

2018 INFORMS International Conference
Taipei, Taiwan, June 17–20, 2018
© 2018 INFORMS| ISBN 978-0-9906153-1-6
https://doi.org/10.1287/infp.2018.XXXX
pp. 000–000

A Deep Learning Approach To Bank Direct Marketing

[a] ***Te-Cheng Hsu,*** [b] ***Jia-Siang Chen***
Department of Electrical Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan,
[a]andy810436@gmail.com, [b]jason06616@gmail.com

[c] ***Wei-Zhu Chen,*** [d] ***Yu-Hsuan Chien,*** [e] ***Sz-Wei Wu,*** [f] ***Mahsa Ashouri,*** [g†] ***Che Lin,*** [h] ***Galit Shmueli***
[cdg]Department of Electrical Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan.
[efh]Institute of Service Science, National Tsing Hua University, Hsinchu 30013, Taiwan,
{[c]jo5j4tp6@gmail.com, [d]eschien0401@gmail.com, [e]sandy.wu@iss.nthu.edu.tw,
[f]mahsa.ashouri@iss.nthu.edu.tw, [g†]che.lin@gmail.com, [h]galit.shmueli@gmail.com}

Abstract With the dramatically increasing amount of data available owing to fast growing financial activities, data about customers has been fast accumulating and becoming one of the most valuable asset for financial holding companies (FHC). These financial activities have created a huge amount of customer behavioral data, which is regarded as important information for direct marketing. We utilized powerful deep learning algorithms combined with oversampling and feature extraction techniques, applied to behavioral financial data, for direct marketing. This paper describes an implementation of prediction models combined with practical business and analytic goals. Real-world data used in model training and testing were collected from a Portuguese bank marketing campaign. Our model achieved a lift index of 91.76%, which is comparable to the random forest (92.03%) and better than naïve Bayes (80.89%) and logistic regression (89.72%). We find that increasing the amount of training data available improves performance of the deep learning algorithm. Therefore, we believe that our model will potentially outperform the random forest with sufficiently large samples. With this close integration of business analytics and deep learning, our model can be thought of as a prototype for direct marketing and can be finally implemented in FHC systems to deal with large amounts of data.

Keywords Deep learning, Data Mining, Direct Marketing, Deep Neural Network, Business Analytics, Lift Index, Bank Marketing

1. Introduction

In recent years, incorporating big data analytics into direct marketing has become increasingly popular due to the fast growth in rich customer-level financial data. For customer segmentation, 360-degree customer relationship management (CRM) systems are built to incorporate various kinds of customer data. In classic CRM systems, customers are segmented using their demographic information (basic customer information such as income and occupation) and financial behavior data (such as history of credit card payment) so that financial holding companies (FHC) or banks can employ this information as a basis to promote specific marketing schemes for each segment of customers.

Deep learning algorithms are now becoming popular due to the fast growth in the amount of available data as well as computing power. One of the most popular deep learning algo-

rithms is deep neural networks (DNNs). DNNs are strong on extracting high-level representations that are powerful in classification and regression (prediction) tasks. A DNN consists of multiple layers of non-linear interactions that can extract succinct representation for complex datasets. The availability of huge amounts of data provides an opportunity for building deep learning algorithms for direct marketing applications, where they can extract useful information from large and complex customer-level financial data.

We adopted an open dataset from the UCI Machine Learning Repository on a marketing campaign by a Portuguese bank for a financial product (term deposits). This dataset contains customer demographic information and information on the customers' behavior in a previous marketing campaign by the bank. We proposed to use this dataset to introduce a novel direct marketing solution that combines deep learning through a business analytics approach. Through integration of state-of-the-art data mining and machine learning techniques under a business analytics framework, we offer a solution that can boost customer satisfaction and lower marketing costs significantly by effectively identifying customers most likely to react positively to the campaign. Unlike traditional machine learning approaches that optimize accuracy or AUC, we use lift index as our model evaluation metric so that we can guarantee increasing the acceptance rate and revenues by the virtue of lift.

The paper is organized as follows. In section 2, we discuss the potential business goal, analytic goal, and corresponding performance evaluation metrics. Deep learning techniques are introduced in section 3. Section 4 shows the result of applying our approach to the Portuguese bank dataset, resulting in performance comparable with a random forest and much better than classifiers such as naïve Bayes. Section 5 discusses our results as well as future work.

2. Financial Goal, Data, and Performance Evaluation

In this section, we describe the dataset, business and data mining goals, and our choice of appropriate evaluation metric.

