



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
Departamento de Ciencias de la Computación
IIC3633 - Sistemas Recomendadores
Segundo Semestre 2018

Review 6

Content-Based Recommendation Systems

Genaro Laymuns

“Dado que generalmente hay más items disponibles en una base de datos de los que se pueden mostrar en una página web, es necesario seleccionar un subconjunto de dichos items para mostrarle al usuario.”

Esta cita, extraída de la introducción del paper, es la principal motivación que hay detrás de un sistema recomendador: determinar el subconjunto óptimo de items que deben mostrarse a un usuario al ingresar a dicha página web. La manera más básica de recomendar items a usuarios consiste en generar *User Profiles*, que preguntan directamente a un usuario sus preferencias para mostrar los items que sean de su propio interés, o bien recomiendan items a usuarios siguiendo reglas básicas (como por ejemplo recomendar la secuela de un libro comprado).

La segunda parte del paper nos habla sobre el proceso de aprendizaje que se debe hacer para generar el perfil de un usuario. En primer lugar se necesita tener una base de datos bien estructurada para entrenar y testear distintos sistemas recomendadores, por lo que es importante controlar la manera en que adquieren nuevos datos. Dependiendo del tipo de interacción entre el sistema y el usuario, la forma de llenar esta base de datos puede ser explícita (obteniendo información de parte directa del usuario), o bien puede ser obtenida haciendo un análisis del comportamiento del usuario. Si bien la información explícita es la fuente más confiable de las preferencias de un usuario, muchos usuarios no están dispuestos a entregar dicha información. En algunos casos es imposible estudiar el comportamiento del usuario, como el caso de “MyBestBets”, donde la página web solo consiste de un despliegue de información y el consumo de los productos no se realiza a través de dicha página web.

El segundo paso consiste en procesar dicha información para generar las recomendaciones. En particular me gustaría hablar del algoritmo *Relevance Feedback and Rocchio’s Algorithm*. Este algoritmo tiene un vector de preferencias de un usuario y en cada iteración este vector es actualizado dependiendo del vector de preferencias anterior más una bonificación con respecto a los items

relacionados con el usuario, y luego restándole puntaje si el ítem no estaba relacionado con el usuario. Siento que este modelo no logra captar bien el comportamiento real de un usuario, dado que existe todo un espectro de gustos posibles, y por esto las bonificaciones y las penalizaciones óptimas obviamente no deben ser iguales para cada uno de los ítems, sino que podrían depender, por ejemplo, de qué tan efectivo haya sido incorporar dicha evaluación al modelo.