Discussion_1

Kumudini Bhave

June 15, 2017

Contents

1 Linked-In Jobs Recommender System

1.1 Summary

Linked-in is worlds largest professional network and is available to registered users only.

With terabytes of data flowing through their systems, generated from member's profile, their connections and their activity on LinkedIn, they have a huge semi-structured data that gets updated real time and growing at a massive pace. Linked-In has used this rich information to provide all kinds of recommendations, be it, members, jobs, groups, companies, articles etc.

1.2 Approach

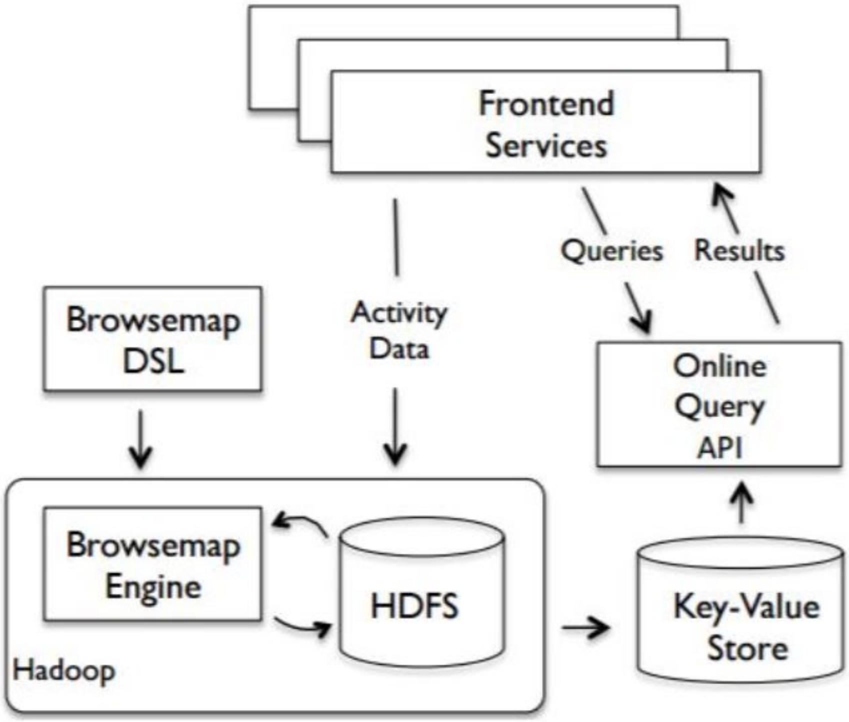
1.2.1 Browsemaps: Collaborative Filtering at LinkedIn

LinkedIn's uses horizontal collaborative filtering infrastructure, known as browsemaps, a hybrid of offline/online system; the system computes a latent co-occurrence graph in batch and serves results to users with low-latency . It is basically item-to-item collaborative filtering that is used for people, job, company, group,and other entity recommendations, where member's browsing histories are used to build a latent graph of co-occurrences of entities.

We would discuss here the Linked-In Jobs Recommender system.

1.2.2 Linkedin Jobs Recommendations

The recsys (Recommender System) takes care of the expired jobs by removing it from browsemaps. The figure below illustrates the Browsemap architecture.



The offline system uses Hadoop for its batch computation engine. Computed browsemaps are bulk loaded into a distributed key-value store, which provides low-latency queries.

Figure 1:

Recruiter Perspective : When a jobs is posted by a recruiter, the linkedin job recsys produces the matches to the jobs poster in realtime , and here it uses content analysis to produce the strong match. 'Similar Profiles' is a section for hiring managers/recruiters who are actively looking for candidates. Here, if one finds a candidate recsys suggests other candidates like the original one in terms of overall experience, specialty, education background and a host of other features.

Applicant/Member Perspective : There is content overlap in 'Similar Jobs' recommendations. while collaborative filtering applied to the section that produces 'People Also Viewed' that tries to match similar member profiles and the job applications/ job views of such members. The job entity has two two types of event , the job view and job apply; here the emphasis is on job apply. When pushing recommendations , for a non-active member, the recsys uses a host of features viz.

- User provided features like 'Title, Specialty, Education, experience amongst others'
- Complex derived features like 'Seniority' and 'Skills', computed using machine learnt classifiers.
- Also features like 'Related Titles' and 'Related Companies'

1.2.3 The Problem Area

The common problem inherent with collaborative filtering is of *cold start*. When a job is posted / new member registers there is not much activity on the new entities. This leads to sparsity and to deal with that (to some extent) , the members browsing history is used to personalise the backfill of any sparse entity.

For eg. For a member who has viewed several jobs, but then lands on a newly posted job with only minimal activity and thus a sparse browsemap. To combat this, the online system surfaces the browsemaps from the jobs the member has previously viewed merged through a reduction function.

Browsemap computation, as any collaborative filtering recommendation, relies solely on user activities and is thus extremely sensitive to the quality and quantity of input data