BEST FRENEMIES: A COMPARISON OF SHELF STOCKING AND WEB SCRAPING

Scott Walker, NMR Group, Inc., Somerville, MA
Melissa Meek, NMR Group, Inc., Somerville, MA
David Barclay, NMR Group, Inc., Somerville, MA

Introduction

As part of a recent residential lighting shelf stocking study for two northeastern states, NMR employed two methods of data collection: traditional shelf stocking store visits and web scraping. This poster compares the methods and results.

Shelf Stocking

A traditional shelf stocking study includes visits to a sample of retail stores, where technicians take a complete inventory of the products of interest. Shelf stocking studies are commonly used as a method to evaluate residential upstream lighting programs. A shelf stocking study requires careful training of technicians, who must collect detailed information on a large selection of lighting products on the shelves of retail stores where store managers may limit the time available to do the inventory. Shelf stocking studies are relatively costly, in that they require travel and lodging expenses and training time. Store visits, however, are the only way to evaluate product display and accurately determine in-store pricing. Also, any type of retailer, depending on the cooperation of store management, can be visited during a shelf stocking study.

Web Scraping

Web scraping refers to software tools that programmatically extract data from websites. A familiar example is Google's search engine, which crawls websites extracting content to present search results. There is pre-made web scraping software available, and there are companies that provide it as a service, but the tools are typically custom-developed for a specific task. In this example, we used web scraping to gather data on lighting products from the websites of two major home improvement retailers.

The most common methods for web scraping rely on open source software tools using the R or Python programming languages. Web scraping also requires knowledge of web technologies like HTML and relational database software such as FileMaker or MySQL. Web scraping is widely used, but may violate a website's terms of use. Websites and web scraping tools can change frequently in ways that require updates to scraping scripts.

A knowledgeable programmer can set up a web scraping script for a particular website in a few days. Scripts can extract any information that the website provides about their products, and can be run automatically on an arbitrary schedule to create time-series data. The data presented on web sites varies widely from site to site in content, quality, and structure, which may require data cleaning and regular updates to the scraping scripts. The potentially large amount of data available can create challenges for cleaning and filtering to extract relevant insights. Generally, only major retailers have web sites that provide extensive, well-formatted product information that justifies the work of webscraping.

Web scraping data can also be used to improve other tools. NMR was able to use the large database of products created via web scraping to populate the tablet-based data collection tool used on the shelf stocking visits, which reduced in-store time for technicians and improved data quality.

Comparison Results

NMR did shelf stocking visits and web scraping runs for the same two major home improvement retailers during the same week. At Retailer 1, the product selection by lamp type was very similar. At Retailer 2, we found a larger share of LED and halogen bulbs on the shelves compared to the web store. Comparing the prices at both retailers, we found that in-store prices were generally lower than online. We concluded that product rebates are inconsistently applied to online prices.

Conclusions

Web scraping and shelf stocking prove to be complementary methods. They both have advantages, insights, and limitations unavailable using one method alone. Shelf stocking visits are difficult and time-consuming, but are the only method to evaluate product display and in-store pricing, and the only way to collect data from all store channels. Web scraping can be done on a regular schedule at low cost, provides a high level of detail on product features, and the method can prove useful for many evaluation tasks.