AUTOMATIC TEMPORAL RETAIL SEGMENTATION FROM BIG DATA

Gavin Smith, The University of Nottingham, UK

James Goulding, The University of Nottingham, UK

Andrew Smith, The University of Nottingham, UK

For further information, please contact Gavin Smith, Dr., The University of Nottingham (gavin.smith@nottingham.ac.uk).

Keywords: Segmentation, Topic mining, Consumer Behavior, Big Data

EXTENDED ABSTRACT

Transactional data captures a significant amount of information regarding consumer purchasing behavior. However, in its raw form such data is both large and unwieldy, obfuscating actionable insights able to inform business decisions. One example of insights that have shown clear business utility is the summarization of consumer behavior into actionable behavioral sub-groups or "customer segments". These segments are normally obtained via the manual or semi-automated specification of a set of behavioral groups based on a pre-selected set of measurable aggregate statistics. These segments are relatively simple to create, interpret and manage. However, the complexity of the human behavior they seek to distil is often incompatible with such approaches, in practice violating many assumptions pre-made to facility the simplicity they provide. Specifically, traditional customer segmentation (e.g. Bock and Uncles, 2002; Boone and Roehm, 2002; Oppewal et al., 2010) suffers from at least one of two key potential weaknesses:

- Segments tend to assume that an individual's behavior is stationary - i.e. it is static over time.
- Segmentation traditionally assumes that each individual can be adequately summarized by placing them in a single behavioral group or assigning a probability of membership to multiple groups.

Both of these assumptions are not only tenuous, but have the potential to lead to non-optimal business decisions. In the former, assuming that an individual's behavior is stationary neglects the well-known behavioral influences of seasonality, holiday periods and life stage changes. Under such an assumption all these differing behaviors are conflated to a single aggregate, temporally

invariant, model – a potentially unrepresentative temporally averaged model which is unlikely to reflect an individual's actual behavior at any given, short term, time frame. In the latter, discriminative and/or insightful non-dominate sub-behavioral patterns are potentially obfuscated or lost.

In this work we present and discuss a new customer analysis technique that tries to remove the coarseness of traditional segmentation approaches, while maintaining conceptual simplicity.

Research Question

Do novel mathematical techniques couple with big retail data enable us to move beyond traditional customer techniques?

Method and Data

A new approach to segmenting consumers is presented and applied to large transactional data set from a UK retail company. Results of the segmentation are presented and evaluated.

This work presents a new customer analysis technique which addresses the limitations of traditional segmentation approaches by representing consumers by the *extent to which they express a set of temporal purchasing behaviors*, derived from mass sets of transactional data. We cast this problem as a variant of finding *latent temporal topics* considered as a recent extension to traditional static topic models (Blei and Lafferty, 2006), implementing our model via Non-negative tensor factorization. In context, under such extensions temporal behavior is simultaneously modeled with purchase behavior over products. Note that these static topic models can themselves be considered an extension to the traditional Dirichlet-multiomial model allowing for in depth summarization not just of a consumer's preferences towards one type of good, but to their purchasing behavior across all products.

Thus the approach does not represent a customer simply as an instance of a single catch-all segment, but as a combination of general underlying temporal behaviors and purchasing trends. Additionally, these building block trends are not static, but reflect varying behavior - providing a significantly richer descriptive mechanism that is still parsimonious and actionable.

Summary of Findings

The significant quantities of consumer transactional data enables significant insight into consumer behaviour. However, traditional segmentation algorithms, which are typically temporally static and focus on single class assignment, currently fail to take full advantage of such data sets and the significant opportunities they afford to decision makers. Acknowledging the complexity of consumers temporal behavior that has been highlighted in consumer research literature, this work presented a technique that uses a variant of dynamic topic modeling based on non-negative tensor factorization. This technique is able to describe individual consumers by an automatically identified set of latent temporal topics - purchasing trends. As demonstrated on a large real world transactional dataset and in contrast to traditional approaches (and dynamic topic models developed for alternative domains), the technique is able to uncover the latent time-varying purchasing behaviours that underpin a market, and express customers as a combination of these building blocks. This adds sophistication to analyses while importantly retaining interpretability.

Key Contributions

1. A new computationally tractable approach to segmenting consumers from 'big data' scale transactional logs that:
   a) forgoes single class (or probabilistic assignment to a class) replacing it with segmentation based on an individual's expression of a set of automatically identified, interpretable temporal purchasing behaviours
   b) enables the identification of temporal purchasing patterns with arbitrary temporal expression including (near)periodic behaviour
2. An evaluation of the technique on a large transactional data set showing the application of the technique in drawing insights about:
   a) general temporal trends in the data set,
   b) summaries of individual's purchase behaviour and
   c) understanding behavioural groups within the population.

Acknowledgements

This work was jointly supported by the EPSRC Neo-demographics: Opening Developing World Markets by Using Personal Data and Collaboration grant, EP/L021080/1 and the EPSRC From Human Data to Personal Experience grant, EP/M02315X/1. The data used is both commercial and personal in nature and is therefore not publicly available. For further information please contact the authors.

Reference are available on request.