INTERACTIVE LEARNING SYSTEMS APPLIED TO THE RETAIL MARKET

Author: António Aristides Romualdo Carvalho
Thesis Advisor: Manuel Fernandez Veiga
PhD Program: Information and Communications Technology of the University of Vigo

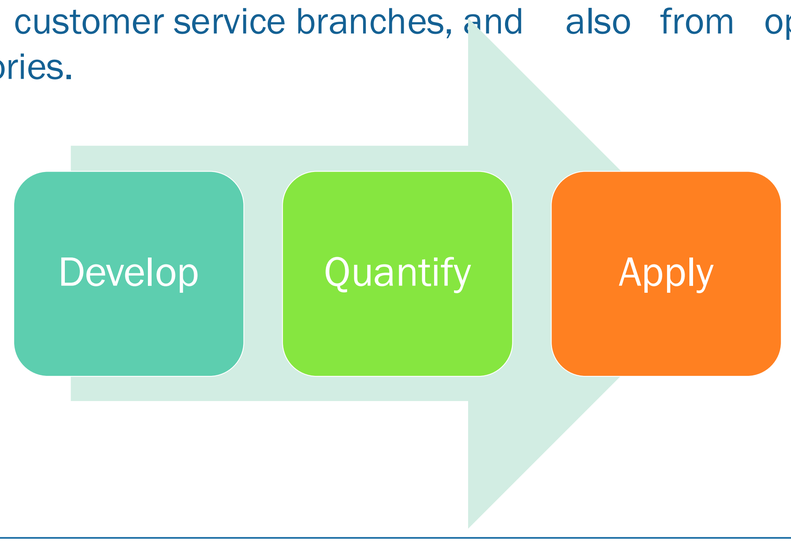
UNIVERSIDADE DE VIGO

Motivation Of The Work

Active learning is the process in which an algorithm is presented with a large pool of unlabeled data and can interactively ask to an oracle - typically, a human expert - for the labels of a subset of the examples, at the algorithm's choice. The ultimate goal is to reduce drastically the number of labels needed to build a statistically significant prediction model. In other words, the number of labels needed to build a hypothesis for the raw data with low error probability. The approach is also generally known as interactive learning, in that it involves a loop in the algorithmic procedure, such that the model is refined as the oracle reveals more "true" labels for the unknown (unlabeled) examples. Clearly, this approach can save lots of time and computational effort in discovering good statistical models for the data, and has important applications, particularly in areas where the underlying space and the estimation problem are hard. Examples come from application areas as health, protein folding, image recognition, etc. For the purposes of this thesis, the application area we are most interested in is crowdsourcing, i.e., the possibility of using collective, distributed feedback to develop good learning algorithms.

Thesis Objectives

- Develop new interactive statistical learning algorithms tailored to the retail markets, based on consumer segments (number of children, gross salary, etc), item locations, suppliers, etc (for example to advise products/items based on the number of children of a consumer).
- Quantify the computational complexity of those algorithms and statistical accuracy, for knowing what kind of machine should be used, as well as trying to optimise the algorithms for the computational complexity to be as lower as possible. On the theoretical side, perform an introductory analysis of the interaction step by using information-theoretical tools.
- Apply the techniques to “real” datasets and implement an algorithmically efficient Interactive/Deep Learning System (ILS), so that all the work made in the below objectives can be verified, with real datasets. Datasets will be collected from actual customer service branches, and also from open data repositories.



Research Plan

- Study retail markets.
- Study interactive learning systems.
- Study R programming language.
- Analyze datasets related with retail markets.

Results and Discussions

The results and discussions (for now) are in a preliminary phase, to delimit the clustering algorithms, as well as metrics to measure the algorithm(s) performance.

