Zero-Coding UMAP in Marketing

A Scalable Platform for Profiling and Predicting Customer Behavior by Just Clicking on the Screen

Takuya Kitazawa
Arm Treasure Data
kitazawa@treasure-data.com

ABSTRACT

Customer Data Platform (CDP) is an integrated customer database operated by marketers. In the context of UMAP, this paper demonstrates a real-world CDP with a special focus on (1) simple and deterministic text-based behavioral profiling technique, and (2) GUI-based versatile tool for predictive analytics. Those functionalities are designed for those who have no expertise in machine learning and natural language processing, so the only thing marketers have to do is clicking some buttons on UI. Meanwhile, their back-end system ensures scalability and utility of the entire workflow from data collection and management to prediction and visualization.

CCS CONCEPTS

• **Information systems → Enterprise applications**; *Information integration*; *Personalization*; *Decision support systems*; Computing platforms; Data warehouses;

KEYWORDS

Customer data platform; user profiling; predictive analytics

ACM Reference Format:

Takuya Kitazawa. 2019. Zero-Coding UMAP in Marketing: A Scalable Platform for Profiling and Predicting Customer Behavior by Just Clicking on the Screen. In *27th Conference on User Modeling, Adaptation and Personalization Adjunct (UMAP'19 Adjunct), June 9–12, 2019, Larnaca, Cyprus.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3314183.3324970

1 BACKGROUND

A rapidly growing need for optimizing day-to-day marketing activities has recently introduced a new concept named **Customer Data Platform (CDP)**. Earley [3] explained that CDP is a centralized place for creating customer profiles, implementing marketing campaigns, and predicting customer behavior in connection with a variety of data and signal sources. In fact, UMAP techniques potentially play an important role in making deeper insights about a large number of customers on CDP. However, the implementation is not straightforward due to the end user's limited technical expertise and complexity of real data. Therefore, this paper provides a practical solution to the unique challenge by demonstrating a commercially available CDP, Arm Treasure Data enterprise CDP[1].

[1]https://www.treasuredata.com/

UMAP'19 Adjunct, June 9–12, 2019, Larnaca, Cyprus

ACM ISBN 978-1-4503-6711-0/19/06.
https://doi.org/10.1145/3314183.3324970

2 DEMONSTRATION

Fig. 1 shows that the CDP stands on a data management layer using open-sourced big data processing tools such as Digdag[2], Presto[3], Apache Hive and Hivemall [4]. These components guarantee scalability and maintainability of the whole system in accordance with the widely recognized best practices [6]. Plus, seamless, accurate, and reliable integration with abundant third-party tools ensures the reproducibility of data modeling as discussed in [5].

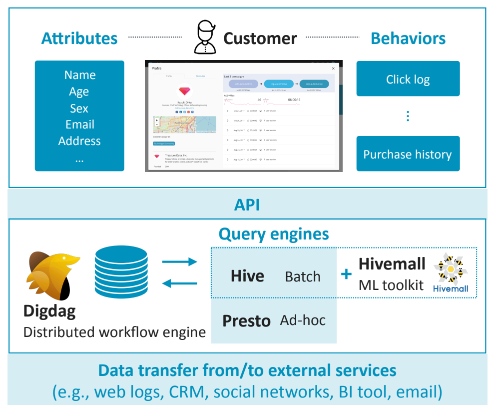


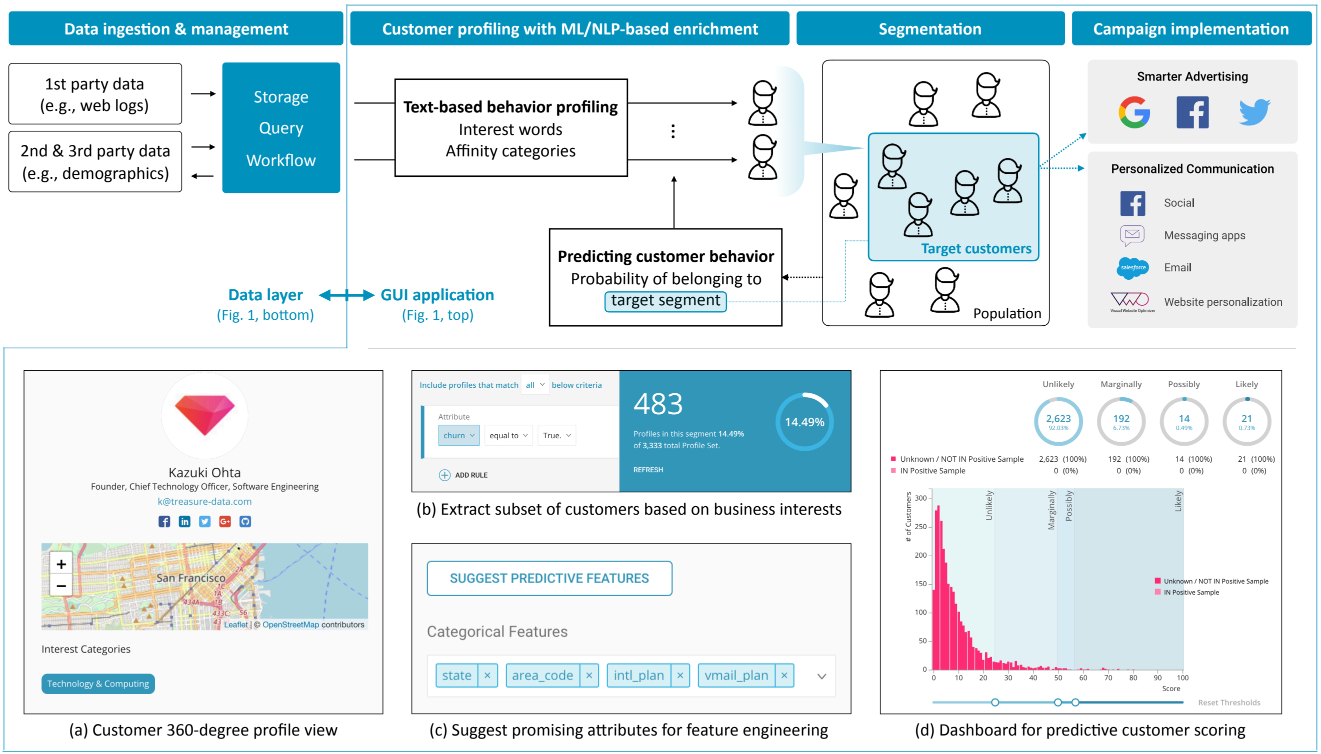
Figure 1: Architecture of the demonstrated CDP. Underlying solid data layer stores massive data on the cloud.

