

Consumer reports data exploration

Andrew Sutcliffe

January 20, 2014

Approach

The idea was to examine if there were any interesting cross-franchise features that could be exploited. By features we mean any rated quality of an item. If we can find features that are cross franchise it may allow to establish similarities between items from different sub-franchises within a franchise. Unfortunately this analysis showed that there were very few similarities at the level of sub-franchises. Similarities on a specifications level is quite similar, although there may seem to be a bit more cross over with specifications.

Finer grain analysis may be interesting if we consider the example of the humidifiers. Humidifiers have two sub-categories either tableTops or console. These two categories share 100% of features. It may be interesting to search through the data base, find such categories and flatten them, adding a feature to them. Although doing this on features may not be as appropriate as doing this on specifications.

Similarity comparison rated features vs specs across different sub-franchises

Only franchises of interest are shown. Franchises with only one sub-franchise that has items are not shown. The scale goes from 0% deep blue to 100% red. The high match in the health category is explained by the fact that they had no specifications for those sub-franchises.

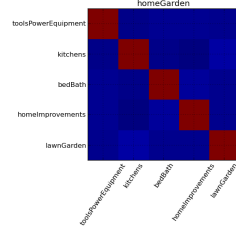


Figure 1: homeGarden features

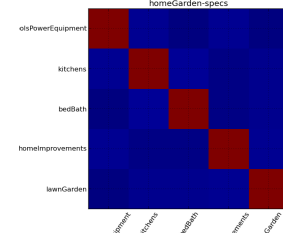


Figure 2: homeGarden specs

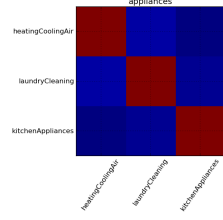


Figure 3: appliances features

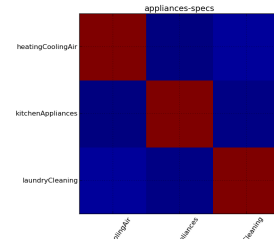


Figure 4: appliances specs

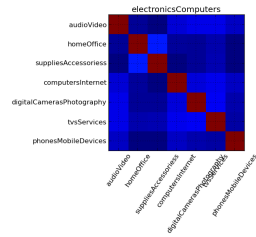


Figure 5: electronicsComputers features

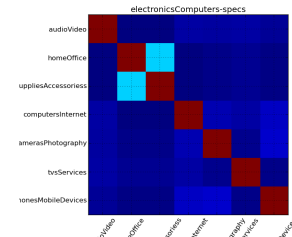


Figure 6: electronicsComputers specs

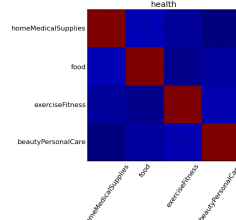


Figure 7: health features

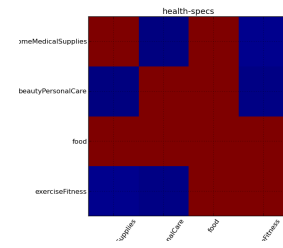


Figure 8: health specs