R Console

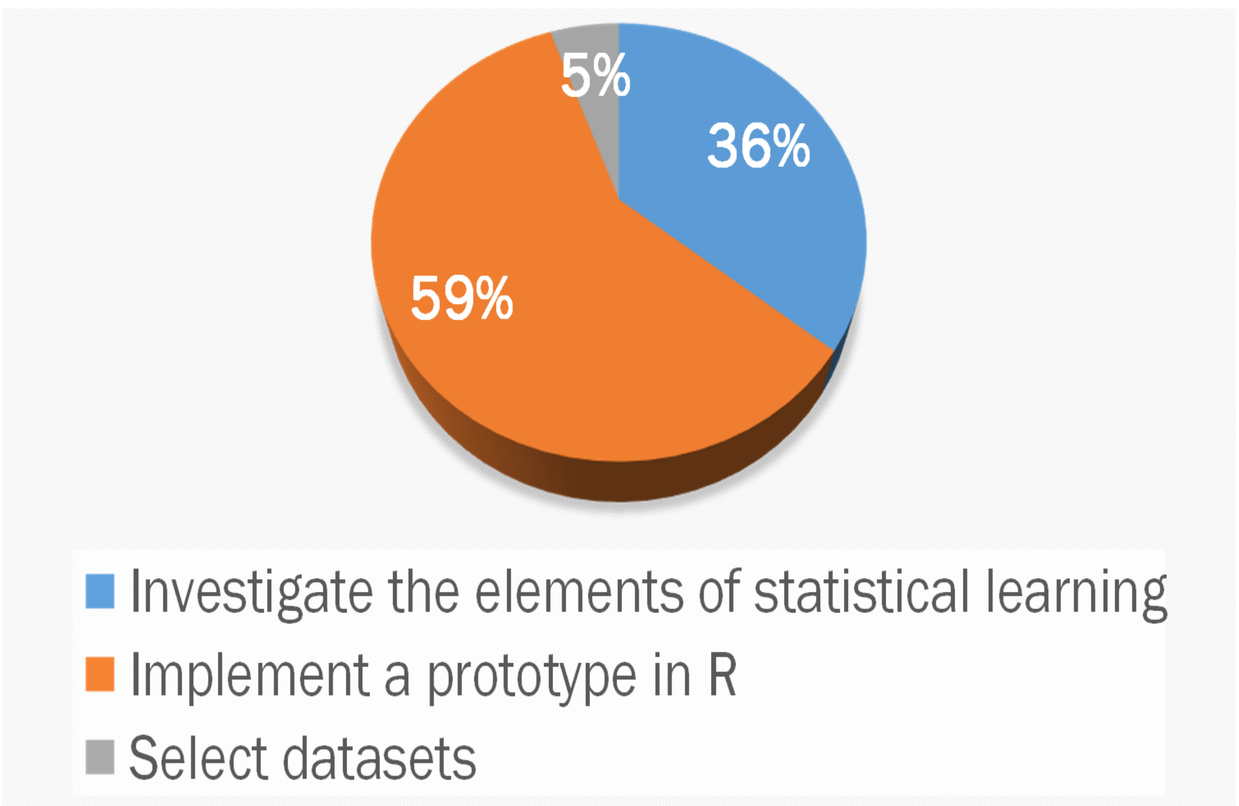
```
> source("C:\\under_construction.R")
> under_construction()
[1] "Under Construction!"
>
```

under_construction.R

```
under_construction <- function( ) {
   sprintf( "Under Construction!" );
}
```

Next Year Planning

- Continuing to investigate the elements of statistical learning (Data Mining, Inference, and Prediction).
- Implement a prototype in R programming language with the machine learning algorithms
- Select new datasets that can be used in the prototype tests.



References

[1] T. Hastie, R. Tibshirani and J. Friedman, The Elements of Statistical Learning, Stanford, California: Springer, 2008.

[2] M. Scherer, "00738578.pdf," 1998. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=738578. [Accessed 01 06 2017].

[3] R. Porter, J. Theiler and D. Hush, "06560028.pdf," Los Alamos National Lab, 2013. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6560028. [Accessed 02 06 2017].

[4] M. Serasinghe and S. Vasanthapriyan, "07924840.pdf," Sabaragamuwa University of Sri Lanka, 2016. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7924840. [Accessed 20 05 2017].

[5] E. Emamjomeh-Zadeh and David Kempe, “A general framework for robust interactive learning”. [Online]. Available: https://arxiv.org/pdf/1710.05422. [Accessed 18 03 2018].

[6] Maria Florina Balcan and Steve Hanneke. “Robust interactive learning”, 2012. [Online]. Available:http://www.cs.cmu.edu/~ninamf/papers/robust_interactive.pdf. [Accessed 18 03 2018].

[7] Steve Hanneke. “Theory of active learning”. [Online]. Available:http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.434.1762&rep=rep1&type=pdf. [Accessed 30 05 2018]