Multi-Market Deal Size Prediction

Evelin C. F. de Amorim
Departamento de Ciência da Computação
Universidade Federal de Minas Gerais
evelinamorim@ufmg.br

ABSTRACT

Daily-deals sites (DDSs) are web portals that offer discount coupons for services or products. Thus, customers are attracted by the captivate discounts. DDSs gain profit by the amount of coupons sold and merchants, who offer services and products, gain profit by increasing the amount of clients . However, DDSs know beforehand the number of coupons sold, can benefit featured offers and bussiness strategies.

Deal size prediction is the name of the task that predicts the number of coupons sold in a given offer. Some research proposed descriptive models to deal size, nonetheless few attempts have been made in order to predict deal size. Due to its practical importance, we propose a novel method to predict deal size. This method is based on an existing method, which considers latent markets in given catalog of deals. The existing method assigns only one latent market for a deal, and we propose to assign the deal to the most likely markets.

We perform experiments in order to detect only error in regression phase. The competition model is still in development, but we explain the competition model we intend to implement . The results show evidences that include all probable markets, for a given deal, increase error. Also, we test another strategy that consider only markets with at least 0.3 of probability to own a given deal.

1. INTRODUCTION

Daily-deals sites (DDSs) are popular web portals that offer discount coupons for services or products. These coupons includes expiration date, therefore it is not available for new customers acquire after a short period of time. As DDSs coupons are available for limited time, it is usual customers miss coupons. Also, customers have decided soon which coupon to buy. Considering that the customers own a limited budget, some coupons will be disregarded in favor of others coupons.

Coupons that provide similar services or products can bring doubt to consumer mind. Thefore, the number of coupons sold of one deal can affect the number of coupons sold of another similar deal. From this ideia, Lacerda et. al [3] proposed a strategy, called *Competitive Business Market Prediction* (CPMB), that assigns each deal in the DDS catalog to a latent market. CPMB strategy is composed by three main steps: the first step divides the catalog of deals into markets by using Latent Dirichlet Allocation (LDA); the second step perfoms regression training in each market; and the third step executes an expectation-maximization procedure in order to consider the competition among the deals in the same market.

Byers et. al [1] performed an experiment in order to evaluate the impact of online social network in the sales of discount coupons of DDSs. Also, Byers builds a linear model to depict the deal size according to the deal features. The authors conclude that predicting deal size is challenging and future research can explore merchant information from reviews web sites.

Although there is practical importance of deal size prediction task, there are a minimal amount of research. For these reason, we proposed a method based on Lacerda's research in order to achieve better results. Nevertheless, our proposal supposes that deals can belong to more than one market. The results show that consider all markets probably introduce predictions errors in deal size prediction and deal only with more likely markets reduce such predictions errors.

2. METHODOLOGY

The methodology that we propose is comprised by three steps: separation of the deals of the catalog into markets, regression training of the deals, and computing the competition between deals in the same market. Althought, our steps are very similar to Lacerda's steps, the first step is modified in order to assign each offer to the most likely markets.

The following subsections explain the methodology in details and the evaluation metric employed to evaluate our results.

2.1 Algorithm

The algorithm is composed by three phases: Market identification, Regression per Market, and Market Competion Weighting. Next, we will describe how these three phases are implemented.

Market Identification. Each deal presents textual features like: title, description, merchant, and so on. These textual features are extracted of each deal and tokenized. NLTK[1] library performs the tokenization of each textual feature of a deal. Next, stop words are removed and a stemmer are applied to tokens. Both stop words and stemmer processes are perfomed by NLTK library, which provides a constant list of words for stop words and the Porter's algorithm for stemming. Also common tokens in the catalog is removed, by common tokens we consider tokens that appear in at least 25% of documents.

After process textual features, Latent Dirichlet Allocation (LDA) method is applied in order to identify the markets. LDA method build probabilistic distributions in order to infer latent topics in a given collection. The probabilistic distribution are assembled based on a defined weight assigned to words. The weights can be, for instance, term frequency in the document. The weigthing scheme in this research is *term spread*, which is defined by the Equation 1.

$$TS(t,d) = \sum_{s \in d} i, \text{where i} \begin{cases} 1 & t \in s \\ 0 & otherwise \end{cases} \quad (1)$$

The implementation of LDA executed in experiments is from Řehůřek and Petr Sojka [4].