2.1. Bank Marketing Dataset

The dataset that we analyzed in this paper is the bank marketing dataset from UCI Machine Learning Repository (`https://archive.ics.uci.edu/ml/datasets/bank+marketing`) [9]. This dataset contains customers' credit card, previous/current campaign contact information and socioeconomic indices for 41,188 customers records (rows). For a detailed descriptions on preprocessing and feature engineering, please refer to section 4.

2.2. Aligning Business Goal with Data Mining Goal

In order to utilized prediction models in FinTech applications, we need to define potential suitable business goals and then transform them into specific data mining goals. A possible business goal for this dataset is to find the customers that will subscribe to the financial product being marketed (term deposits) given their financial behavior and sociodemographic information and previous campaign outcomes. Therefore, the data mining goal is to predict whether an existing bank customer will finally subscribe to a term deposit. More specifically, we consider the scenario of batch prediction, where for a group of customers we are interested in detecting those with the highest probability of subscribing to a term deposit.

2.3. Performance Evaluation

In typical classification tasks, accuracy, F1-score, and area under the receiver operating characteristic (ROC) curve (AUC), are the main criteria for performance evaluation. However, those criteria fail to genuinely meet practical direct marketing application goals. In practical application scenarios, we will only conduct marketing to a small subset of customers who are

predicted to be most favorable to the marketed product, as determined by our prediction model. We therefore are interested not in classification, but rather in *ranking*. In ranking, we are only interested in accurately predicting the "top-tier" records which are most likely to be positives [11]. We are interested to know whether the set of top-scoring customers from our prediction model contains a large portion of true buyers. A performance measure that captures this goal is lift.

Lift [10] has been extensively used in direct marketing applications to model customer churn and promotion targeting [1] [2]. Moreover, lift is a suitable performance measure even in extremely imbalanced datasets where the class of interest (positives) is rare, and therefore we are interested in identifying those rare records by ranking them as having high probability of being positive [8].

2.4. Lift Index

Marketing department will only send out offers to a limited set of customers they think are most likely to react positively. This limited action is aimed at reducing marketing costs and improving customer satisfaction by avoiding contacting uninterested customers. Typically, direct marketing generates customer acceptance scores, orders them from high to low, and those customers with the highest scores are contacted. The performance of such a *ranking* process can be evaluated using performance measures such as *lift* or *lift ratio* and the corresponding *lift chart* (or *gains chart*). We can write lift (ratio) as a function of n, the number of customers in the top scoring group, as

$$\text{lift}(n) = \frac{\#\text{true buyers in the top-n scoring customers}}{n}. \tag{1}$$

A lift chart can then be plotted from (1) by plotting lift(n) as a function of n (normalized by the total number of customers). *Decile lift charts* plot the average lift (numerator of (1)) for each n deciles, where each decile consists of 10% of the total customers. *Lift index* is defined as

$$\text{lift index} = \sum_{i=1}^{10} w_i \times S_i. \tag{2}$$

where S_i's are the decile lift in decile i and w_i's are prespecified weights. We can think of the lift index as an averaged lift indicating how well our model is on giving high scores to those true buyers. An example of weights can be $(w_1, w_2, \ldots, w_{10}) = (1.0, 0.9, \ldots, 0.1)$, meaning that we care more about the first few deciles which are of practical significance in direct marketing. We selected lift index as our performance evaluation metric (see section 4).

3. Deep Learning

3.1. Existing Deep Learning Techniques

Neural networks are now becoming increasingly popular. Adding depth (layers) to the architectures is beneficial for extracting high-level abstractions from raw data [5]. Depending on the application circumstances, various architectures have been designed to solve many practical problems. YouTube utilize fixed-sized feature vectors for different categories of features and then combines them using DNN in their recommender system [4]. Google Apps manipulates categorical variables to form sparse feature vectors and uses both DNN and logistic regression to make predictions [3]. However, DNNs greatly suffer from imbalanced datasets. In such cases, DNNs tend to simply label all records as the majority class, which leads to high accuracy. To overcome this issue, oversampling techniques were adopted to create a label-balanced dataset at the training phase. Synthetic Minority Oversampling TEchnique (SMOTE) and K-means can be further combined to suppress unwanted noise [7]. Restricted Boltzmann Machine (RBM) is extensively applied to extract features and was shown to be a solution to the cold start problem in recommender systems [12].

