Learning to Predict the Cost-Per-Click for Your Ad Words

Chieh-Jen Wang and Hsin-Hsi Chen
Department of Computer Science and Information Engineering
National Taiwan University
Taipei 106, Taiwan
{d96002, hhchen}@csie.ntu.edu.tw

ABSTRACT
In Internet ad campaign, ranking of an ad on search result pages depends on a cost-per-click (CPC) of ad words offered by an advertiser and a quality score estimated by a search engine. Bidding for ad words with a higher CPC is more competitive than bidding for the same ad words with a lower CPC in the ad ranking competition. However, offering a higher CPC will increase a burden on advertisers. In contrast, offering a lower CPC may decrease the exposure rate of their ads. Thus, how to select an appropriate CPC for ad words is indispensable for advertisers. In this paper, we extract different semantic levels of features, such as named entities, topic terminologies, and individual words from a large-scale real-world ad words corpus, and explore various learning based prediction algorithms. The thorough experimental results show that the CPC prediction models considering more ad words semantics achieve better prediction performance, and the prediction model using the support vector regression (SVR) and features from all semantic levels performs the best.

Categories and Subject Descriptors
H.3.5 [Information Storage and Retrieval]: Online Information Services – commercial services; H.2.8 [Database Management]: Database Applications — data mining; J.4 [Social and Behavioral Sciences]: Economics

General Terms
Algorithm, Performance, Measurement, Economics

Keywords
CPC prediction, ad ranking, search engine optimization.

1. INTRODUCTION
The commercial values of ads for advertisers depend on their click through rates (CTRs). In sponsored search, an ad CTR is connected with a notion of ad ranking. That is, ad clicks usually occur at the first and the second ranks [4]. In other words, the ad CTRs decrease when ads are shown at lower ranks on search result pages for the sake of reduced visual attention. An ad rank is determined by a function [13] shown as follows:

$$RankScore = QualityScore \times CPC$$

An ad with a higher *RankScore* will be shown at a higher rank. Two factors determine a *RankScore*. One is the *QualityScore*, which depends on various factors in a search query, ad words, ad texts, etc. Some methods [4][13] extract features such as terms in ad contents, ad types, etc. to estimate the *QualityScore*. The other is the *CPC* of ad words, which is bidden by an advertiser. The CPC shows how much the advertiser is willing to pay a publisher (e.g., a search engine) for each click on the ad. Concretely, when a user clicks on the sponsored link on search result pages, the user is directed to the advertiser's landing page. The publisher then charges the advertiser for sending the user to the landing page. The price is called cost-per-click (CPC). In sponsored search, search engines need an auction mechanism for ad words and the mechanism is usually based on the generalized second-price (GSP) auction. That is, if a user clicks on a sponsored link at position *k* on a search result page, the sponsored advertiser is charged by the search engine an amount equal to the next highest bidding, i.e., the bidding of the advertiser at position (k+1).

When advertisers plan to advertise consumers via the advertising services of search engines, they have to select the CPCs of their ad words. The CPC of ad words is an important factor due to its direct impact on the *RankScore* in terms of ad ranking. How to select an appropriate CPC of ad words for maximizing benefit under a budget constraint is indispensible [9]. Advertisers may suffer from the "stupid money" called by the Business Week if they bid ad words too intensively. To deal with this issue, advertisers might be good to shift their budgets to other ad words of similar advertising objectives. The cost and the effectiveness of ad words are extremely hard to predict for advertisers knowing nothing about the bidding reference.

Some Internet advertising services provide estimated CPCs for ad words which have been offered, but not all ad words have been recorded for reference. In this paper, we propose various models to predict a CPC of new ad words with features extracted from a large-scale real-world ad words corpus. Features on three different levels including the named entity, the topic terminology and the word levels are proposed to capture different types of semantics in ad words. Features on the named entity level represent the named entities in ad words identified and categorized by using AlchemyAPI [1]. Features on the topic terminology level represent topic terminologies and senses in ad words mined by using WordNet [15]. Features on the word level describe the information from individual words themselves.

