



reSonate

Presented By: Sam Hawala
Sam.hawala@resonate.com

An SVD application



AGENDA

- I. Problem to solve
- II. Use Cases
- III. Data / Issues
- IV. SVD

I. Problem to solve

- Finding people with similar online behavior
- A persistent problem in the digital marketing industry is the superabundance of web sites people visit, and the many ways visitors interact with the sites.
- Online behavior data are sparse.

Zipf's law: site frequency is almost proportional to rank

$$f_n \propto 1/n^a$$

I. Problem to solve

1. We need a sufficient number of web sites to model the whole online population of 250M.
2. Any automatic data processing approach will have diminishing returns with the total number of web sites we observe.
3. No matter how much data we observe during training, we will invariably see something new by the time we deploy the model.

II. Use-Cases

Use-Case 1

An Ad campaign for a **retail outlet**.

The goal: **drive clicks** and essentially make a purchase.

The client was judging us on a backend ROAS model for online sales.

We optimized toward **clicks**.

Use-Case 2

A **visitor bureau** campaign to get people to sign up for travel brochures.

We optimized toward **conversions** for brochure sign ups.

II. Use-Cases

Use-Case 3

An ad campaign to get people to **sign up for 3 different certification classes**.

We optimized toward **conversions** which were counted when someone submitted a lead (name, email) to receive more information about the courses.

Use-Case 4

An ad campaign to **drive awareness of an industry** and the products that they make that people use and need every day.

Drive **clicks** to the landing page for people to learn more about what the industry does.

III. Issues & Approach

1. Techniques based on eigen-decomposition allow for a **transformation of the data matrix** into a set of orthogonal vectors, each with an associated "strength", or eigenvalue.
2. The data matrix is compressed by discarding the less significant dimensions and allowing the **critical underlying features** to reveal themselves.
3. **Compression and understanding**: meaningful generalities can help you represent your data with fewer numbers, finding a way to represent your data in fewer numbers can often help you find meaningful generalities.

IV. Issues & Approach

1. **Operationally** the task is to match the behavior of a new user with observed behavior of all users of interest (e.g. who visited a tag.)
2. **SVD** discovers the higher-order (latent classification) structure in the association of behavior with users
3. **Latent classification** of behavior tries to overcome the problems of direct matching of behavior by using statistically derived profiles instead of individual behavior for retrieving similar users.
4. **Performance:** data show that profiles (eigenvectors) are more robust indicators of web behavior than any individual's behavior.

IV. Issues & Approach

1. **Operationally** the task is to match the sites of a new user with observed sites of all users of interest (who visited a tag.)
2. **SVD** discovers the higher-order (latent classification) structure in the association of sites with users
3. **Latent classification** of sites tries to overcome the problems of direct matching of sites by using statistically derived profiles instead of individual sites for retrieving similar users.
4. **Performance** data show that profiles (eigenvectors) are more robust indicators of web behavior than any individual site.

III. Singular Value Decomposition

$$A = U\Sigma V^T$$

1. U and V are matrices of orthogonal left and right singular vectors
2. Σ is a diagonal matrix of singular values.
3. Preserving only the k largest singular values (and hence only the first k columns) provides a least-squared error, rank- k approximation to the original matrix A . k is much smaller than the more than a million unique behaviors.
4. The subspace of eigenvectors represents important associative relationships between online users and their behavior, that are not evident in behaviors taken one-at-a-time.



Resonate Performance Report

Campaign: [REDACTED]

Date: 26 Feb 2018

Sales Rep: [REDACTED]

Account Manager: [REDACTED]

Display Performance by Audience

Strategy	<u>Impressions Run</u>	<u>Clicks</u> ▼	<u>Ad Server CTR</u>
LAM1291722	156,284	175	0.11 %
LAM1294216	24,746	8	0.03 %
LAM1294214	7,450	6	0.08 %
Total	188,480	189	0.10 %

1 - 3 of 3 items

KHAG0118A Click

BASELINE: Online Adult Population | CONTA

DEMOGRAPHIC HIGHLIGHTS

% COMPOSITION for highest indexing



52% are Female



23% are 35-44 years of age



23% with Household Income of \$50 to 75K



34% with College Degree



61% are Married



61% do not have Children

KHAGO0118A Click SLAM

BA | CONTA



Survey Question

What is your gender?



