Link

Paper link - `https://arxiv.org/pdf/1609.02489.pdf`

Brief

Trying to predict whether the customer will purchase a product or not by supplying the product's data both image and attributes.

Authors

Christian Bracher, Sebastian Heinz, Roland Vollgraf - Zalando Research

Date

8 September 2016

Input

1. Image 177×256
2. "expert labels" such as color, pattern, fabrics composition etc. The price labels were created by k-mean clustering. Label information was preprocessed into one-hot vectors.

Label

$\Pi_{ij} \in \{0, 1\}$ for i-th articles, j-th customer. 0 - hasn't bought, 1 - has bought. The data is devided into 4 parts (refer to figure 1).

1. training data for both customer and articles - Π^{tt}
2. validation data for articles feature (articles unseen, trained customers). - Π^{vt}
3. validation data for customer parameter (customers unseen, trained articles). - Π^{tv}
4. all validation (new article, customers from Π^{vv})

Network

For better understanding, please refer to figure 4

1. **Attribute network**
   4-layers connected neural network supplied by one-hot attributes(labels). The target is to extract the features from labels.

2. **Image network**
   CNN network, Alexnet according to the paper.

3. **Combined network**
   From 2 models above, concatenate the extracted feature and pass them through FC-256, Relu, dropout-0.2 then the final value is called "fDNA" (simply feature). Let's the input data (image, labels) is ϕ_i and the θ is the parameters. Therefore f_i (i-th fDNA for i-th article) is $f_i = f(\phi_i, \theta)$

Prediction & Loss

Assume that we extracted the fDNA f_i from the combined network, we model the probability of purchase the article-ith from customer-jth by

$$p_{ij} = \sigma(f_i \cdot w_j + b_j)$$

where w_j, b_j is a factor associated with customer j-th (Each customer has their own parameters). The loss is calculated by the mean cross entropy loss.

Training

1. **Training**
   Use the training data Π_{tt} to update both network weight θ and customer weight w_j, b_j

2. **Article Validation**
   Straightforwardly pass the article validation data Π_{vt}.

3. **Customer Validation**
   Freeze the network parameter θ but update the the customers weight w_j, b_j by passing data Π_{tv}.

4. **All validation**
   Similar to part 2, we pass data Π_{vv}. This step is for judging whether that the customer validation from part 3 generalizes well to unseen articles.

Evaluation

Because the ratio of purchase is very low $E[P(\Pi_{ij} = 1)] = 1.14 \times 10^{-4}$, the overall prediction quality should not be determined by 0 or 1 instead it can be expressed by receiver operating characteristic (ROC) analysis (refer to Fig 6) and the score is the area under the ROC curve, called AUC score (refer: `https://www.medcalc.org/manual/roc-curves.php`)