Mining the Long Tail of Search Queries
Finding Profitable Patterns

Michael Meisel, Maik Benndorf and Andreas Ittner
Professur Informatik Verteilte Informationssysteme, Hochschule Mittweida, Technikumplatz 17, Mittweida, Germany

Keywords: Data Mining in Electronic Commerce, Mining Text and Semi-structured Data.

Abstract: Many search engine marketing campaigns contain a lot of different search queries with a low frequency referred as "Long Tail". It is not possible to draw reliable conclusions about the performance of a specific search query with low frequency regarding a business goal because of its limited sample size. In this paper we present a method for finding profitable patterns in the long tail of search queries. The method aggregates search queries based on mined patterns and rejects the non profitable groups. We applied our method to a search engine marketing campaign with over 10,000 different search queries and performed an offline test and an online A/B-test to measure the performance of the method.

1 INTRODUCTION

Search engine marketing (SEM) is a form of advertising where companies promote their products based on customers search queries. Advertisers select a set of keywords where their adverts should be placed and a bid price they are willing to pay. The search engine determines for each search query the matching keyword definitions and places the adverts in dependence of the bids of the competitors.

The frequency distribution referred as long tail is well known ((Anderson, 2004), (Anderson, 2006)) and rests upon the Pareto Principle where 80% of the effects come from 20% of the causes. This principle applies in many different areas as in the distribution of search queries in search engine marketing campaigns, where a small number of search queries generate the main part of the traffic and a large number of search queries generate the remaining, smaller part. The competition and costs in bidding for the keywords matching the few search queries with a lot of traffic is much higher as for the keywords matching search queries with little traffic. Because of that a trend emerged to target search engine marketing campaigns on the search queries in the long tail (B. Skiera, 2010). Summing up the traffic generated by thousands of cheap keywords can be very profitable but optimizing a campaign with a large number of low frequency search queries to increase e.g. a conversion rate is very difficult. It is not possible to draw reliable conclusions about the quality of a specific search query when the sample size is too small. We developed a method to overcome this problem by aggregating single search queries in the long tail to larger sets of queries and providing the corresponding keyword definition. Thus it becomes possible to optimize a search engine marketing campaign in terms of a defined metric by identifying profitable keyword definitions with a statistically more significant size and rejecting the unprofitable keyword definitions.

2 RELATED WORK

The idea of the long tail was popularized by Chris Anderson ((Anderson, 2004), (Anderson, 2006)) to describe the demand for niche products. Anderson noted that a substantial fraction of revenue is generated from those niche products and argues that the "future of business is selling less of more" (Anderson, 2006). Brynjolfsson et al. further analyzed the anatomy and economics of long tail markets ((E. Brynjolfsson, 2007), (E. Brynjolfsson, 2011)).

In search engine marketing there has been a lot of interest from researchers on forecasting the success of single keywords by analyzing the relations between bid, rank and click-through rate ((J. Feng, 2007), (A. Ghose, 2009)). Rusmevichientong et al. postulated an adaptive bidding algorithm for identifying profitable keywords where click-through rates, costs and profits of the keywords were known in advance

Meisel M., Benndorf M. and Ittner A..
Mining the Long Tail of Search Queries - Finding Profitable Patterns.
DOI: 10.5220/0004521602250229
In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval and the International Conference on Knowledge Management and Information Sharing* (SSTM-2013), pages 225-229
ISBN: 978-989-8565-75-4

(P. Rusmevichientong, 2006), which is problematic with low frequency keywords. In addition long tail search engine marketing campaigns became very popular and attained attention in research (B. Skiera, 2010). Skiera et al. argue in their empirical study that focusing on the long tail is not that profitable because the top 20% of keywords in terms of search volume covered already 94.32% of the conversions (B. Skiera, 2010). Nonetheless our tests show that it can be valuable to find profitable pattern in the long tail.

Another research stream in web search focuses on finding concepts in search queries and relationships between search and interest ((G. Xu, 2009), (M. Pasca, 2007)). As far as we know there has been no work on aggregating search queries in the long tail from a search engine marketing perspective.