54% are Female



18% are 65+ years of age



27% with Household Income of \$25 to 50K



12% with Post-grad Degree

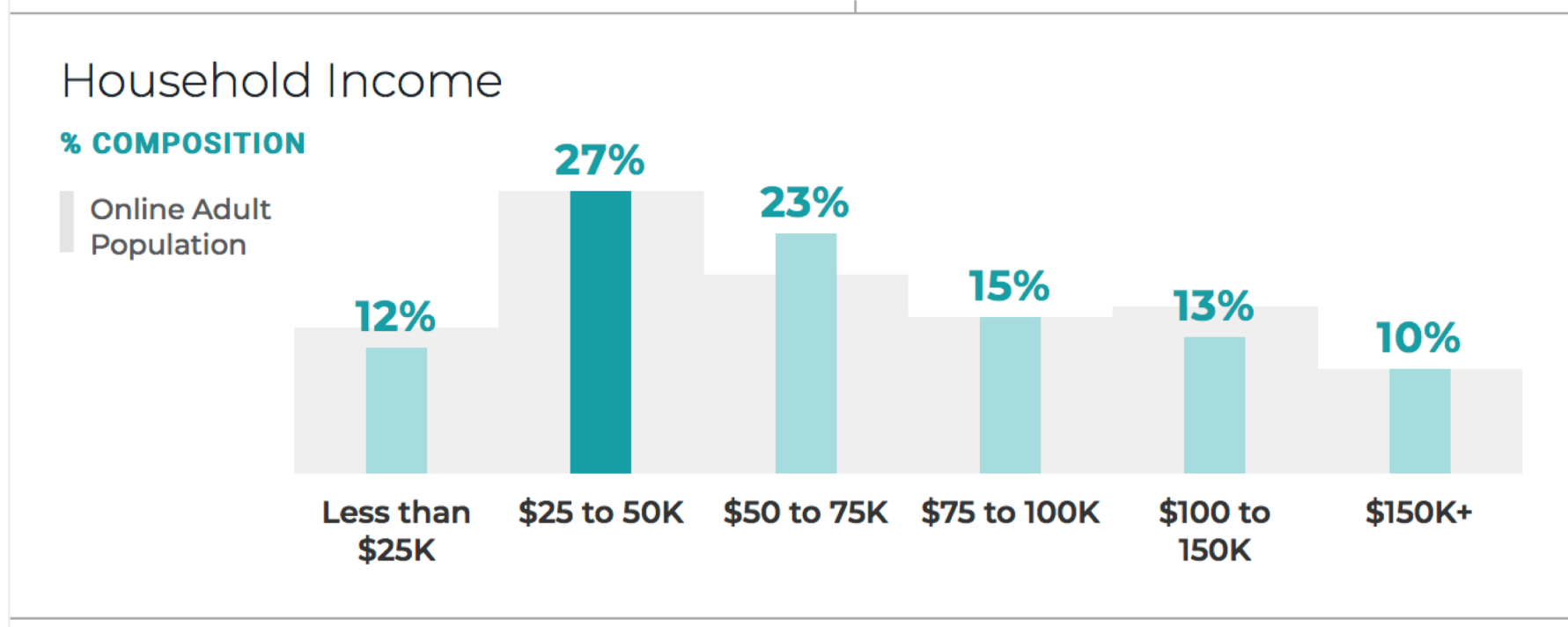


12% are Divorced



61% do not have Children

Income Distribution



Income Distribution of Look-Alikes