Regression per Market. After the processing of textual features, each deal is assigned to the most probable markets. Next, one regression model is trained in historical data for each market. The regression method is trained in the following deal features:

- City of the deal;
- The absolute value of the discount;
- Wether the deal sold out;
- Number of the days which the deal is available for costumers;
- Day of week that the deal starts;
- Category in the DDS web site.

Each regression model predicts the deal size of the test set, when one deal belongs to more than one market, then a weighted mean is computed according to Equation 2, where $\sigma(d,t)$ is the predicted value for the deal d, if the deal belongs to market t.

$$pred(d) = \sum_{d \in S} p(S) \times \sigma(d, S) \quad (2)$$

Prediction step is applied only to deals of the day and we employ and SVR as regression method. There are as many SVR's classifiers as latent markets.

[1] www.nltk.org

Market Weighting. After the building of regressor classifiers, weights are computed to consider the impact of the market competition in deal size prediction. The processes of computation of market weights start with the partition of the catalog into k markets, so that $S = \{S_{m_1}, S_{m_2}, \cdots, S_{m_k}\}$. Each deal $q \in S$ belongs to one or more markets, and we call S_q the subset of markets that owns q. In this step we consider the prediction as an weigthted mean of predictions. For each market $S_q^j \in S_q$ we compute the Equation 3.

$$\sigma^j(q) = \alpha \times \frac{\sum_{\forall d^j \in S_q^j} f(d^j) \times \rho(S_q^j)}{|S_q^j|} + (1 - \alpha) \times f(q) \quad (3)$$

The variable $\rho(S^j)$ is unknown and is estimated by the expectation-maximization (EM) algorithm. The Equation 4 calculates ρ in each iteration of EM procedure.

$$\rho(S_q^j) = \frac{\sum_{\forall d^j \in S_q^j} \sigma(d^j)}{\sum_{\forall d \in S} \sigma(d)} \quad (4)$$

Then, the prediction for a given deal q is depicted by the weighted mean of σ^j predictions for q. The weight is, again, the probability of q belongs to the market S_q^j.

The loss function for the expectation-maximization is *Root Mean Square Error*(RMSE), which is defined by Equation 5.

$$RMSE(y, \sigma) = \sqrt{\frac{1}{n} \sum_{1 \geq i} (y_i - sigma(q))^2} \quad (5)$$

, where $\sigma(q)$ is the weithed mean of $\sigma^j(q)$ and n is the number of predicted deals.

2.2 Evaluation

There are many ways to evaluate recommender systems. However, our task is not the recommendation itself but a support task for recommender systems. Also, the sizes of the deals can vary greatly , hence to evaluate our prediction, we normalized our dataset using standard deviation of whole collection. Otherwise, our predictions errors reach high values and we are unable to identify the potential of our proposal. We adopt the *Root Mean Square Error* (RMSE) as evaluation metric, which is defined by Equation 5.

3. EXPERIMENTS AND RESULTS

The implementation of this research produced a different outcome reported by the original [3]. The reasons for this outcome may vary: different parameters for griding search in SVR, different dataset, differences in the features of dataset, details about expectation-maximization from Lacerda's thesis, and so on. In face os these issues, we performed alternative experiments. The following sections describe the dataset employed and the experiments performed.

3.1 Dataset

Dataset employed in the experiments was provided by Byers et. al [1]. There are two DDSs web portal in Byer's collection: Groupon and Living Social. As we demand textual features, and Byers collection's has only structural features of the deals, we crawled the web pages from mentioned DDSs portals. Nonetheless, some deals are unavailable, therefore the number of the deals we tested is different from the numbers reported by Byers.

The deals from living social dataset were collected between March 21st and July 3rd, 2011. The textual features from these dataset were crawled in September 24th, 2014. From the original dataset, we excluded nineteen deals, resulting in 2590 deals. The deals from groupon were tested, however due to time constraints only one test was performed. From the original groupon dataset, we excluded five deals, resulting in 16685 deals.

We depicted the distribution of deal size of both datasets in Figure 1. The distributions of deal sizes resemble an exponential distribution: few deals with low deal sizes.

