LOCALITY-SENSITIVE HASHING FOR MARKET SEGMENTATION

Paul Drozd[1]

[1]Department of Intelligent Systems, Belarussian State University, Belarus
drozdps@gmail.com

Customer (or market) segmentation is the process of dividing customers into groups based on common characteristics so that any business can market to each group effectively and appropriately. This research suggests new method for customer segmentation which have demonstrated great results on real business examples.

Our approach eliminates the instability of machine learning for customer segmentation using Locality-Sensitive Hashing algorithm (abbreviated as LSH), which is commonly used as the nearest neighbor search (NNS) [1]. LSH utilizes a specific set of bad hash-functions. The main feature of these hash functions is that they generate collisions on similar objects so neighboring points are tend to lie in the same hash bucket, which is strictly forbidden for regular applications of hash functions, such as cryptography (Fig. 1).

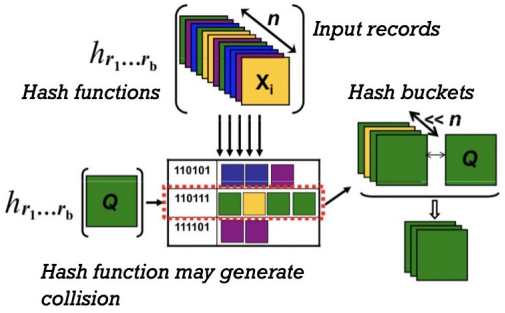

Fig. 1. Locality-Sensitive Hashing algorithm

Then Bisecting K-means algorithm (abbreviated as BSK) is used for processing the values of the hash functions in a metric space. This technique significantly increases machine learning stability and demonstrates better performance than other clustering algorithms [2].

The program implementation of the suggested LSH+BSK approach was developed using Apache PredictionIO technology stack. The sample use-case scenario of our market segmentation system is given below (Fig. 2).

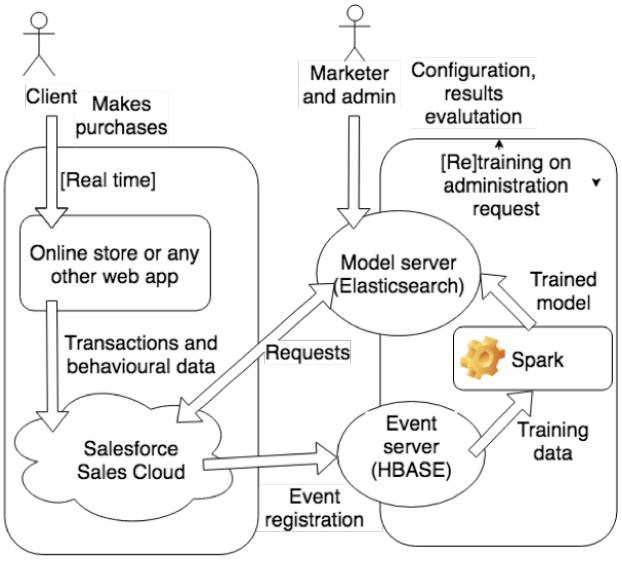

Fig. 2. Customer segmentation system use-case scenario

The system is able to handle both behavioural data (for example, the average session time a customer spends surfing your e-commerce website) and transactional records (for example, the most expensive transaction amount for a customer within a specified period of time). Our machine learning algorithms are completely automated, so the system can be used by marketers, financial analysts, product managers and other non-technical specialists.

[1] Fern, X.Z., Clustering ensembles for high dimensional data clustering / X.Z. Fern, C.E. Brodley // In Proc. International Conference on Machine Learning / ed. T. Fawcett. Washington DC, 2003. P.178-185.

[2] Drozd P. Kohonen's neural networks for Customer Segmention / P. Drozd // Open Readings 2018 / ed. E. Skliutas. Vilnius, 2018. P.93.