## Summary of the Paper: "Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method"

* **مشكلة البحث:** تقدم المواقع الإلكترونية عادةً العديد من الخيارات أمام المستخدمين ، مما يجعل من الصعب و يستغرق وقتًا طويلاً العثور على المنتج أو المعلومات الصحيحة.
* **الهدف:** تطوير نظام توصية باستخدام أقرب جار (KNN) بناءً على بيانات نقرات المستخدم من موقع ويب لقارئ RSS ) Really Simple Syndication or Rich Site Summary.
* **الطريقة:** يتم تدريب طريقة تصنيف KNN لتحليل بيانات نقرات المستخدم في الوقت الفعلي ، وتصنيفها إلى مجموعات مستخدمين وتقديم توصيات مصممة خصيصًا لهم.
* **الخطوات:**
  + استخراج وتنظيف بيانات عناوين RSS للمستخدمين وتنسيقها.
  + تجميع البيانات في جلسات ذات مغزى وإنشاء مستود بيانات.
* **النتائج:** يعد تصنيف KNN بسيطًا ومتسقًا وسهل التنفيذ ، مما يجعله مناسبًا للتوصيات في الوقت الفعلي مع القليل من المعرفة المسبقة حول توزيع البيانات.
* **معايير الجودة والكفاءة:** يتم التقييم بناءً على المحتوى والعرض وسهولة الاستخدام والوصول إلى المعلومات المطلوبة ووقت انتظار المستخدم.
* **النهج:**
  + تصنيف المستخدمين بناءً على بيانات النقر الحالية ومطابقتها بمجموعات المستخدمين.
  + تقديم خيارات تصفح مصممة خصيصًا باستخدام استخراج بيانات استخدام الويب.
* **البيانات:**
  + تم استخدام قاعدة بيانات وصول مستخدمي RSS على مدار شهرين.
  + استخراج البيانات ومعالجتها مسبقًا وتجميعها في جلسات لإنشاء مستود بيانات.
* **الأسلوب:**
  + تم استخدام تصنيف أقرب جار لتحليل معلومات عنوان URL من قاعدة بيانات مستخدمي RSS.
  + تم التقييم باستخدام جلسات اختبار نموذجية ، وعرض النتائج والتحليل.
* **النتائج:**
  + يقدم نموذج KNN توصيات دقيقة ، ويتفوق على خوارزميات التصنيف الأخرى.
  + حقق معدل دقة يزيد عن 70٪ ، مما يعني أن أكثر من 70٪ من الأخبار الموصى بها تتماشى مع متطلبات المستخدم الفورية.
  + يحسن من تصفح المستخدم من خلال تقديم معلومات ذات صلة دون طلبات صريحة ، مما يحسن تجربة المستخدم بشكل عام.

**الآثار المترتبة:**

* يمكن أن تساعد نتائج الدراسة مصممي ومشرفي المواقع الإلكترونية في التخطيط للترقيات لتحسين التنقل في الموقع وتلبية احتياجات المستخدمين بشكل أكثر كفاءة.
* يمكن أن يؤدي اعتماد نموذج KNN إلى تحسين كبير في دقة نظام التوصية وجودته ، مما يؤدي إلى تجربة ويب أكثر جاذبية وسهولة في الاستخدام.

Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method

## ملخص وتنظيم البحث

#### العنوان

استخدام K-Means و أقرب جار (KNN) في استخراج بيانات استخدام مواقع الأفلام ونظام التوصية

#### الملخص

* **الهدف:** بناء نظام توصية للأفلام باستخدام خوارزميتي K-Means و أقرب جار (KNN).
* **مجموعة البيانات:** مجموعة بيانات MovieLens من موقع Kaggle.
* **المكونات:** مفاهيم التعلم الآلي وأنظمة التوصية والخوارزميات (K-Means، KNN، التصفية التعاونية، التصفية القائمة على المحتوى).
* **التقييم:** يتم قياس أداء النظام باستخدام جذر متوسط مربع الخطأ (RMSE)، حيث تم تحقيق أفضل قيمة RMSE وهي 1.081648.