3 METHOD

Our method is a recursive algorithm (Algorithm 1). It aggregates search queries in dependence on a given target metric and provides viable keyword definitions for later deployment in Google Adwords. Typical target metrics for optimizing search engine marketing campaigns are click-through rate, conversion rate, cost-per-click and cost-per-conversion. The algorithm has four different parameters:

- training data (T)
- offset phrase (o)
- minimal size (mS)
- quality measure (mQ).

The training data for the algorithm has to contain a list of different search engine queries and the required attributes for calculating the target metric (e.g. number of impressions and number of clicks for click-through rate as target metric). The algorithm divides the whole set of available search queries into distinct subsets in dependence of the most frequent phrase in the set containing the offset phrase. The subsets are further spitted till the stop criterion is reached. The stop criterion is the minimal size of the subset. The target metric determines the attribute for calculating the minimal size (e.g. target metric conversion rate determines number of clicks as minimal size because the conversion rate depends on the number of clicks). If a subset cannot be divided any further the target metric in the subset is calculated. If the target metric in the subset fulfils the requirements of the quality measure the keyword definitions for the subset are generated.

The requirements of the target platform where the optimized search engine marketing campaign should

Algorithm 1: Keyword Definition Generation.

procedure GETKEYWORDS(T, o, mS, mQ)

Find the most frequent phrase mfp in the available search queries T containing o.

Divide T into three subsets A, B, C with:
$\forall a \in A,\ mfp = a$
$\forall b \in B,\ mfp \subset b$
$\forall c \in C,\ mfp \not\subset c$

if $(getSize(A) > mS$ **then**
  if $getQuality(A) > mQ$ **then**
    Generate a new keyword definition containing the most frequent phrase (mfp).
  end if
end if
if $getSize(B) > mS$ **then**
  $getKeywords(B, mfp, mS, mQ)$
else
  if $getQuality(B) > mQ$ **then**
    Generate a new keyword definition containing the most frequent phrase (mfp).
  end if
end if
if $getSize(C) > mS$ **then**
  $getKeywords(C, o, mS, mQ)$
else
  if $getQuality(C) > mQ$ **then**
    Generate a new keyword definition containing the most frequent offset phrase (o).
  end if
end if

end procedure

be deployed limit the design of possible keyword definitions. Therefore a single keyword definition can only consist of a phrase (single words in a specific order) with four different modifiers (Table 1). Table 2 shows an example set of possible keyword definitions describing a profitable subset of matching search queries found by the algorithm.

Table 1: Allowed modifiers for keyword definitions.

Positive exact match	Matches if a search query equals the phrase: [phrase].
Positive phrase match	Matches if a search query contains the phrase with optional words before or after the phrase: ”phrase”.
Negative exact match	Doesn’t match if the search query equals the phrase: -[phrase].
Negative phrase match	Doesn’t match if the search query contains the phrase: -”phrase”.

Table 2: Example of keyword definitions describing a profitable subset.

”flights online”
-”free flights online”
-”compare flights online”
-”flights online tracking”
-”flights online game”

4 OFFLINE EXPERIMENT

We applied the algorithm to a search engine marketing campaign with a distinct long tail distribution of the search queries (Figure 1). The campaign was carried out over the span of 365 days and contained 10,232 different search queries where the advertisement was clicked at least one time. About 80% of the search queries occurred only once during the whole time. The campaign was not homogenous over the period. During the running time the campaign owner had already made some adjustments to eliminate bad performing keywords. We took this into account when we divided the data into training and test sets.

The target metric of our analysis was the cost-per-conversion. A conversion was defined as the registration of a new user to the website after he had clicked on the advertisement. The data contained the following attributes on a daily basis: search query, number of clicks, number of conversions, total costs.

We divided the available data into training and test sets to evaluate our method. It was not possible to split the data linearly because of its non-homogeneous nature. Therefore we selected randomly 290 days for training and 75 days for testing. The algorithm generated a set of keyword definitions from the training data regarding the given values for the minimal size and the quality measure. The selected target metric for optimization determined the number of clicks as measure for the minimal size because cost-per-conversion is directly correlated to the conversion rate which depends on the number of clicks. The quality measure to score the profitability of a generated keyword definition was a maximum value for the cost-per-conversion the keyword definition obtained in the training data. Afterwards we applied the presumably profitable keyword definitions to the test data and calculated different evaluation metrics (Table 3). We run several tests as three-fold cross validation with different parameter settings to explore their influence.

