

الخلاصة

كمية المعلومات في العالم حاليا تزداد بسرعة هائلة وأسرع من قابلية معالجتها . جميعا يدرك الشعور عند صدور كتب جديدة، مقالات صحفية و مؤتمرات تنعقد سنويا لهذا نحتاج تكنولوجيا تساعد على تقليص الحواجز والإعلان عن مثل هذه المعلومات . والوقت الآن مهم لخلق تكنولوجيا تساعدنا على اختيار مثل هذه المعلومات المهمة من بين كم هائل من المعلومات. واحد هذه الطرق التي تساعد هو الترشيح المتعاون . واهم خوارزمية له هي الذاكرة المعتمدة الترشيح المتعاون والتي تتنبأ بأفضلية المستخدم اعتمادا على تشابه مع مستخدم آخر في قاعدة البيانات. من ناحية ثانية أدى النمو الهائل للمستخدمين والعدد الكبير في المنتجات إلى مشاكل في خوارزمية الذاكرة المعتمدة على الترشيح المتعاون وتتلخص تحديد الأشخاص (الطلبات) الصحيحة للاستخدام وذلك لتقليل كلفة التنفيذ والخرن الزائد وغالبا لتحسين الدقة بشكل عام وذلك بتجنب الرديء وغير الملائم ويتم ذلك باستخدام الاختيار اللحظي الذي يلعب دور مهم في استخلاص البيانات. حددنا عملنا هنا في اختيار الطلبات (المستخدمين) من نوعيه عاليه وتجنب الطلبات من نوعيه واطئه من قاعدة بيانات الأسواق التي استخدمناها. طبق جزء التقليل من ذاكره لمعتمدة لترشيح متعاون. واستخدم لهذا الجزء من الخوارزمية الاختيار اللحظي ليعمل على تقليص قاعدة البيانات . في خوارزمتنا هناك قيمة تحديد (ك) والتي تعتبر خطوه مهمة في خوارزمية التقليل ومن تجاربنا وجدنا إن أفضل القيم (ك)

تتراوح بين (0.046 - 0.048).

Abstract

The amount of information in the world is increasing far more quickly than our ability to process it. All of us have known the feeling of being overwhelmed by the number of new books, journal articles, and conference proceedings coming out each year. Technology has dramatically reduced the barriers to publishing and distributing information. Now it is time to create the technologies that can help us sift through all the available information to find, that which is most valuable to us. One of the most promising such technologies is collaborative filtering.

One popular algorithm is the Memory-Based Collaborative Filtering, which predicts a user preference based on his or her similarity to other user (instance) in the database. However, because of the tremendous growth of users and the large number of products, Memory-Based CF algorithms have problem in deciding the right instances to use. In order to reduce executive cost and excessive storage, and possibly to improve the generalization accuracy by avoiding noise and overfitting, instance selection is used. It plays an important role in data mining application. We focus our work here on selecting instances (users) with high quality and removing instances with low quality from our DB of supermarket we used. The part of reduction for Memory-Based Collaborative Filtering is applied, and for this part of the algorithm we used instance selection to make reduction for the database. In our algorithm we have threshold value K , which is used to decide the reduction of users. After the experiment we can decide that the best value K is between (0.046-0.048).