عنوان الرسالة: حل مشكلة البداية الباردة في أنظمة الاقتراح باستخدام معلومات سياق باللغة العربية

اسم الطالب: نهى سعيد أحمد الغامدي

اسم المشرف: د. فاطمة عسيري

المستخلص:

إحدى التحديات في أنظمة الاقتراح هي مشكلة البداية الباردة، وتعني عدم قدرة نظام الاقتراح على صنع مقترحات دقيقة نظرًا لعدم توافر معلومات كافية عن تفضيلات المستخدم. نتيجة لذلك، تم تكييف معلومات السياق واستخدامها لحل مشكلة البداية الباردة عند تسجيل أي مستخدم جديد في النظام عبر اقتراح منتجات أو خدمات اعتمادًا على المعلومات التي تصف حال المستخدم.

طريقتنا المقترحة لحل هذه المشكلة هي عبر استغلال خدمة التقويم. يمكن الوصول إلى التقاويم لمعرفة اهتمام المستخدم من أحداثه المسجلة، ومن ثم اقتراح منتجات لها علاقة بالأحداث القادمة. تركيزنا يصب على تحليل النصوص العربية داخل التقويم لأن اللغة العربية تقبع في المركز السادس من حيث عدد المتحدثين بها كلغة أولى. لذلك، بنينا مصنف نصي متعدد الفئات لتصنيف أحداث التقويم، ثم تحققنا من صحة المصنف عبر التقييم من حيث التحزيم المتشابك. وصلت دقة مصنفنا إلى ٨٨،١٠% وقيمة ف١ وصلت إلى ٨٨،٣٨%. في الخطوة التالية، قمنا باختبار المصنف على بيانات حقيقية من تقاويم مستخدمين متعددين ووصلت الدقة إلى ٧٦،٧٢%. قمنا بالتحقق من اسباب هذه النتيجة، ووجدنا عدة مؤثرات منها: كتابة كلمات إنجليزية بحروف عربية، والكلمات العامية. أخيرًا، قمنا بتسليط الضوء على عدة اتجاهات بحثية لحل القصور في هذا البحث.

Title: Solving Cold Start Problem in Recommender Systems Using Arabic Contextual Information

Student name: Nuha Saeed Ahmed Alghamdi

Supervisor: Dr. Fatmah Assiri

Abstract:

One of the challenges in recommender systems is cold start problem, which is the inability of recommender system to make accurate recommendations due to unavailability of enough information about user's preferences. However, context information is adapted to solve cold start problem when new user is using the system by recommending items or products based on information defining a user situation.

Our proposed method to solve this problem is through the utilization of calendars. They can be accessed to find user's interest based on registered events, then products that are related to the upcoming events are recommended to the user. We focus on analyzing Arabic text since Arabic language is the sixth most spoken language in the world. We built a multi-class text classifier to classify calendar's events, the classifier trained on a dataset that is collected from Wikipedia, and we validated our classifier using cross validation. Our classifier reached an accuracy of 88.10% and an F1 score of 88.38%. Testing the classifier on actual data reached an accuracy of 76.72%. We investigated the reasons for such a result, and we found many factors that impacted them including, but not limited to, English events that are written by Arabic letters and names. We highlighted some research directions to tackle existing limitations.