Identifying Ridership Patterns in an Urban Bicycle Sharing System via Poisson Mixture Models

Eric Buras\textsuperscript{1}, Hans Engler\textsuperscript{2,⋆}

\textsuperscript{1} Rice University
\textsuperscript{2} Georgetown University, Washington, DC
⋆ Contact author: engler@georgetown.edu

Keywords: Poisson mixture models, soft clustering, bicycle sharing systems, community detection

Washington, DC has had a bicycle sharing system since 2010. There are several hundred self-service rental stations where riders can pick up and drop off bicycles, either as long-term “registered” members (typically, commuters) or as short-term “casual” users (typically, tourists). There are now about 10,000 rides per day during spring and summer, making this the second largest such system in North America after New York City. Worldwide, there are more than 600 such systems. The Washington, DC system makes its past ride data as well as the current status of all stations available to the public.

In this contribution, a class of Poisson mixture models is used to identify ridership patterns. Specifically, consider any two stations \(i, j\) in the system, and consider a set of time intervals \( \{I_t\} \) that together cover a 24 hour day, labeled e.g. by start hours \(t \in \{0, 1, \ldots, 23\}\). It is postulated that rides from \(i\) to \(j\) by a specific ridership group (e.g. registered riders) on a specific type of day (e.g. on weekdays) that start (or end) during \(I_t\) follow a Poisson distribution with intensity \(\alpha_{ij}\lambda_{kt}\) for some \(k \in \{1, \ldots, K\}\), where \(\alpha_{ij}\) describes the overall daily traffic volume from \(i\) to \(j\), \(k\) indicates membership of the station pair \((i, j)\) in one of \(K\) unknown clusters, and \(t \mapsto \lambda_{kt}\) denotes a daily intensity pattern specific to this cluster which is also unknown. This approach follows a class of models introduced in [1], where ridership patterns in the Paris bicycle share system were analyzed.

Using a suitable implementation of the EM-algorithm in \(R\), cluster memberships and associated daily intensity patterns are identified for a large fraction of daily rides, indicating that the ridership community has settled on a robust and stable usage pattern. Clusters are readily identified as morning commutes, afternoon commutes, or evening trips to and from neighborhoods with social activities. The patterns are especially complex near subway stations where riders transfer between two modes of transportation. The methodology also allows short term predictions of bicycle availability at individual stations. Results will be visualized with \textit{shiny}.

This work is based on the undergraduate honors thesis of the first author which was supervised by the second author.

References