



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

Review 7

Factorization Machines

Genaro Laymuns

En sistemas recomendadores se tiene una matriz de relaciones usuario-ítem que suele tener una estructura muy sparse. Para estimar la matriz de relaciones usuario-ítem existen distintos algoritmos que requieren distinta complejidad computacional para ser implementados. Uno de los algoritmos más utilizados corresponde a “Support Vector Machine” (SVM), que consiste simplemente en suponer que los datos son generados a partir de un polinomio de orden d (dependiendo de cada problema), donde las variables de dicho polinomio corresponden a la información relacionada a cada usuario.

Dado el modelo de SVM, está comprobado que la complejidad de dichos métodos es del orden de $O(kn^d)$, donde k es una constante que depende de la densidad del dataset. La complejidad no lineal que se obtiene de forma natural en SVM implica que se tengan grandes limitaciones con respecto al tiempo requerido para estimar dichos coeficientes. Dada la gran cantidad de datos que generalmente hay en dichos datasets, esta limitación de tiempo restringe ampliamente a los algoritmos que se puedan utilizar para estimar dichos coeficientes. Esta limitación es lo que motiva principalmente el estudio de “Factorization Machines” (FM). Estos modelos estiman los coeficientes de dicho polinomio a partir de productos internos entre vectores $v_i \in \mathbb{R}^k$. Además de las ventajas algorítmicas de FM con respecto a SVM, FM permite entender de una manera más intuitiva la relación entre las preferencias entre distintos usuarios, viéndolo como el producto interno (altamente relacionado con la semejanza) entre el vector que representa a cada usuario.

Los modelos basados en FM presentan dos grandes ventajas con respecto a los SVM. La primera de ellas es que funcionan mucho mejor en datasets sparse (esto se ejemplifica en el mismo paper con un ejemplo en que el dataset tiene solo un dato). La segunda gran ventaja es que, para estimar los coeficientes de dicho polinomio la complejidad es de $O(kn)$, es decir, corresponde a un algoritmo con complejidad lineal, lo cual abre las puertas a muchos algoritmos de *online learning*.

En el paper solo hablan de que la mejora en la complejidad algorítmica abre las puertas para

utilizar el método del gradiente estocástico en el contexto de *online learning* con distintas funciones de pérdida, pero el hecho de que la complejidad sea lineal permite actualizar los coeficientes del polinomio utilizando distintos esquemas de trabajo. Un ejemplo es representar los coeficientes como elementos de una serie de tiempo y así analizar la evolución de los gustos de los usuarios, en el caso de que los ratings contengan el *timestamp* de cuando fueron realizados. Esto nos da la gran ventaja de que existe una gran cantidad de métodos altamente eficientes en el contexto de *online learning* para estimar el valor de distintas series de tiempo.