Case Study: A New Direction for Personals

Using mathematical proximity algorithms to determine "best match" in real time

OVERVIEW

A major player in the online dating space was looking to develop a personals application that focused on free text matching. The trick was to find a way to determine which users were compatible based on highly variable data, and to do it in real-time.

- Implemented and refined complex statistical matching algorithms to determine how textual inputs relate to each other, and to predict how individual users feel about each other
- Developed real-time map-reduce architecture that allowed non-trivial algebraic computation over the dataset as users type
- Designed, deployed, and managed high-capacity sharded infrastructure to host the application
- Ongoing data mining and analytics to keep model up to date and to shape output around real-life distributions

THE CHALLENGE

This was a new kind of personals application, one that didn't focus on pre-set attributes but instead matched people as they typed and chose their own way of expressing themselves. This posed a unique set of challenges:

- Examining the published research on proximity matching, and adapting them based on how they perform in the real-world and in a new context
- Finding ways to continuously search the entire database of users over thousands of discrete attributes and return results in a fraction of a second
- Ensure performance over time, both in the responsiveness of the application and in the accuracy of the mathematical prediction models

Thumbtack was called upon because of its experience in building highly available systems as well as its expertise in statistical analysis. The team included a cadre of engineers and a statistics PhD to build and maintain the application and models.

THE SOLUTION

Thumbtacks solution involved engineering, hosting, and statistical analysis.

Statistician analyzed behavior of algorithm over time, the distribution of data as it arrives, and provided refinement of prediction models as data is gathered.

Engineering team built a dynamic web application plus matrix generation and computation clusters.

Deployment team tuned and hosted end-to-end system for zero-downtime deployments, as well as monitoring and metrics for algorithmic integrity.

COMPONENTS

The application had components to perform a variety of tasks.

- Dynamic web application, user controls, and administrative consoles.
- Custom in-memory map-reduce cluster to perform algebraic manipulation in real-time.
- Social aggregation services to provide rich Facebook interaction & incorporating "friend of friend" searches.
- Notification services to control and throttle message events across the social graph.

RESULTS

The application is currently live and can handle hundreds of thousands of pages per node per hour, each of which performs multiple scans over the computation cluster. The prediction model is deployed and producing common sense suggestions to new users, and metrics indicate a steady increase in prediction quality.

INTERESTED IN MORE?

For more information on Thumbtack's services or the kinds of engagement models offer, please contact us at:

Thumbtack Technology
68 Jay Street, Suite 201
Brooklyn, NY 11201
(877) 433-3835
sales@thumbtack.net