business sensitivity, we omit some details regarding to feature computation. Instead, we give a high-level description the key feature categories.

Expertise scores For every search result, we compute sum of his or her expertise scores, which are described in Section 2.1, on the skills in the query. The higher the sum is, the better the result matches the information need in terms of skills.

Textual features These features capture the textual similarities between the query and different sections of result profiles, such as, current titles, past titles, current companies, past companies, etc.

Geographic features Talent search on LinkedIn is highly personalized. For instance, given the same query, recruiters in New York City and in Montreal are interested in very different results. Thus, location plays an important role in personalizing search results. We create multiple features capturing this.

Social features Another important aspect of personalization is to capture how the results socially relate to the searcher. We leverage a variety of the signals on LinkedIn, such as common friends, companies, groups and schools to generate features in this category.

Interested reader could find more details on how the personalized ranking is trained in our recent work [2].

4.2 Career Trajectory Similarity

Different from keyword based matching methodology, Career Trajectory Similarity (CareerSim) ascertains a similarity between two profiles by leveraging the trajectory information encoded in series of positions held by the individuals through their careers. Due to the space limitation, we focus on the two key steps of the CareerSim framework here.

Profile Modeling To capture the trajectory information, CareerSim models every individual member profile as a sequence of nodes, each of which records all information within a particular position of member's career, such as company, title, industry, time duration, and keyword summary.

Similarity Computing At the node (position) level, similarity is ascertained by using a generalized linear model but other approaches could be easily substituted. Then at the sequence (profile) level, sequence alignment method is employed to find an optimal alignment between pairs of nodes from the two career paths.

To the best of our knowledge, CareerSim is the first framework to model professional similarity between two people taking into account their career trajectory information. More details about the CareerSim framework can be found in our recent paper [6]. Given a set of ideal candidates IC, the similarity between a result r and the candidate set is simply the average over the individual ones as shown in Equation 3. We posit that using the temporal and structural features of a career trajectory for modeling similarity between a result and the ideal candidates provides a good complement to the signals in the personalized ranker. This also gives a direct similarity between a result and ideal candidates.

$$f_2(r, IC) = \frac{\sum_{c \in IC} CareerSim(r, c)}{|IC|} \tag{3}$$

4.3 Search Ranking Demo

Following the scenario described in Section 3.3, after the query builder constructs a query from input profiles, the query is used to retrieve matched results. Then, the personalized ranker takes the query and the searcher's information into account to score each of the matched results. At the same time, the ideal candidates (Satya and Ryan) are also treated as another input for the CareerSim model to obtain a score measuring career trajectory similarity between each result and the candidates. The final ranking results based on a combination of the two scores are shown in the low-right section of Figure 3. As illustrated in the figure, the top results are similar to the ideal candidates. They are all software engineers with the same seniority level ("staff") and from the same or similar companies ("LinkedIn", "Google" or "Twitter").

During the search section, the searcher is still allowed to interact with the query via deleting or adding any of the entities. After every query edition, the retrieval system and the ranker are triggered and the result ranking is refreshed.

5. CONCLUSIONS

In this work, we present the next generation of talent search at LinkedIn: Search by Ideal Candidates. In this new search paradigm, instead of constructing a highly complicated query describing criteria of a hiring position, the searcher simply inputs one or a few ideal candidates for the position, e.g., existing members in the team. Our system will automatically build a query from the ideal candidates and then retrieve and rank results.

For query building, we present approaches based on collaborative filtering to generate skill and company facets. For result ranking, we propose a ranking function combining a personalized search ranker and a career trajectory similarity model. As of this writing, the product is being launched to a set of pilot customers and it will be ramped to other customers early 2016.

6. REFERENCES

[1] L. Ballerini, X. Li, R. B. Fisher, and J. Rees. A query-by-example content-based image retrieval system of non-melanoma skin lesions. In *Medical content-based retrieval for clinical decision support*, pages 31–38. Springer, 2010.

[2] V. Ha-Thuc, G. Venkataraman, M. Rodriguez, S. Sinha, S. Sundaram, and L. Guo. Personalized expertise search at linkedin. In *Proceedings of the 4th IEEE International Conference Big Data (BigData)*. IEEE, 2015.

[3] J. J. Rocchio. Relevance feedback in information retrieval. In *The SMART Retrieval System*. Prentice-Hall, Englewood Cliffs NJ, 1971.

[4] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pages 285–295. ACM, 2001.

[5] L. Wu, S. Shah, S. Choi, M. Tiwari, and C. Posse. The browsemaps: Collaborative filtering at linkedin. In *RSWeb*. ACM, 2014.

[6] Y. Xu, Z. Li, A. Gupta, A. Bugdayci, and A. Bhasin. Modeling professional similarity by mining professional career trajectories. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data mining (KDD)*. ACM, 2014.