Forecasting payment terminals failure

Patrick

1/19/2020

Project summary

Say the company we work for manufactures and sells electronic devices in a B2B context while offering after sales services. In some sectors, shutdown time might be extremely costly for clients and handling device reparations requires maintening a sufficient stock of spare parts at hand. Yet, producing and storing those parts comes with a cost which is why inventory levels should be about just high enough to cover the demand. This project explores the idea of quantifying and forecasting the occurence of device failures and therefore the future reparation needs using as input historical data on product sales and reparations (as a proxy for failure occurences).

Literature reviews on the topic (Auweraer, Boute, and Syntetos 2019; Dekker et al. 2013) stress the poor performence of traditional time series forecasting methods for the task of predicting spare parts demand. A first challenge is that the occurence of failures is erratic and intermitent. Another is that our expectations on the number of failures should obviously take into account the number of devices currently on the market, which we will refer to as the *installed base*, as well as its expected evolution, rather than naïvely extrapolating patterns observed in historical failure data.

The method explored here consists in three steps: First, a time-to-failure probability distribution will be estimated for each device part and failure type based on historical failure data. In a second phase, a forecast of future failures is performed based on the current state of the installed base by exploiting the fact that the expected number of part failures in time `t` is the sum of the individual probabilities to fail in time `t` for all parts in service. Finally, the exercise will be repeated in a setting where the future size and composition of the installed base is estimated.

Literature takeaways

The proposed method follows the literature's recommendations to think of the installed base as the driving factor for failures and therefore to exploit its current but also future state for the purpose of building expectations on future failures. Besides, it follows the subset of this literature that estimates time-to-failure probability distributions for each particular product instead of single-point estimates. What distinguishes the present attempt is the use of a non-parametric density estimation technique to build these probability distributions while generally a specific distribution is assumed. I also explore how statistical information can be combined with business knowledge to build pragmatic expectations on the future evolution of the installed base.

Data

```
source_df <- read.csv("source_df.csv")
```

The data used as source for the analysis is generated for the purpose of this project (see `Dataset generation: commented code`). The minimum requirement for this methodology to work is that one can gather the following information for (most) devices that have been sold in the past:

- its model;
- the date at which it has been sold;
- the reparation date, if any;

- the description of the failure (i.e. part that failed and reason), if any.

The dataset used here contains 5000 records of that type. In total, 3000 devices of model `A` were sold between January 01 2017 and December 31 2019 and 2000 devices of model `B` were sold between January 01 2018 and December 31 2019. For example, the 5 first lines of the dataset look like:

X	terminal_model	date_sold	failure_description	failure_date
1	A	2019-07-22	NA	NA
2	A	2018-12-04	NA	NA
3	A	2019-01-24	Key not responding	2019-05-31
4	A	2019-06-25	NA	NA
5	A	2017-10-19	Key not responding	2019-09-06

were *failure_description* and *failure_date* take the value `NA` for devices that have not been sent back for reparation. 4 different types of failures were recorded with the following frequency:

failure_description	n
Key not responding	752
Screen flickering	80
Screen shutdown	314
NA	3854

Exploration and density estimation

Now let's start exploring the dataset as if we didn't know its generating process. First we'd be interested in understanding the lifetime at which failures generally occur. A first look reveals that the typical time-to-failure is quite different from part to part, hence from device to device.

References

Auweraer, Sarah Van der, Robert N Boute, and Aris A Syntetos. 2019. "Forecasting Spare Part Demand with Installed Base Information: A Review." *International Journal of Forecasting* 35 (1). Elsevier: 181–96.

Dekker, Rommert, Çerağ Pinçe, Rob Zuidwijk, and Muhammad Naiman Jalil. 2013. "On the Use of Installed Base Information for Spare Parts Logistics: A Review of Ideas and Industry Practice." *International Journal of Production Economics* 143 (2). Elsevier: 536–45.