Moreover, we investigate which feature levels have more effects on predictions. In other words, we would like to know whether features containing more ad words semantics are more accurate in the CPC prediction or using only words surface features is still workable. Various CPC prediction models learned by different prediction algorithms with features extracted from different semantic levels are explored. The experimental results show that the CPCs of ad words are related to ad words semantics after testing on a large-scale real-world ad words corpus. The CPC prediction model using the support vector regression (SVR) and features from all semantic levels performs the best.

The rest of this paper is organized as follows. The related work is presented and compared in Section 2. The experimental dataset used in this study is described in Section 3. The semantic feature representations and the CPC prediction models are introduced in Section 4. Experimental results are shown and discussed in Section 5. Lastly, Section 6 concludes the remarks.

Copyright is held by the author/owner(s).
CIKM'12, October 29– November 2, 2012, Hawaii, USA.
ACM 978-1-4503-1156-4/12/10...$15.00.

2. RELATED WORK

Many related literatures have been proposed to analyze the ad words auctions in the economics and the computer science community. Most of the work focused on the bidding strategy of ad words rather than the prediction. Bajari and Hortascu [5] studied the factors which may affect the bidding price and utilized econometric techniques to build bidding models. Animesh *et al.* [3] explored different profiles of bidding strategies and examined the relationship between advertisers' bidding strategies and their firm quality in the online advertising. Edelman and Ostrovsky [7] aimed at strategic behavior analyses of the ad words auction on Yahoo and Google. Kitts and Leblanc [10] estimated a distribution of the final prices in the ad words auctions and then used the distribution to optimize bidding strategies. Cary *et al.* [6] used a greedy bidding strategy for the ad words auction to balance revenue, convergence and robustness properties of bidding strategies.

Many studies with respect to the price prediction of item auctions such as PDA [8], digital camera [12], etc. were proposed in the online auction. Wellman *et al*. [14] introduced a game theory-based trading agent (TA) for the price prediction in a travel domain like airline, hotel, and ticket prices. They created several virtual participating TAs. Each TA took into account evidences from different sources for bidding. All TA simulated competitive bidding via a random walk process. The prices were generated for prediction if the competitive bidding was done. In Internet ad campaign, literatures on the CPC prediction of ad words were comparatively fewer. The study proposed by Kitts and Leblanc [10] was a work that is similar to the CPC prediction implicitly. They simulated the pay-per-click (PPC) auctions of ad words based on the TA technology. The TAs considered various factors such as the advertiser's specifications, the clicks estimation, the markets estimation, etc. to simulate the competitive bidding process for the PPC prediction. The CPC and the PPC are nearly synonymous but the CPC is more often used when referring to the Internet advertising. The PPC is a pay by a search engine to another publishers (e.g., personal websites, blogs, etc.) when an ad automatically scattered by the search engine on the publisher websites is clicked. The search engine will split the earnings from the CPC to the PPC for publishers. However, the study in the PPC prediction of ad words was performed with artificially generated data and was not verified on the real-world data.

As mentioned before, the key challenge for advertisers is to select an appropriate CPC of their ad words. This paper proposes a learning-based approach to deal with this issue. The contributions of this work are three-fold. First, the CPC prediction model can overcome the *cold-start* problem, where advertisers wish to refer CPCs of their ad words that no one has yet been recorded in a database for the decision support. To the best of our knowledge, our work is the first attempt to propose prediction models for selecting CPCs of new ad words. Second, we employ the natural language processing (NLP) techniques to mine ad words semantics and construct learning-based models for the CPC prediction. Our proposed prediction models can precisely predict the CPC of ad words without referring to external auction factors, i.e., the starting price, the bid increment, the reserve option, etc. Finally, we explore and analyze the CPC prediction of ad words from an analytical and an empirical perspective. In other words, we verify the prediction models empirically by a large-scale real-world ad words corpus obtained from a major search engine. The results show that the prediction models achieve promising results.