3.2. Deep Neural Network

Deep learning algorithms are now becoming overwhelmingly popular due to the dramatic growth in data availability as well as computing power. Deep learning algorithms are suitable for extracting high-level features automatically from large amount of data. In this paper, we propose a DNN-based classification model to achieve the data mining goal defined in section 2.2. DNN consists of multiple layers where each layer reads input features (data) or outputs from the previous layer as its inputs. These quantities are then fed into a non-linear activation function. An example neural network may be formulated as

$$\begin{aligned} X_l &= \sigma\left(X_{l-1}W_l + b_l\right), \\ l &= 1, 2, \ldots, L, \end{aligned} \tag{3}$$

where L is the number of layers, W_l's are the weights, b_l's are the biases, $\sigma(x) = \frac{1}{1+exp(-x)}$ is an element-wise sigmoid function, and X_L can then acts as scores of specific customers belonging to each class (true buyers and non buyers). These non-linear relationships are able to extract informative high-level features from raw data. Deep learning algorithms are reported to perform well on large and complex data; therefore, we evaluate the potential of a DNN-based model to outperform other commonly-seen powerful models, such as the random forest (see section 4.2).

3.3. SMOTE Oversampling

As mentioned previously, DNNs greatly suffer from imbalanced label datasets. SMOTE is a widely-used technique that generates synthetic minority samples. We therefore made the training set balanced by generating synthetic records for the minority class so that they have the same number as the majority one. Validation and test sets were left unmodified, reflecting the imbalanced population. The reason for not using oversampling with replacement is that we found this caused overfitting for the DNN. Undersampling was also excluded from our list since we only have a relatively small amount of data available in this dataset.

3.4. Restricted Boltzmann Machine

Restricted Boltzmann Machine (RBM) is an energy-based two-layer model. It consists of a visible layer and a hidden layer. The neurons in the visible layer and hidden layer are connected through some weights but there are no direct interactions between any two visible/hidden units. They form a bipartite graph. The pretrained hidden layer units were used to initialize our model (see section 4).

4. Direct Marketing Solution Based on Deep Learning

We adopted DNN as our predictive algorithm and we pretrained the first layer with RBM to effectively extract useful information from the raw data. SMOTE oversampling and feature engineering were performed during the training process. In the following, we describe benchmark models for performance comparison. We display lift curves as well as report their corresponding lift indices.

4.1. Benchmark Models and Dataset Preprocessing

Categorical variables are expanded into one-hot vectors using dummies. In addition to DNN, we also included L2-regularized logistic regression, naïve Bayes, and the random forest for comparison. For regularized logistic regression, naïve Bayes, and DNN, we applied SMOTE oversampling to ensure balanced sample sizes; whereas for the random forest, we left the raw data unprocessed since we have found that the random forest can deal with label imbalance naturally and achieve better performance. All columns were standardized using the training set distribution.

FIGURE 1. Lift curves.

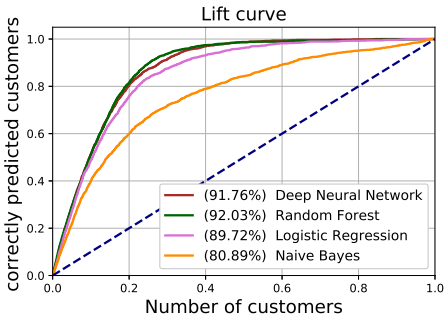


FIGURE 2. ROC curves.

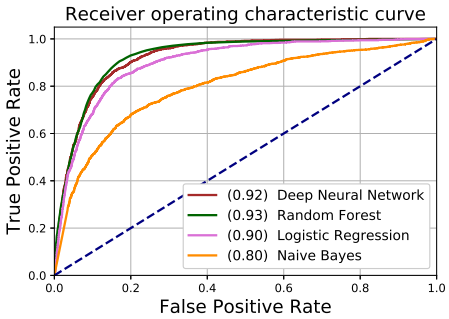


TABLE 1. Lift indices of models.

Model	Lift index
DNN	91.76%
Random Forest	92.03%
Logistic Regression	89.72%
Naïve Bayes	80.89%

TABLE 2. Lift indices for different amounts of training data (DNN).

%data	Deep Neural Network
100%	91.76%
70%	91.33%
30%	90.33%
10%	89.23%

The dataset was divided into three subsets: 60% for training, 10% for validation, and 30% for test. The ratio of label 1 and 0 was preserved in each subset (before oversampling). All models were trained using the training set (no early stop for DNN and trained until 50 epochs). Note that we only created a balanced training set and left the validation and test sets intact in order to have an estimate from the real data distribution. Grid search was performed to select the optimal set of hyper-parameters by selecting the smallest error rate model evaluated on the validation set. All the results reported were obtained through evaluation on the test set.

