





























A RNN query returns all objects whose  $k$  nearest neighbors contain the query object. The classical motivating example [12] of RNN is the decision support task of identifying the optimal location for a new store. Given several location choices, the strategy is to pick the location that can attract the maximum number of customers. A RNN query returns the customers who are likely to choose the new store because of its geographical proximity over the existing stores. There are several approaches for the implementation of RNN queries [13], [14]. There are also algorithms that work for higher dimensionality data sets [15]. What we suggest in our work is quite different from the similarity introduced and utilized in RNN queries, where the results are based solely on objects' characteristics while users' preferences are not taken into account.

Item-based collaborative filtering techniques [16] may share a similar intuition, but in contrary to our methods, they suggest that customers have a *taste* of some products and thus rate them. Suppose, the case of a product that recently launched in the market or a product is under designing during its manufacturing process. In both cases, there would be no ratings expressing customers' opinions, making a collaborative filtering algorithm inapplicable. On the other hand, our framework does not require any previous knowledge about users' opinions for the products because they express in a more general way their preferences providing a weighting factor for each attribute of products, which is different than rating individual products.

## 9 CONCLUSION

In this paper we introduced a user-centric similarity framework in which the similarity of products is assessed by taking into account user preferences. We demonstrated via examples and through our experiments that user-centric similarity search can yield quite different results than using conventional metrics that only look at the products, in isolation to the preferences their customers have expressed. We identified two interesting query types and we proposed efficient algorithms for their execution. We also discussed optimizations that help reduce execution times.

## REFERENCES

[1] A. Rajaraman and J. D. Ullman, *Mining of massive datasets*. Cambridge: Cambridge Univ. Press, 2012.

[2] K. Georgoulas and Y. Kotidis, "Towards Enabling Outlier Detection in Large, High Dimensional Data Warehouses," in *Proceedings of SSDBM*, 2012, pp. 591–594.

[3] H. Becker, M. Naaman, and L. Gravano, "Learning similarity metrics for event identification in social media," in *Proceedings of WSDM*, 2010.

[4] A. Vlachou, C. Doulkeridis, Y. Kotidis, and K. Nørnvåg, "Reverse Top-k Queries," in *Proceedings of ICDE*, 2010.

[5] A. Guttman, "R-Trees: A Dynamic Index Structure for Spatial Searching," in *Proceedings of SIGMOD*, 1984, pp. 47–57.

[6] V. Hristidis, N. Koudas, and Y. Papakonstantinou, "PREFER: A System for the Efficient Execution of Multi-parametric Ranked Queries," in *Proceedings of SIGMOD*, 2001, pp. 259–270.

[7] N. Roussopoulos, S. Kelley, and F. Vincent, "Nearest Neighbor queries," in *Proceedings of SIGMOD*, 1995, pp. 71–79.

[8] R. Fagin, "Combining Fuzzy Information: an Overview," *SIGMOD Record*, vol. 31, no. 2, pp. 109–118, 2002.

[9] R. Fagin, R. Kumar, and D. Sivakumar, "Comparing Top-k Lists," *SIAM J. Discrete Math.*, vol. 17, no. 1, pp. 134–160, 2003.

[10] A. Vlachou, C. Doulkeridis, K. Nørnvåg, and Y. Kotidis, "Identifying the Most Influential Data Objects with Reverse Top-k Queries," *PVLDB*, vol. 3, no. 1, pp. 364–372, 2010.

[11] F. Korn and S. Muthukrishnan, "Influence Sets Based on Reverse Nearest Neighbor Queries," in *Proceedings of SIGMOD*, 2000, pp. 201–212.

[12] E. Aichert, C. Böhm, P. Kröger, P. Kunath, A. Pryakhin, and M. Renz, "Efficient Reverse k-Nearest Neighbor Search in Arbitrary Metric Spaces," in *Proceedings of SIGMOD*, 2006, pp. 515–526.

[13] F. Korn, S. Muthukrishnan, and D. Srivastava, "Reverse Nearest Neighbor Aggregates Over Data Streams," in *Proceedings of VLDB*, 2002, pp. 814–825.

[14] K. C. K. Lee, B. Zheng, and W.-C. Lee, "Ranked Reverse Nearest Neighbor Search," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 7, pp. 894–910, 2008.

[15] A. Singh, H. Ferhatosmanoglu, and A. S. Tosun, "High Dimensional Reverse Nearest Neighbor Queries," in *Proceedings of CIKM*, 2003, pp. 91–98.

[16] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proceedings of WWW*, 2001.



**Konstantinos Georgoulas** received the BSc degree from the Department of Informatics of the Athens University of Economics and Business in 2002. In 2008 received the MSc from the Department of Informatics of the University of Ioannina. He is currently a PhD Student at Athens University of Economics and Business. His research interests include query processing and data management in large-scale systems.



**Akriki Vlachou** received the BSc and MSc degrees from the Department of Computer Science and Telecommunications of University of Athens in 2001 and 2003, respectively, and the PhD degree in 2008 from the Athens University of Economics and Business (AUEB). She is currently a post-doctoral researcher at the University of Piraeus. Her research interests include query processing and data management in large-scale distributed systems.



**Christos Doulkeridis** received the BSc degree in electrical engineering and computer science from the National Technical University of Athens and the MSc and PhD degrees in Information Systems from the Department of Informatics of Athens University of Economics and Business. He is currently an assistant professor in the Department of Digital Systems of the University of Piraeus. His research interests include parallel and distributed data management, and data analytics.



**Yannis Kotidis** received the BSc degree in electrical engineering and computer science from the National Technical University of Athens, the MSc and PhD degrees in computer science from the University of Maryland. He is an associate professor in the Department of Informatics of Athens University of Economics and Business. Between 2000 and 2006, he was a senior technical specialist at the Database Research Department of AT&T Labs Research in Florham Park, New Jersey. His main research interests include

large scale data management systems, data warehousing and sensor networks.