3. AN AD WORDS CORPUS

An ad words corpus consisting of 2.43 million pairs of unique ad words and the corresponding CPCs from a major search engine is adopted. The CPC is the least amount of money to show ads at the first rank on search result pages based on the GSP auction. In an ad words-CPC pair, ad words contain different levels of semantics that may affect the determination of CPC. Consider some examples. Two pairs of ad words denoting the same theme at different locations have different CPCs, e.g., ("car accident lawyers Los Angeles", 79.5) and ("San Francisco car accident lawyers", 72.94), where the first and the second elements in the pairs denote ad words and its corresponding CPC, respectively. Similarly, two pairs of ad words denoting different themes at the same locations have different CPCs, e.g., ("Los Angeles drunk driving lawyer", 80.71) and ("motorcycle accident lawyer Los Angeles", 58.45). Clearly, "DUI lawyers Southern California" and "birth injury lawyers Philadelphia" touching on different themes at different locations have different CPCs, i.e., 82.26 and 71.44, respectively. Thus, segmenting ad words into a group of meaningful units, finding their senses, and measuring the relatedness of CPCs with the meaningful units and their senses are three major steps in the CPC prediction.

4. CPC PREDICTION MODELS

In this section, we propose several CPC prediction models which integrate various prediction algorithms and features extracted from the ad words corpus with various semantic considerations.

4.1 Representation of Ad Words

We may have several alternatives to segment ad words into meaningful units. For example, the ad words "Los Angeles drunk driving lawyer" consists of 5 individual words, i.e., 1 named entity ("Los Angeles"), and 3 topic terminologies ("drunk driving lawyer", "drunk driving", and "lawyer"). The following describes how to identify them and their types/senses.

Named Entity Level (N). Named entities (NEs) such as people, companies, organizations, cities, etc. are key components in ad words. We employ AlchemyAPI to identify NEs from ad words, and tag each NE with type(s) along with a relevance score. AlchemyAPI uses a two-level hierarchal structure to categorize each NE into NE type(s). For example, the NE "Blizzard" is identified as a type "Company" and a subtype "Video Game Publisher" of the type "Company". Total 815 NE types defined in AlchemyAPI are used. After named entity recognition, total 143,006 NEs are identified. Table 1 shows the top 10 highly-frequent NE types. The location related NE types frequently appear in ad words. That reflects the location is a key factor for advertisers to organize their ad words. In particular, the NE type with respect to disease is at the seventh rank. That demonstrates that people pay more and more attention to the health care so that many health related NEs are shown in ad words. The weights of NE features are set to binary, while the weights of NE type features are determined by their relevance scores.

Topic Terminology Level (T). In ad words bidding, advertisers offer CPCs based on the topics of advertising objectives like "car accident", "drunk driving", etc. We remove those words marked with NE tags in the ad words corpus, and then employ AlchemyAPI again to identify topic terminologies. The extracted topic terminologies form a lexicon. Two selection strategies – say, the longest first (LF) and all the combinations (ALL), are used to segment ad words by using the lexicon. In the above example, only a topic terminology ("drunk driving lawyer") is selected by the LF strategy, but three topic terminologies ("drunk driving lawyer", "drunk driving" and "lawyer") are selected by the ALL strategy. WordNet 3.0 is employed to determine the senses of topic terminologies. Here, we postulate that the head of a topic terminology is always at its final position. Then, we consult WordNet with the head of each topic terminology after stemming, and assign it a synset. In this way, a set of topic terminologies with the same synset are put into the same sense cluster. Finally, there are 273,721 topic terminologies and 8,153 sense clusters. The weights of topic terminology features are set to binary. In contrast, a topic terminology may be ambiguous, so that the weights of the sense cluster features are determined by a normalized frequency count of all the matching synsets.

Word Level (W). Every word in ad words forms a feature. The weights of word features are set to binary. Besides, the length of ad words is also considered as a feature.