#### مسار المعالجة

1. **إدخال المستخدم:** يوفر المستخدم معرف المستخدم والنوع الاجتماعي والعمر ورمز PIN.
2. **معالجة البيانات مسبقًا:** تتم معالجة البيانات الخام مسبقًا وتحويلها إلى إطارات بيانات باستخدام NumPy و Pandas.
3. **التجميع:** يتم استخدام طريقة مجموع المربعات المجمعة ضمن المجموعة لتحديد العدد الأمثل للمجموعات لتجميع K-means.
4. **مصفوفة الفائدة:** يتم بناء مصفوفة فائدة مجمعة، والتي تحدد متوسط التقييم الذي يمنحه المستخدمون لكل مجموعة.
5. **حساب التشابه:** يتم استخدام معامل ارتباط بيرسون لحساب التشابه بين المستخدمين بناءً على مصفوفة الفائدة.
6. **التنبؤ:** يستخدم KNN مصفوفة الفائدة المجمعة والتشابه للتنبؤ بتوصيات الأفلام للمستخدم الذي أدخل بياناته.

#### خوارزمية النظام المقترح

1. **استيراد المكتبات:** استيراد المكتبات الضرورية (NumPy، Pandas، Matplotlib، sklearn).
2. **قراءة البيانات:** قراءة بيانات CSV إلى إطارات بيانات.
3. **تقسيم البيانات:** تقسيم البيانات إلى مجموعات تدريب واختبار.
4. **إنشاء مصفوفة الفائدة:** إنشاء مصفوفة فائدة تشير إلى المستخدم الذي قام بتقييم كل فيلم.
5. **تحديد العنقود:** استخدام طريقة WCSS لتحديد العدد الصحيح للمجموعات.
6. **مصفوفة الفائدة المجمعة:** تحديد مصفوفة الفائدة المجمعة بعد تطبيق تجميع K-means.
7. **مصفوفة التشابه:** تطبيق معامل ارتباط بيرسون على مصفوفة الفائدة لحساب تشابه المستخدم.
8. **التطبيع:** تطبيع القيم في مصفوفة الفائدة.
9. **دالة التنبؤ:** تقوم دالة التخمين (Guess()) بالتنبؤ بتقييمات الأفلام لأفضل N مستخدم مشابه.
10. **المقارنة:** مقارنة التقييمات المتوقعة بالتقييمات الفعلية في مجموعة الاختبار.
11. **التقييم:** حساب جذر متوسط مربع الخطأ (RMSE) لتقييم دقة النموذج.

#### الخاتمة

* **تعريف التعلم الآلي:** طريقة تحليل البيانات التي تعمل على أتمتة بناء النماذج التحليلية، مما يتيح للأنظمة التعلم من البيانات واتخاذ القرارات بأقل تدخل بشري.
* **النظام المقترح:** نظام توصية للأفلام باستخدام K-Means و أقرب جار (KNN)، تم تنفيذه باستخدام لغة البرمجة Python ومجموعة بيانات MovieLens.
* **الأداء:** يحقق جذر متوسط مربع خطأ (RMSE) أفضل مقارنة بالتقنيات الموجودة، مع عدد أقل من العناقيد.
* **التحسينات المستقبلية:** إمكانية التحسين الإضافي عن طريق تحسين عملية التجميع وتضمين تقنيات إضافية.