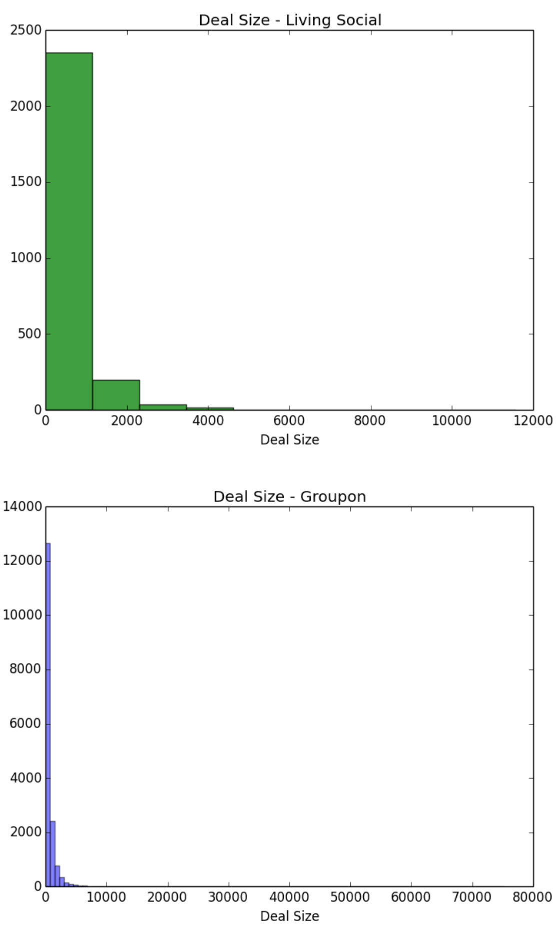

Figure 1: Distribution of Deal Size in the Dataset

3.2 Experiments

Due to the time constraints and difficulties in the interpretation of the strategy's description, it was not possible perform the experiments as we wish.

Lacerda employed an SVR classifier for each market and then an EM method. Also, Lacerda performed a grid search to find the best SVR classifier for each market and SVR rbf kernel was employed. However, we considered linear kernel for grid search as well. The linear kernel performs efficiently in datasets that are linearly separable. The parameters for our grid search were $C = \{0.01, 0.1, 1, 5, 10, 50\}$ and $\gamma = \{0, 0.001, 0.0000001\}$.

The EM procedure was not clear in Lacerda's P.h.d. thesis. The definition of the EM method [2] comprises the likelihood of an mixture statiscal distribution. However, the definition provided is unclear whether the likelihood is applied in the method. Although the description of the marketing competition phase is vague, we attemped to implement the procedure described. Nevertheless, we failed in reproducing the EM step.

Despite of the problems in reproducing Larceda's research, we attempt to evaluate the RMSE for SVR strategy. The ideia is to test the hypothesis that employing best market to predict deal size can be improved considering all the possible market or considering the markets above a threshold.

Table 1 shows the average of RMSE for thirty markets, nine executions for each strategy and 95% of confidence for Living Social Dataset. Also, the deal size are normalized by standard deviation of whole dataset. The threshold tested was 0.3, i.e., only the markets that has the probility of 0.3 or above to one given deal is considered in the weighted mean of such deal. The difference between Best-Market Strategy and All-Market is inconclusive, as well as the difference between Best-Market and All-Market. However, the difference between All-Market and Threshold-Market is significant. The threshold of Threshold-Market probably removes unlikely markets, and then it reduces errors that All-Markets strategy includes when consider unlikely markets.

Table 1: RMSE for all the strategies and 30 markets in Living Social Dataset

Strategy	RMSE
Best	1.06
All	1.12
Threshold	0.99

The Groupon dataset was tested only once, therefore the results in Table 2 are just for curiosity purposes.

Table 2: RMSE for all the strategies and 50 markets in Groupon Dataset

Strategy	RMSE
Best	0.87
All	1.07
Threshold	1.00

The results we obtain are slightly better that Lacerda reports, however our dataset is different and we also experimented linear kernel in our grid search.

The Figure 2 presents a graph of the error for different values of markets in Living Social Dataset. The behavior of both strategies are similar to the Lacerda's strategy. We observe that the behavior of *Best* strategy is consistently better than *All* strategy. While the *threshold* strategy is better than *All* strategy. The *Best* strategy perform better than the other two strategies.