4.2 Prediction Algorithms

Several machine learning algorithms such as the linear regression (LR) [8], the polynomial regression with degrees 2 and 3 (abbreviated as PR(2) and PR(3), respectively) [8], the regression tree (RT) [8], the genetic programming (GP) [11], the neuro-fuzzy (NF) [12], the back-propagation neural network (BPN) [12] and the support vector regression (SVR) [2] have been adopted to predict the auction prices in the online auctions. Various CPC prediction models are constructed by these prediction algorithms and combinations of features extracted from different semantic levels as mentioned in Section 4.1. For RT, the parameter of minimum number of observations is set to 10. For GP, the population size and the run generations are set to 2000 and 25, respectively. The elitism selection scheme is used and the fitness is measured the mean absolute percentage error (MAPE). For NF, the fuzzy associative memory (FAM) is adopted to determine its parameters, including the relative importance of each fuzzy rule and the membership function. For BPN, several options of the neural network configurations are tested, and the hidden layer, learning rate and momentum are set to 1, 0.8 and 0.2, respectively. For SVR, the RBF kernel is used and the grid algorithm is adopted to determine the best parameters: *c*, *gamma* and *ε*.

5. EXPERIMENTS AND DISCUSSIONS

We randomly partition the ad words corpus into two equal sets for training and testing. Three evaluation metrics, i.e., mean square error (MSE), mean absolute error (MAE), and mean absolute percentage error (MAPE), defined as follows are considered.

$$MSE = \frac{\sum_{i=1}^{n}(G_i - P_i)^2}{n},\ MAE = \frac{\sum_{i=1}^{n}|G_i - P_i|}{n},\ MAPE = \frac{\sum_{i=1}^{n}|G_i - P_i|/G_i}{n}$$

where G_i is the real value of sample i, P_i is a predicted value of sample i, and n is the number of samples.

Various CPC prediction models constructed by different prediction algorithms, combinations of features extracted from different semantic levels, i.e., named entity (N), topic terminology (T) and word (W), and two selection strategies, i.e., the longest first (LF) and all the combinations (ALL) are explored. Using features on the word level (W) only is a baseline. Features extracted from different semantic levels are compared. Figures 1-3 show the performance of CPC prediction models generated by different prediction algorithms and feature level combinations on MSE, MAE and MAPE evaluation metrics, respectively. The tendency is similar on the three evaluation metrics. For the six feature combinations, using SVR performs the best among all the prediction algorithms on the three evaluation metrics significantly (p-value<0.001 in t-test). The results reflect integrating features on the three levels (W+N+T) with the ALL selection strategy

Table 1. Top 10 frequent NE types

No	Type	Subtype	Frequency
1	Country	Location	13,051
2	Country	Region	9,733
3	Person	TV Actor	9,639
4	Person	Film Actor	5,846
5	Person	Athlete	5,242
6	Person	Musical Group Member	5,148
7	Health Condition	Disease	4,402
8	Geographic Feature	Location	4,330
9	Company	Operating System Developer	4,278
10	Company	Video Game Publisher	4,008