Table 3: Evaluation metrics.

Cost-per-Conversion	The cost-per-conversion gained by the search queries matching the selected keyword definitions in the test set.
Relative Costs	The costs caused by the search queries matching the selected keyword definitions in the test set in relation to the total costs in the test set.
Relative Conversions	The number of conversion gained by the search queries matching the selected keyword definitions in the test set in relation to the total number of conversions in the test set.
Gain	Relative Conv.-Relative Costs

The first chart (Figure 2) shows the achieved gain on the test set over different values for the minimal size (number of clicks) with the quality measure (maximum cost-per-conversion) fixed to 4.70. The maximum gain occurred at minimal size=40. Smaller values for minimal size were indicative of over fitting the keyword definitions to the train data whereas larger numbers lead to more unspecific keyword definitions and decreasing gain.

The second chart (Figure 3) shows the dependency between maximum cost-per-conversion and relative conversions and relative costs measured on the test data with minimal size fixed to 40. The relative conversions and costs decreased with the maximum cost-per-conversion. This is because of the decreas-

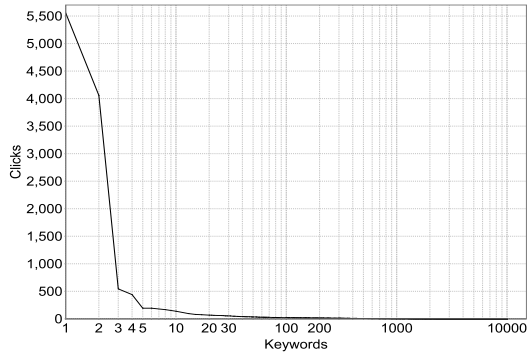

Figure 1: Distribution of search queries.

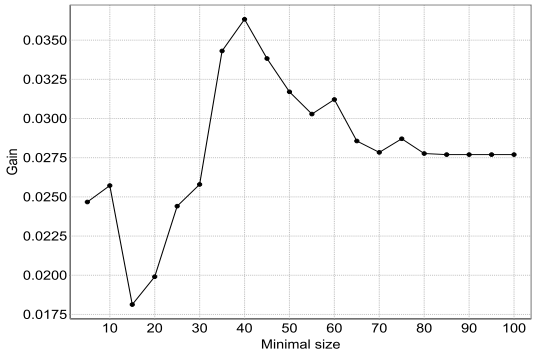

Figure 2: Gain to minimal size.

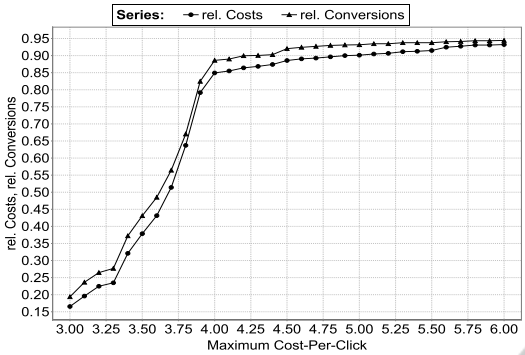

Figure 3: Maximum cost-per-click to relative conversions and relative cost.

ing number of generated keyword definitions with decreasing cost-per-conversion. The largest slope of the curve appeared around the average value of cost-per-conversion in all test sets. The gain was positive over all tested values. The largest gain occurred at a maximum cost-per-conversion of 3.60. The curves for other values of minimal size looked similar but with smaller values for gain.

Table 4 shows the target metric cost-per-conversion (CpConv) and relative number of conversions (relConv) we obtained in the test set with different parameter settings. The cost-per-conversion in all test sets without optimization averaged 3.47.

Table 4: Selected results.

minimal size	maximum CpConv	CpConv	relConv
35	3.10	2.81 (-19%)	24.0%
35	3.70	3.06 (-12%)	50.5%
35	4.10	3.31 (-4.6%)	89.4%

5 ONLINE A/B-TEST

After the offline experiments we conducted an online A/B-test to verify the results. The control group was made up of the settings from the original search engine marketing campaign we had used for training. The treatment group was made up of the keyword definitions our algorithm generated with minimal size set to 40 and maximum cost-per-conversion to 4.10 (Table 4). We elected this parameter setting because it obtained the largest gain with a high value for the relative number of conversions generated by the keyword definitions. Both groups had the same size in terms of available budget and were identical despite of the keyword definitions. From the results in the offline experiment we expected a 4.6% lower cost-per-conversion in the treatment group.