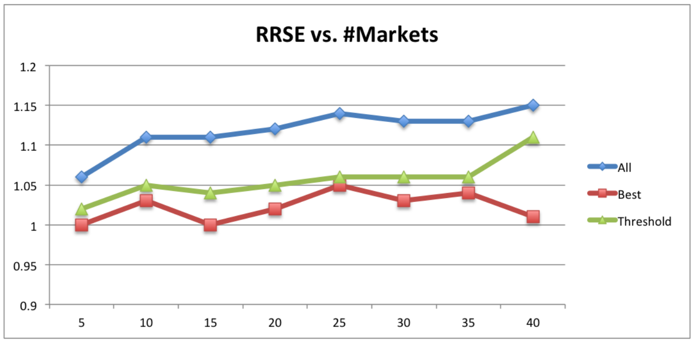

Figure 2: RMSE vs. #Markets for Living Social Dataset

The errors for each topic in one execution in Living Social dataset is presented by Table 3.2. In some markets, All strategy performed well, this usually happens when LDA defined a miscellaneous market and also there is a high standard deviation in such markets. However, in most markets Threshold and Best Strategy are superior. Also, in all execution always there are markets that own more deals than others markets. This is expected, since the dataset itself present some numerous categories, and although category is different from the topic concept, it is an evidence that the dataset is unbalanced.

4. CONCLUSION AND DISCUSSION

We proposed a modification in Lacerda's research in order to obtain better results. However, our strategy seens simple to consider the participation of one deal in different markets. It is possible that the errors of all markets introduce more errors in the final result.

More tests must be perfomed as well. The groupon dataset was not tested and more tests can produce more insightful clues about more sophisticate techniques, which can consider multi-market deals but avoid to introduce errors like the strategy proposed here.

Although the strategy here may be introducing errors in the regression prediction, we may investigate whether is this really the case. The threshold strategy also presented an improvement over the initial hypothesis that all likely markets can contribute to deal size prediction. Also, despite the fact that the difference between Best-Market strategy and Threshold-Market strategy is not significant, the Threshold-Market performed consistently well over all executions. These results encourage us to formulate more elaborate strategies that consider multi-market deals.

Table 3: Error by Topic in one Execution in Living Social Dataset

Topic	#docs	Best	All	Threshold
0	95	0.57	0.65	0.54
1	64	0.89	1.11	1.10
2	27	1.08	1.24	1.17
3	35	0.48	0.56	0.39
4	35	1.25	1.43	1.56
5	61	2.01	2.08	2.24
6	24	0.78	0.88	0.73
7	32	0.43	0.56	0.55
8	95	0.43	0.50	0.41
9	13	2.07	2.17	2.19
10	56	0.43	0.47	0.37
11	67	1.16	1.27	1.27
12	21	1.01	1.24	1.26
13	36	0.91	1.05	1.12
14	59	0.84	0.86	0.80
15	260	1.01	1.26	1.11
16	93	1.13	1.27	1.30
17	32	0.64	0.89	0.77
18	17	1.93	1.99	0.55
19	22	0.77	0.73	0.54
20	59	0.42	0.53	0.43
21	25	0.53	0.65	0.64
22	24	1.16	1.33	0.65
23	23	1.19	1.08	0.89
24	130	1.64	1.61	1.63
25	38	0.86	1.00	0.53
26	48	0.50	0.63	0.38
27	25	0.55	0.70	0.68
28	37	1.44	1.41	1.59
29	31	0.68	0.74	0.71

The Table 3.2 shows that the markets own few deals to train in SVR. The errors can be higher due to few deals in each SVR has to consider in the training step. In this case, the groupon dataset can present different results since there are more deals and also more features to SVR.

Also, the vague description of the expectation-maximization strategy of Lacerda's P.h.d. thesis, become difficult to implement the strategy in time. Maybe if we employ the multi-market strategy in EM algorithm, the outcome can be improved, even though the multi-market strategy performed worse compared to best-market for the regression module.

More experiments are being made, like the variation in the threshold of the *threshold* strategy and Groupon dataset.

5. REFERENCES

[1] J. W. Byers, M. Mitzenmacher, and G. Zervas. Daily deals: prediction, social diffusion, and reputational ramifications. In *Proceedings of the fifth ACM international conference on Web search and data mining*, WSDM '12, pages 543–552, New York, NY, USA, 2012. ACM.

[2] T. Hastie, R. Tibshirani, J. Friedman, T. Hastie, J. Friedman, and R. Tibshirani. *The elements of statistical learning*, volume 2. Springer, 2009.

[3] A. Lacerda, A. Veloso, R. L. Santos, and N. Ziviani.

Context-aware deal size prediction. In *String Processing and Information Retrieval*, pages 256–267. Springer, 2014.

[4] R. Řehůřek and P. Sojka. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta, May 2010. ELRA.