية المستخدمين في العثور على العناصر المهمة حسب اهتماماتهم و رغباتهم فهي تعمل علي تصفية البيانات باستخدام تقنيات تصفية البيانات مثل تقنية التصفية التعاونية و تقنية التصفية بناءً علي المحتوي، تستخدم كل تقنية طريقة مختلفة عن الأخري لتقديم المقترحات، لذلك كل تقنية تمتاز بمزايا تختلف عن الأخري، مثلاً تعتمد تقنية التصفية التعاونية علي تقييمات مستخدمي الجوار و بالتالي يمكنها تقديم مقترحات متنوعة، بينما تعتمد تقنية التصفية بناءً علي المحتوي علي مواصفات المحتوي مما يمكنّها من تقديم مقترحات للعناصر المضافة حديثاً للنظام. تناولت هذه الورقة التحديات التي تواجه كلا التقنيتين و قد استعرضت عدد من الدراسات التي اقترحت بعض الحلول لمواجهة تلك التحديات، الإ أن هذه الدراسات واجهت تحديات أخري مثل دقة و موثوقية المقترحات في حالة التجميع، و جودة البيانات المستخدمة لملء مصفوفة التقييمات، و بذلك تجلي مدي الحاجة لمزيد من البحث و التطوير في هذا الإتجاه و ذلك بغرض زيادة دقة و موثوقية نظم التوصية، و تقديم توصيات ذات علاقة بالمستخدم.

المصادر و المراجع

Abdollahpouri, H., Burke, R. and Mobasher, B. 2019. *Managing Popularity Bias in Recommender Systems with Personalized Re-ranking*, The Thirty-Second International Florida Artificial Intelligence Research Society Conference (FLAIRS-32).

Adamopoulos, P., and Tuzhilin, A. 2014. *On Over-Specialization and Concentration Bias of Recommendations: Probabilistic Neighborhood Selection in Collaborative Filtering Systems*. In 8th ACM Conference on Recommender Systems: RecSys 2014 Association for Computing Machinery (ACM). https://doi.org/10.1145/2645710.2645752.

Aggarwal,C. 2016. *Recommender Systems: The Textbook* .1st. ed. Springer Publishing Company, Incorporated.

Burke, R. 2002. *Hybrid Recommender Systems: Survey and Experiments*. User Modeling and User-Adapted Interaction. 12. 10.1023/A:1021240730564.

Çano, E. and Morisio, M. 2017. *Hybrid Recommender Systems: A Systematic Literature Review*. Intelligent Data Analysis. 21. 1487-1524. 10.3233/IDA-163209.

Cantador, I., Elliott, D. and Jose, J. 2009. *A Case Study of Exploiting Data Mining Techniques for an Industrial Recommender System*.

Cho,Y. , Kim,J. and Kim,S. 2002. *A personalized recommender system based on web usage mining and decision tree induction*. Expert Syst Appl, 23, pp. 329-342.

Covington, P., Adams, J. and Sargin, E. 2016. *Deep Neural Networks for YouTube Recommendations*. In Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16). Association for Computing Machinery, New York, NY, USA, 191–198. DOI:https://doi.org/10.1145/2959100.2959190.

Das, A., Datar, M., Garg, A. and Rajaram, S. 2007. *Google news personalization: scalable online collaborative filtering*, Proceedings of the 16th International Conference on World Wide Web (WWW '07) (New York), ACM Press, pp. 271–280.

Davidson, J., Liebald, B., Liu, J., Nandy,P., Vleet,T., Gargi,U., Gupta,S., He,Y., Lambert, M., Livingston, B. and Sampath, D. 2010. *The YouTube video recommendation system*. In Proceedings of the fourth ACM conference on Recommender systems (RecSys '10). Association for Computing Machinery, New York, NY, USA, 293–296. DOI:https://doi.org/10.1145/1864708.1864770.

Decampos, L., Fernández-Luna, J., Huete, J. and Rueda-Morales, M. 2010. *Combining content-based and collaborative*

Kim, M. and Im, I. 2014. *Resolving the 'Gray sheep' Problem Using Social Network Analysis (SNA) in Collaborative Filtering (CF) Recommender Systems*. Journal of Intelligence and Information Systems. 20. 137-148. 10.13088/jiis.2014.20.2.137\.

Konstan, J., Miller, B., Maltz, D., Herlocker, J., Gordon, L. and Riedl, J. 1997. *GroupLens: applying collaborative filtering to usenet news*. Communications of the ACM 40(3), 77–87.

Kula, M. 2015. *Metadata Embeddings for User and Item Cold-start Recommendations*. 1448.

Li, Q., Kim, B. 2003. *An approach for combining content-based and collaborative filters*, In Proc. of IRAL2003.