Table 5: Results Online A/B-Test.

Group	Conversions	CpConv	Variance CpConv
Control Group	255	2.67	2.10
Test Group	265	2.25	1.39

The results of the A/B-test (Table 5) indicated a lower cost-per-conversion in the treatment group as in the control group. The t-value for the Welch-test was 3.61 ($p-value = 0.00017$) which fulfils a significance level of 0.05. Thus the null hypothesis that the cost-per-conversion in both groups was the same can be rejected in favor of the alternative hypothesis that the cost-per-conversion in the treatment group were lower as in the control group. So the online experiment approved the results of our offline experiment.

From the offline experiment we expected the total number of conversion in the test group to be lower than in the control group. We cannot conclusively explain why the number of conversions in the test group was actually higher than in the control group. The relatively small number of total conversions in the online test and small expected difference in the number of conversions might be an explanation that this happened by chance.

6 CONCLUSIONS

In this paper we provided a method to aggregate low frequency search queries from the long tail into profitable keyword definitions. We could show that it is possible to indentify profitable groups of search queries in the long tail regarding an optimization goal like cost-per-conversion. Although the method is pretty simple we consider the obtained improvements on the target metric in the online experiment as relevant for practical usage. We could lower the cost-per-conversion of a SEM campaign significantly without loss of reach. On the downside the generated keyword definitions were partially unintelligibly and none self-explanatorily. This could be a problem in practice because it makes SEM campaigns difficult to control e.g. in terms of a targeted advertising of a single product in shops with multiple products.

As search engine marketing campaigns provide data about different dimensions like advertisement position, time, origin and channel we see potential for further work on this topic. The consideration of additional dimensions reduces the sample size of search queries even more and aggregating low frequency search queries is a practical method to overcome this problem.

REFERENCES

A. Ghose, S. Y. (2009). An empirical analysis of search engine advertising: Sponsored search in electronic markets. In *Management Science Volume 55 Issue 10, Pages 1605-1622*. INFORMS.

Anderson, C. (2004). The long tail. In *Wired Magazine 12(10), Pages 170-177*. Wired Magazine.

Anderson, C. (2006). *The Long Tail: Why the Future of Business is Selling Less of More.* Hyperion, New York.

B. Skiera, J. Eckert, O. H. (2010). An analysis of the importance of the long tail in search engine marketing. In *Journal Electronic Commerce Research and Applications Volume 9 Issue 6, Pages 488-494*. Elsevier Science Publishers.

E. Brynjolfsson, Y. Hu, D. S. (2011). Goodbye ff pareto principle, hello long tail: The effect of search costs on the concentration of product sales. In *Journal Management Science Volume 57 Issue 8, Pages 1373-1386*. INFORMS.

E. Brynjolfsson, Y. Hu, M. D. S. (2007). From niches to riches: Anatomy of the long tail. In *MIT Sloan Management Review Volume 47 Issue 4, Pages 67-71*. MIT.

G. Xu, S. Yang, H. L. (2009). Named entity mining from click-through data using weakly supervised latent dirichlet allocation. In *KDD '09 Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, Pages 1365-1374*. ACM.

J. Feng, H. K. Bhargava, D. M. P. (2007). Implementing sponsored search in web search engines: Computational evaluation of alternative mechanisms. In *INFORMS Journal on Computing Volume 19 Issue 1, Pages 137-148*. INFORMS.

M. Pasca, B. v. D. (2007). What you seek is what you get: extraction of class attributes from query logs. In *IJCAI'07 Proceedings of the 20th international joint conference on Artifical intelligence, Pages 2832-2837*. Morgan Kaufmann Publishers Inc.

P. Rusmevichientong, D. P. W. (2006). An adaptive algorithm for selecting profitable keywords for search-based advertising services. In *Proceeding EC '06 Proceedings of the 7th ACM conference on Electronic commerce, Pages 260-269*. ACM.