4.2. Model Comparison

We included naïve Bayes, logistic regression, and the random forest for comparison. Naïve Bayes and logistic regression have significant worse performance than the DNN, while the random forest had the same ranking ability as the DNN. This can be observed from the lift curves (and ROC curves, as a secondary measure) in Figures 1 and 2. The DNN achieved a lift index of 91.76% which is better than the naïve Bayes (80.89%) and logistic regression (89.72%). DNN and random forest performed equally well. This result means that the DNN performed well and is able to incorporate our business analytics goal of ranking based on financial behavioral data even on a relatively small dataset. Table 1 tabulates the lift indices for all models.

4.3. Varying Amounts of Training Data

Next, we varied the amount of training data and re-trained the DNN model. As the data increase, we observed an increase of 3% in the lift index when the amount of data increased from 10% to 100% (Table 2). This indicates that the DNN model has the potential to perform even better than the random forest if eventually we have more data. In practice, real-world financial data can indeed provide us with a huge amount of records and a richer and more complex feature set, which gives deep learning algorithms a great chance to shine.

5. Discussion

5.1. Direct Marketing based on Deep Learning

In this paper, we have build a DNN-based model for direct marketing by incorporating lift index as our evaluation metric. Unlike traditional machine learning performance evaluation, we ranked the customers using probabilities from our model to ensure that most of the true buyers are recalled. Through novel feature engineering and extraction methods, we are able to achieve performance comparable to the powerful random forest using DNN, even on a dataset with a relatively small sample size and feature set (relative to deep learning applications). Results further show that the DNN model improves with larger amounts of data. Hence, we think that our model can produce even better results when larger and richer datasets are available. We show that the DNN model can be fit into a business analytics framework and that it is flexible enough to be incorporated into existing FHCs' marketing models in order to provide increased revenues and improved customer service.

5.2. Future Work

We found that pretraining our DNN using RBM greatly improves performance. We will use an even deeper energy-based model that may possibly provide pretraining over all layers in the DNN. For instance, Deep Boltzmann Machine (DBM) [6] and Deep Belief Network (DBN) are two candidates. However, DBM is reported to have nicer properties and is much easier to train compared to DBN [5] due to the undirected edges.

Feature engineering with domain knowledge may also bring about a performance boost. Including customer segmentation information may provide nice representations for customers and therefore the ranking task will become much easier.

Bootstrapping is now becoming a popular technique to produce robust predictions. With pretraining, we are able to speed up DNN training and hence boostrapping becomes faster to execute. Such ensemble methods will make the rankings for customers more robust where the rankings will not vary a lot when trained with different data partition.

References

[1] Bekkar M, Djemaa HK, Alitouche TA (2013) Evaluation measures for models assessment over imbalanced datasets. *J Inf Eng Appl* 3(10).

[2] Burez J, Van den Poel D (2009) Handling class imbalance in customer churn prediction. *Expert Systems with Applications* 36(3):4626–4636, ISSN 09574174.

[3] Cheng HT, Koc L, Harmsen J, Shaked T, Chandra T, Aradhye H, Anderson G, Corrado G, Chai W, Ispir M, Anil R, Haque Z, Hong L, Jain V, Liu X, Shah H (2016) Wide & Deep Learning for Recommender Systems .

[4] Covington P, Adams J, Sargin E (2016) Deep Neural Networks for YouTube Recommendations .

[5] Goodfellow I, Bengio Y, Courville A (2016) *Deep Learning* (MIT Press).

[6] Goodfellow I, Mirza M, Courville A, Bengio Y (2013) Multi-prediction deep boltzmann machines. *Advances in Neural Information Processing Systems*, 548–556.

[7] Last F, Douzas G, Bacao F (2017) Oversampling for Imbalanced Learning Based on K-Means and SMOTE .

[8] Ling CX, Li C (1998) Data mining for direct marketing: Problems and solutions. *KDD*, volume 98, 73–79.

[9] Moro S, Laureano R, Cortez P (2011) Using data mining for bank direct marketing: An application of the crisp-dm methodology. *Proceedings of European Simulation and Modelling Conference-ESM'2011*, 117–121 (Eurosis).

[10] Piatetsky-Shapiro G, Masand B (1999) Estimating campaign benefits and modeling lift. *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, 185–193 (ACM).

[11] Shmueli G, Bruce PC, Yahav I, Patel NR, Jr KCL (2017) *Data Mining for Business Analytics: Concepts, Techniques, and Applications in R* (John Wiley & Sons), ISBN 978-1-118-87936-8, google-Books-ID: ETwuDwAAQBAJ.

[12] Yedder HB, Zakia U, Trajković L (2017) Modeling prediction in recommender systems using restricted boltzmann machine .