**SVD-based Incremental Approaches for Recommender Systems**

## ملخص وتنظيم البحث

#### العنوان

استخدام تحلل قيمة مفردة (SVD) الب bertaئي لأنظمة التوصية

#### الملخص

* **المشكلة:** تتطلب كمية المعلومات الهائلة على الإنترنت أنظمة توصية فعالة لتوفير محتوى مخصص.
* **الحل:** اقتراح خوارزمية تدريجية تسمى Incremental ApproSVD، تجمع بين Incremental SVD و ApproSVD، للمعالجة القابلة للتطوير والفعالة.
* **التقييم:** تمت مقارنة الخوارزمية مع Incremental SVD على مجموعات بيانات MovieLens و Flixster، مما يدل على أداء متفوق في دقة التنبؤ ووقت التشغيل.

#### المساهمات

1. **خوارزمية تدريجية:** تقترح خوارزمية Incremental ApproSVD للتنبؤ الديناميكي بالتصنيفات غير المعروفة مع إضافة عناصر جديدة.
2. **القابلية للتطوير والكفاءة:** الخوارزمية سهلة التنفيذ وتوفر وقت تشغيل أقل مقارنة بـ Incremental SVD.
3. **تحليل الخطأ:** يوفر حداً أعلى للخطأ بين التصنيفات الفعلية والمتوقعة، مما يدل على فعالية الخوارزمية.
4. **التحقق التجريبي:** تظهر التجارب مزايا الخوارزمية المقترحة على مجموعات بيانات MovieLens و Flixster تحت معايير مختلفة.

#### المنهجية

1. **تحويل البيانات:** تحويل بيانات تصنيف المستخدم-الفيلم الثابتة إلى مصفوفة تصنيف المستخدم-الفيلم B1 وتحويل البيانات الجديدة إلى مصفوفة B2.
2. **عينة المصفوفة:** استخراج عدد ثابت من الأعمدة من B1 و B2 لتكوين مصفوفات أصغر C1 و C2.
3. **تحلل قيمة مفردة الب bertaئي:** يحسب تحلل قيمة مفردة لـ [C1، C2] بناءً على تحلل قيمة مفردة لـ C1.
4. **التنبؤ:** يستخدم متجهات تفضيلات المستخدم ومتجهات ميزات الفيلم للتنبؤ بالتصنيفات. يوصى بالأفلام ذات التصنيفات المتوقعة في النصف العلوي من النطاق.

#### منصة تجريبية

1. **مجموعات البيانات:**
   * **MovieLens:** 943 مستخدمًا و 1682 فيلمًا بـ 100000 تصنيف.
   * **Flixster:** 8465 مستخدمًا و 9602 فيلمًا بـ 5326788 تصنيفًا.
2. **مقياس التقييم:** متوسط الخطأ المطلق (MAE) وجذر متوسط مربع الخطأ (RMSE).
3. **التحقق المتقاطع الخماسي:** تقسم التصنيفات إلى خمسة أقسام للتدريب والاختبار.

#### النتائج والمناقشات

* **المقارنة:** تتفوق Incremental ApproSVD على Incremental SVD في كل من دقة التنبؤ ووقت التشغيل على مجموعات بيانات MovieLens و Flixster.
* **تحليل الخطأ:** يوفر التحليل الرياضي حداً أعلى للخطأ، مما يضمن قوة الخوارزمية.

#### الخاتمة

* **خوارزمية تحلل قيمة مفردة الب bertaئي:** تستغل قيم وموحدات القيمة المفردة السابقة للتحديث الفعال.
* **خوارزمية Incremental ApproSVD:** تجمع بين ApproSVD مع Incremental SVD لتحسين الأداء.
* **التحليل الرياضي:** يثبت دقة التوقعات.
* **التحقق التجريبي:** يوضح تفوق Incremental ApproSVD على Incremental SVD في دمج دقة التنبؤ ووقت التشغيل على مجموعتي بيانات كبيرتين.

تقدم هذه الورقة طريقة قوية وقابلة للتطوير وفعالة لتحسين أداء أنظمة التوصية في سياق البيانات الضخمة، حيث تلبي الحاجة إلى توصيات دقيقة وفي الوقت المناسب في مشهد معلومات دائم التطور.