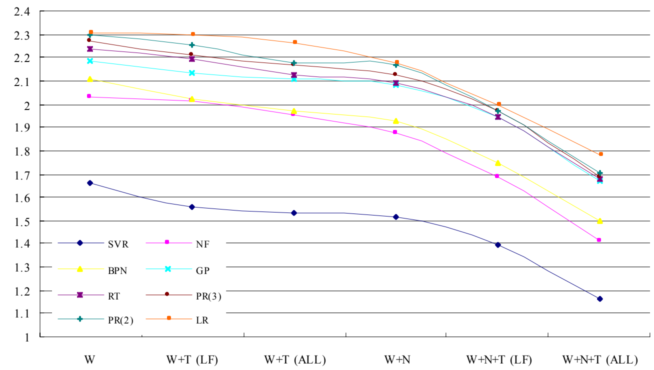

Figure 1. MSE performance on different prediction models

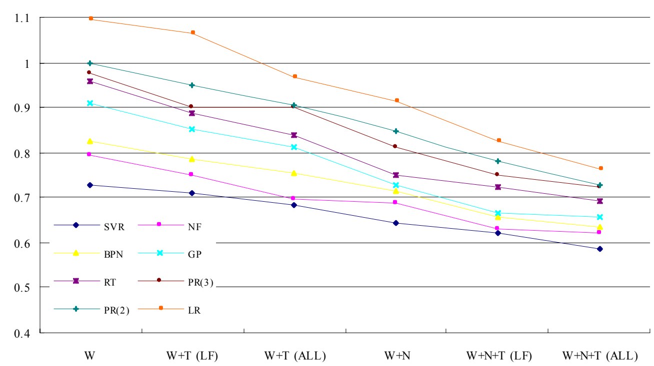

Figure 2. MAE performance on different prediction models

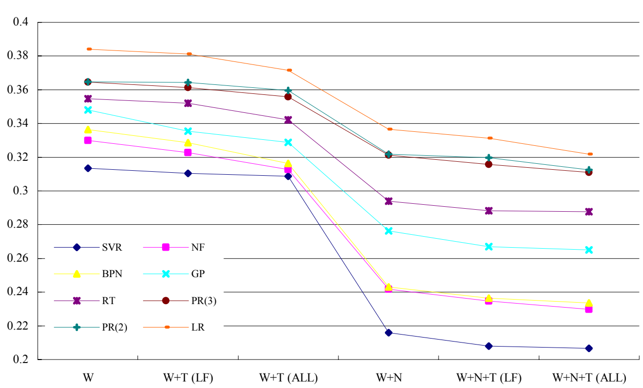

Figure 3. MAPE performance on different prediction models

performs better than the other feature combinations significantly (p-value<0.001 in t-test) for every prediction algorithm.

Table 2 further lists and analyzes the experimental results of the best prediction algorithm (i.e., SVR). Using features on the word and the topic terminology levels (W+T) improves 0.99% and 1.53% of MAPE performance with the LF and the ALL selection strategies, respectively, compared to the baseline. This is because a topic terminology is a conceptual representation of words. For example, "lawyer" and "attorney" on the topic terminology level have an identical sense, but they are different on the word level. Using features on the word and the named entity levels (W+N) improves 31.10% of MAPE performance. That demonstrates both NE and NE type are good indicators for prediction. We find that ad words containing NE "California" generally have higher CPCs than those containing "Mississippi" on the same theme. That meets our expectation because California's gross state product (GSP) is higher than Mississippi's in recent years. Similarly, we find that ad words containing NE type "city" have higher CPCs than those containing "continent". That demonstrates CPCs of more precise words are higher than those of general words. The ALL selection strategy is better than the LF strategy. It may be because the LF strategy is a special case of the ALL strategy, and the ALL strategy considers more representations of topic terminologies. Integrating features on the three levels (W+N+T) with the ALL selection strategy performs the best. The model achieving MSE 1.1674, MAE 0.5879 and MAPE 0.2067 is better than the other models significantly (p-value<0.001 in t-test).

Table 2. Impact of Features on Different Semantic Levels

Feature Level	MSE	MAE	MAPE	Improvement (%)		
W	1.6584	0.7261	0.3135	—	—	—
W+T (LF)	1.5542	0.7085	0.3104	6.28	2.42	0.99
W+T (ALL)	1.5342	0.6841	0.3087	7.49	5.78	1.53
W+N	1.5189	0.6419	0.2160	8.41	11.60	31.10
W+N+T (LF)	1.3956	0.6220	0.2080	15.85	14.34	33.65
W+N+T (ALL)	1.1674	0.5879	0.2067	29.61	19.03	34.07

Table 3 shows the performance of the best prediction model under different ranges of CPCs. Total 97.67% of CPCs are in the range from 0.01 to 10 USD. The results show our proposed model performs well within this rage.