On the other hand, our web-based GUI application covers those technical details with dedicated views for end users. Fig. 2 and the rest of this section describe how marketers interact with their customer data on the application. Most importantly, no coding is needed at every stage as follows.

First, once CDP users specified desired data and signal sources, unified customer profiles are systematically populated by aggregating static attributes and time-stamped behavioral data. At the same time, the system conducts simple, deterministic text-based customer profiling on titles and descriptions of customer's visited web pages as needed. Consequently, customer records are enriched with their ***interest words*** and ***affinity categories*** as seen in an individual's profile view like Fig. 2 (a). Unfortunately, we cannot explain the detail of the profiling technique due to space limitations, but it basically relies on (i) TF-IDF weighting for interest word extraction in a similar way to [2], and (ii) a large word-to-category mapping table generated from the Wikipedia corpus.

Subsequently, marketer defines a subset of customer profiles (i.e., customer segment) as illustrated in Fig. 2 (b), with a flexible

[2]https://www.digdag.io/
[3]https://prestodb.io/



(a) Customer 360-degree profile view

(b) Extract subset of customers based on business interests

(c) Suggest promising attributes for feature engineering

(d) Dashboard for predictive customer scoring

Figure 2: End-to-end UMAP workflow implemented with the demonstrated CDP. Demo scenario follows this entire flow[4].

choice of filtering conditions. The CDP then enables to implement marketing campaigns on arbitrary customer segment. For example, if you create a segment consisting of dormant customers, you may want to send a special offer to them for customer reactivation.

Moreover, an advanced use case of the platform is to implement look-ahead-based marketing campaigns by predicting unseen customer's behavior. In the previous example, it means that marketers can preliminarily find out customers who are likely to become dormant *before* they actually stop using your service. The predictive analytics capability, namely ***predictive customer scoring***, technically solves binary classification problem on customer profiles, and the classifiers eventually give the probability of belonging to a specific "target segment" (e.g., "set of customers who have not logged in for a year" for dormant customer prediction). The probabilities are finally visualized on a dashboard view as Fig. 2 (d) with some auxiliary metrics such as evaluation accuracy and feature importance. In the end, marketers can easily and effectively perform predictive marketing campaigns.

As the scenario above suggests, the CDP simply requires users to click buttons on UI at each step of customer profiling, segmentation, predictive analytics, and campaign execution. It should be noted that, while practitioners normally spend a huge amount of time on feature engineering, the predictive analytics functionality makes the task semi-automated by suggesting features as captured in Fig. 2 (c). The process internally takes a heuristics-based approach with single-column profiling [1] on customer attributes.

[4]A supplemental screencast is available at: https://youtu.be/iwbqb5D2uPw

3 DISCUSSION

As an example of productized UMAP solution, we have demonstrated the enterprise-grade implementation of CDP that provides a solid data layer and marketer-friendly GUI-based application. From an algorithmic point of view, the CDP employs limited conventional UMAP techniques like TF-IDF weighting and binary classification to make the application as explainable, scalable, and accurate as possible for non-expert users. On that point, future challenges might be related to a field of interactive and explainable user modeling e.g., based on topic modeling and clustering.

4 ACKNOWLEDGMENTS

The author would like to thank Kazuki Ohta, Sadayuki Furuhashi, Takahiro Tanaka, Tsubasa Tomoda, and all contributors to this project for their professional work. We would also like to thank the users of our CDP solution for providing valuable feedback.

REFERENCES

[1] Z. Abedjan et al. 2015. Profiling Relational Data: A Survey. *The VLDB Journal* 24, 4 (2015), 557–581.
[2] J. Beel et al. 2017. TF-IDuF: A Novel Term-Weighting Scheme for User Modeling based on Users' Personal Document Collections. In *Proc. of iConference 2017*.
[3] S. Earley. 2018. The Role of a Customer Data Platform. *IT Professional* 20, 1 (2018), 69–76.
[4] T. Kitazawa and M. Yui. 2018. Query-Based Simple and Scalable Recommender Systems with Apache Hivemall. In *Proc. of ACM RecSys 2018*. 502–503.
[5] P. Sugimura and F. Hartl. 2018. Building a Reproducible Machine Learning Pipeline. `arXiv:1810.04570 [cs.LG]` (2018).
[6] M. Zinkevich. 2017. Rules of Machine Learning: Best Practices for ML Engineering. http://martin.zinkevich.org/rules_of_ml/rules_of_ml.pdf