Lops, P., Musto, C., Narducci, F., Gemmis, M., Basile, P. and Semeraro, G. 2011. A *Semantic Content-based Recommender System for Cross-Language Personalization.*

Musto,C., Greco,C., Suglia,A. and Semeraro,G. 2016. *Ask me any rating: A content-based recommender system based on recurrent neural networks*. In Proceedings of the IIR.

Nguyen, S. 2019. *Model-Based Book Recommender Systems using Naïve Bayes enhanced with Optimal Feature Selection*. 217-222. 10.1145/3316615.3316727.

Pazzani, M., Muramatsu, J. and Billsus, D. 1996. *Syskill and Webert: Identifying Interesting Web Sites*. In: Proceedings of the Thirteenth National Conference on Artificial Intelligence and the Eighth Innovative Applications of Artificial Intelligence Conference, pp. 54–61. AAAI Press / MIT Press, Menlo Park.

Ricci, F., Rokach, L., Shapira, B. and B. Kantor, P. 2010. *Recommender Systems Handbook* (1st. ed.). Springer-Verlag, Berlin, Heidelberg.

Singh, V., Mukherjee, M. and Mehta, G. 2011. *Combining a Content Filtering Heuristic and Sentiment Analysis for Movie Recommendations*. 10.1007/978-3-642-22786-8_83.

Sinha,B., and Dhanalakshmi, R. 2019. *Evolution of recommender paradigm optimization over time*. ISSN 1319-1578, https://doi.org/10.1016/j.jksuci.2019.06.008. (http://www.sciencedirect.com/science/article/pii/S1319157819304963).

recommendations: a hybrid approach based on Bayesian networks. International Journal of Approximate Reasoning, 51(7), 785–799.

Degemmis, M., Lops, P. and Semeraro, G. 2007. *A content-collaborative recommender that exploits wordnet-based user profiles for neighborhood formation*. User Modeling and User-Adapted Interaction 17(3), 217–255.

Degemmis, M., Lops, P., Musto, C., Narducci, F. and Semeraro, G. 2015. *Semantics-Aware Content-Based Recommender Systems*. 10.1007/978-1-4899-7637-6_4.

Fazziki, A., Aissaoui, O., Alami, Y., Allioui, Y. and Benbrahim, M. 2019, *A new collaborative approach to solve the gray-sheep users problem in recommender systems*. Third International Conference on Intelligent Computing in Data Sciences (ICDS), Marrakech, Morocco, pp. 1-4.

Funk, S. 2006 .*Netflix Update: Try this at Home*. https://sifter.org/~simon/journal/20061211.html.

Gao, C., He, X., Gan, D., Chen, X., Feng, F., Li, Y., Chua, T. and Jin, D. 2018. *Learning Recommender Systems from Multi-Behavior Data.*

Gautam, A., Bedi, P., Singh, R. and Sharma, C. 2014. *Using Novelty Score of Unseen items to Handle Popularity Bias in Recommender Systems*. Proceedings of 2014 International Conference on Contemporary Computing and Informatics, IC3I 2014. 10.1109/IC3I.2014.7019608.

Geetha,G., Safa,M., Fancy, C. and Saranya, D. 2018.*A Hybrid Approach using Collaborative filtering and Content based Filtering for Recommender System*. Journal of Physics: Conference Series. 1000. 012101. 10.1088/1742-6596/1000/1/012101.

Hofmann, T., *Latent Semantic Models for Collaborative Filtering*, 2004, ACM Transactions on Information Systems 22 89–115.

Jannach, D., Zanker, M., Felfernig, A. and Friedrich G .2010. *Recommender Systems – An Introduction*. Cambridge University Press.

Jayech, K. and Mahjoub, M. 2012. *New approach using Bayesian Network to improve content based image classification systems*. International Journal of Computer Science Issues. 7.

Son, J., and Kim, S. 2017. *Content-based filtering for recommendation systems using multiattribute networks*. Expert Systems with Applications, 89, 404-412. https://doi.org/10.1016/j.eswa.2017.08.008.

Sun, D., Li, C. and Luo, Z. 2011. *A content-enhanced approach for cold-start problem in collaborative filtering*. 10.1109/AIMSEC.2011.6010230.

Wang, B., Ye, F. and Xu, J. 2018. *A Personalized Recommendation Algorithm Based on the User's Implicit Feedback in E-Commerce*. Future Internet 10: 117.

Xue, G., Lin, C., Yang, Q., Xi, W., Zeng, H., Yu, Y. and Chen, Z. 2005. *Scalable collaborative filtering using cluster-based smoothing*. In Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '05). Association for Computing Machinery, New York, NY, USA, 114–121. DOI:https://doi.org/10.1145/1076034.1076056.