Table 3. Prediction Performance in Various CPC Ranges

CPC	#Samples	Distribution (%)	MSE	MAE	MAPE
0.01~1	648,823	53.2233	0.0526	0.1823	0.2039
1.01~10	541,827	44.4463	0.7000	0.6146	0.1998
10.01~20	21,812	1.7892	12.536	2.9618	0.1977
20.01~30	4,546	0.3729	31.8205	4.8148	0.1937
30.01~40	1,415	0.1161	96.3122	9.0743	0.2631
≧ 40.01	637	0.0522	149.5558	10.7075	0.2276

6. CONCLUSIONS AND FUTURE WORK

In, this paper, we propose various models learned from a large-scale real-world ad words corpus for predicting CPCs of new ad words. We explore the features from different semantic levels. Considering more ad words semantics, such as named entities and topic terminologies, achieves better prediction performance than words only. Integrating features on all semantic levels and adopting all the combinations (ALL) selection strategy achieves better performance in every prediction algorithm. Using SVR prediction algorithm is better than the other six prediction algorithms significantly. In the future, considering other semantic information to predict CPC more precisely and recommending ad words along with their CPCs will be investigated.

7. ACKNOWLEDGEMENTS

This work was partially supported by National Science Council, Taiwan under grant NSC99-2221-E-002-167-MY3.

8. REFERENCES

[1] AlchemyAPI: *http://www.alchemyapi.com/*.

[2] Andreou, P.C., Charalambous, C. and Martzoukos, S.H. 2009. European option pricing by using the support vector regression approach. In *Proceedings of the 19th International Conference on Artificial Neural Networks: Part I*, 874–883.

[3] Animesh, A., Ramachandran, V. and Viswanathan, S. 2010. Quality uncertainty and the performance of online sponsored search markets: an empirical investigation. *Information Systems Research*. 21, 1, 190–201.

[4] Ashkan, A., Clarke, C.L.A., Agichtein, E. and Guo, Q. 2009. Estimating ad clickthrough rate through query intent analysis. In *Proceedings of the 2009 IEEE/WIC/ACM International Conference on Web Intelligence*, 222–229.

[5] Bajari, P. and Hortaçsu, A. 2003. The winner's curse, reserve prices, and endogenous entry: empirical insights from eBay auctions. *The RAND Journal of Economics*. 34, 2, 329–355.

[6] Cary, M., Das, A., Edelman, B., Giotis, I., Heimerl, K., Karlin, A.R., Mathieu, C. and Schwarz, M. 2007. Greedy bidding strategies for keyword auctions. In *Proceedings of the 8th ACM Conference on Electronic Commerce*, 262–271.

[7] Edelman, B. and Ostrovsky, M. 2007. Strategic bidder behavior in sponsored search auctions. *Decision Support Systems*. 43, 1, 192–198.

[8] Ghani, R. 2005. Price prediction and insurance for online auctions. In *Proceedings of the 11th ACM SIGKDD international Conference on Knowledge Discovery in Data Mining*, 411–418.

[9] Hou, L., Wang, L. and Li, K. 2007. Prediction of keyword auction using Bayesian network. In *Proceedings of the 8th International Conference on E-commerce and Web Technologies*, 169–178.

[10] Kitts, B. and Leblanc, B. 2004. Optimal Bidding on Keyword Auctions. *Electronic Markets*. 14, 3, 186–201.

[11] Kovalchuk, Y. and Fasli, M. 2008. Adaptive strategies for predicting bidding prices in supply chain management. In *Proceedings of the 10th International Conference on Electronic Commerce*, 1–10.

[12] Lin, C. S., Chou, S., Weng, S. M. and Hsieh, Y. C. 2011. A final price prediction model for English auctions: a neuro-fuzzy approach. *Quality and Quantity*. 1–15.

[13] Neshat, H.S. 2011. Ranking of new sponsored online ads. In *Proceedings of the 4th International Conference on the Applications of Digital Information and Web Technologies*, 7–12.

[14] Wellman, M.P., Reeves, D.M., Lochner, K.M. and Vorobeychik, Y. 2004. Price prediction in a trading agent competition. *Journal of Artificial Intelligence Research*. 21, 1, 19–36.

[15] WordNet: *http://wordnet.princeton.edu/*.