A Social Network-Based Inference Model for Validating Customer Profile Data

Sung-Hyuk Park, Soon-Young Huh, Wonseok Oh, and Sang Pil Han

Abstract

Drawing from the social and relational perspectives, this study offers an innovative conceptualization and operational approach regarding the validation of self-reported customer demographic data, which has become an essential corporate asset for harnessing business intelligence. Specifically, based on social network and homophily paradigms in which individuals have a natural tendency to associate and interact frequently with others with similar characteristics, we constructed a relational inference model to determine the accuracy of self-administered consumer profiles. In addition, to further enhance the reliability of our model’s prediction capability, we employed the entropy mechanism that minimizes potential biases that may arise from a simple probabilistic approach. To empirically validate the accuracy of our inference framework, we obtained and analyzed over 20 million actual call transactions supplied by one of the largest global telecommunication service providers. The results suggest that our social network-based inference model consistently outperforms other competing mechanisms (e.g., weighted average and simple relational classifier) regardless of the criteria choice (e.g., number of call receivers, call duration, and call frequency), with an accuracy rate of approximately 93 percent. Finally, to confirm the generalizability of our findings, we conducted simulation experiments to validate the robustness of the results in response to variations in parameter values and increases in potential noise in the data. We discuss several implications related to business intelligence for both research and practice, and offer new directions for future studies.

Keywords: Customer profile, data quality, business intelligence, inference model, social network, query processing system, simulation experiment