Automatically Conceptualizing
Social Media Analytics Data via Personas

Soon-Gyo Jung, Joni Salminen, Jisun An, Haewoon Kwak, Bernard J. Jansen

Qatar Computing Research Institute, HBKU
{sjung, jsalminen, jan, hkwak, bjansen}@hbku.edu.qa

Abstract

Social media analytics is insightful but can also be difficult to use within organizations. To address this, we present Automatic Persona Generation (APG), a system and methodology for quantitatively generating personas using large amounts of online social media data. The APG system is operational, deployed in a pilot version with several organizations in multiple industry verticals. APG uses a robust web and stable back-end database framework to process tens of millions of user interactions with thousands of online digital products on multiple social media platforms, including Facebook and YouTube. APG identifies both distinct and impactful audience segments for an organization to create persona profiles by enhancing the social media analytics data with pertinent features, such as names, photos, interests, etc. We demonstrate the architecture development, and main system features. APG provides value for organizations distributing content via online platforms and is unique in its approach to leveraging social media data for audience understanding. APG is online at https://persona.qcri.org.

Introduction

Social media analytics is used for audience analysis. However, this data has disadvantages in certain contexts, being voluminous and granular. One method of simplifying the presentation of this data is via the use of personas. Personas are employed in domains from software development to system design to marketing in order to describe and communicate about core users, customers, or audiences. One can think of personas as mental shortcuts that present the who, what, and why of a set of individuals. Personas can assist in defining organizational strategy, develop new products, and improve customer operations. Personas are integrated into many design, product, marketing, content, and advertising development processes.

The Automatic Personas Generation (APG) system leverages privacy-preserving aggregated data of audience interactions with digital content posted on major online social media and other analytics platforms. APG collects, processes, and decomposes this actual user data and then enriches the results with descriptive attributes to produce data-driven persona profiles. APG can produce personas from millions of user interactions within a matter of days, which is far shorter time than traditional persona creation.

Here, we present an overview of APG, including architecture and development. We then highlight the core features of APG, with other features left for the demo. We discuss use cases, uniqueness, and commercial efforts.

System Overview

APG uses a stable, robust, and scalable structure employing (a) the Flask web framework to support the front-end applications, services, and application programming interfaces (APIs), (b) PostgreSQL database for back-end data storage, processing, and (c) Python libraries including Pandas and scikit-learn for data analysis (Jung, et al, 2017).

APG accesses the specific online social media platforms, (e.g., Facebook, YouTube) via the analytics API of the account holders. Typical user profile data from these platforms are demographic variables of gender, age, country location, and which site the user comes from, although at an aggregated level. Via the API, the APG system collects the detailed interactions of users with the online content. This is accessible only to managers of a particular social media account (e.g., YouTube channel).

To automatically generate personas from this data, our methodology (Jung, et al, 2017) applies a sequential approach, consisting of:

- identifying distinct user interaction patterns,
- linking these patterns to demographic groups,
- identifying impactful demographic groups from the data,
- creating base personas via demographic attributes,
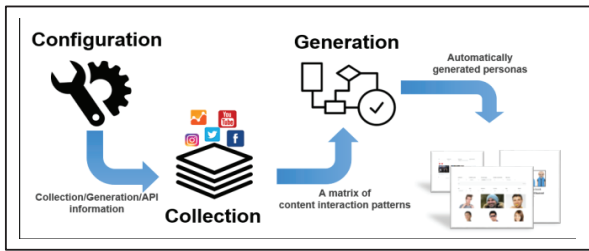- enriching these shell personas to create a complete persona.

Figure 1: APG flowchart from server configuration to data collection, persona generation, and users accessing system.

With the complete data flow shown in Fig. 1, we now present some key features of APG to demonstrate system capabilities, with others shown during the live demonstration.

System Features

As shown in Fig. 2, a user can select the number of personas to generate, from a minimum of 5 and a maximum of 15. The user can also select the desired data source and is given options for viewing the personas. The listing of the personas, combined with the mouseovers, affords the user an overview of the audience segments in terms of demographic information, as shown in Fig. 2. By hovering over a persona image, the basic attributes are displayed, in this case, *'Neelix*, *United States*, *3*, *Male'*. The end user can also change content collections or the number of personas, or apply different filterings, such as data set, topics of interest, ethnicity, country, age, or gender. A click on one of the images in the persona listing will display the corresponding persona profile.

There are a multitude of other features and capabilities that will be presented during the live demonstration, including comparing personas between periods, comparing current user segment size for personas against potential audience size, and displaying personas related to each item of content (Salminen, et al, 2018).

Results and Discussion

Use Cases for APG

APG is currently deployed, in pilot version, in multiple organizations and is used to understand audiences on multiple social platforms, design engaging digital content, and communicate more effectively concerning audiences, users, and customers.

Comparison to Other Systems

To the best of our knowledge, APG is the first deployed system to use large amounts of online behavioral user data for quantitative and automatic persona generation.
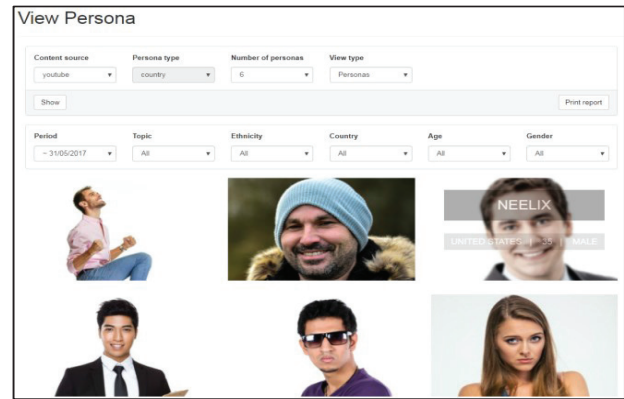

Figure 2: APG for selecting the content source platform (e.g., Facebook, YouTube), the number of personas to generate (5 to 15), and the view type (various formats of viewing the relationship between content and personas). Th options in the screen capture are set to YouTube and 6 personas.

Commercialization Efforts

Presently, APG is in beta use with three major multi-national corporations, one major non-profit organization, and two SMEs in the retail vertical, with efforts underway for full deployment. The APG project is the recipient of development funding to take the system to Technology Readiness Level 8 (TRL-8). APG has an evaluation license in place and is participating in an Entrepreneur in Residence (EIR) program, with start-up funding in progress.

Conclusion

In this demonstration, we will show the APG capability for automatically creating personas from large scale, quantitative, aggregated user data from major social media platforms. The personas are generated from real data reflecting real user behavior and are updated automatically. APG can be applied to a wide range of contexts and can be used in a stand-alone mode or used in conjunction with or to supplement personas created via qualitative means.

References

Jung, S., An, J., Kwak, H., Ahmad, M., Nielsen, L., & Jansen, B. J. (2017, 6-11 May). *Persona Generation from Aggregated Social Media Data.* Paper presented at the ACM Conference on Human Factors in Computing Systems 2017 (CHI2017), Denver, CO.

Salminen, J., Nielsen, L., An, J., Jung, S.G., Kwak, H., and Jansen, B. J. (2018) Is More Better?": Impact of Multiple Photos on Perception of Persona Profiles. ACM CHI Conference on Human Factors in Computing Systems (CHI2018), Montréal, Canada, 21-26 April.