Merging Multiple Information Sources in Federated Sponsored Search Auctions

(Extended Abstract)

Sofia Ceppi
Politecnico di Milano
Milano, Italy
ceppi@elet.polimi.it

Enrico H. Gerding
Southampton University
Southampton, UK
eg@ecs.soton.ac.uk

Nicola Gatti
Politecnico di Milano
Milano, Italy
ngatti@elet.polimi.it

ABSTRACT

The recent increase of domain–specific search engines, able to discover information unknown by general–purpose search engines, leads to their federation into a single entity, called *federated search engine*. In this paper, we focus on how it can effectively merge sponsored search results, provided by the domain–specific search engines, into a unique list. In particular, we discuss the case in which the same ad can be provided by multiple sources, which requires information about the ad to be merged. We approach the problem of merging and sharing the revenue using mechanism design techniques. The main impossibility result we obtain points out there exists no mechanism that satisfies the customarily required properties. Thus, we present several mechanisms that violate at most one of these properties, and we experimentally analyze them using a real–world Yahoo! dataset.

Categories and Subject Descriptors

I.2 [**Artificial Intelligence**]: Miscellaneous

General Terms

Economics

Keywords

Mechanism Design, Sponsored Search Auctions

1. INTRODUCTION

Recently, we can see an increasing number of domain–specific search engines (DSSEs), e.g., bravofly.com, booking.com. Their advantage is that, for their specific domain, they are able to scour the *deep web* finding information (hidden in e.g. databases) that current general–purpose engines are unable to discover. This naturally leads to a new search paradigm where *federated search engines* (FSEs) integrate search results from heterogeneous DSSEs [1, 4] with the aim of providing the users with a 'one–stop shop'. However, similar to their general–purpose counterpart, DSSEs rely on revenue from sponsored search pay–per–click auctions. Currently, publishers, that use organic search results from general–purpose search engines, can use this service for their own websites for free, but in return the publishers require to also show the ads of the search engine. The revenue from clicks is then shared in a fixed way between the publisher and the search engine. However, this solution is not practical for an FSE since it then needs to display a separate list of ads for each domain–specific search engine. This is especially an issue if the same ad appears in more than one search engine (i.e. ads are shared among DSSEs). Although there is considerable literature on sponsored search auctions, to our knowledge [2] is the only other paper that considers the problem of merging sponsored search results for an FSE. However, the strong assumption is made that ads cannot be shared, which we relax in this paper.

Appears in: *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, Conitzer, Winikoff, Padgham, and van der Hoek (eds.), 4-8 June 2012, Valencia, Spain.

2. FEDERATED SEARCH ENGINE

Background The solution to the problem highlighted before, proposed in [2], is as follows. The FSE merges a selection of the ads into a coherent and unique list that it will display. To do this effectively, it needs detailed information about the ads that are known only by the DSSEs, i.e., the *qualities* (which is used to calculate an ad's click probability) of each ad, as well the *values* (i.e., the amount that the advertisers pay the search engines when their ad is clicked).

The authors of [2] approached the problem using *mechanism design* techniques. They obtained the following results. If the click probabilities are not influenced by the presence of other ads (i.e., there are no externalities), the standard VCG mechanism is *dominant–strategy* incentive compatible for this setting (i.e., truthfully elicits the values and qualities from the DSSEs). Furthermore, in the case of externalities between the ads, incentive compatibility can still be achieved using an *execution–contingent* VCG mechanism, where the payment is conditional on the realization of events (in this case the actual ads clicked by the user). However, this can only be achieved in *ex–post*, which requires others to be truthful (and thus is slightly weaker than having dominant strategies). Furthermore, although both mechanisms are weakly budget balanced (i.e. the FSE does not make a loss), this is only in expectation w.r.t. to events.

The New Challenge The work described above is based on the strong assumption that ads cannot be shared. When this assumption does not hold, as it commonly happens in practice, the nature of the problem changes fundamentally. Specifically, if an ad appears in multiple DSSEs, the FSE needs to merge all the reports received for this ad in order to accurately predict its click probability and produce efficient allocations. However, this could incentivize a DSSE

to report a lower quality for some of its low–valued ads (for which it does not expect to gain any profit) purely in order to discredit the same ads from its competitors (thereby improving the allocation for its remaining ads and/or reducing the payment). Worse still, it could fabricate ads and simply pretend to have the same ads as its competitors. Our aim is to solve this challenge using mechanism design techniques.

3. CONTRIBUTIONS AND RESULTS

Our contributions and results are as follows:

1) We extend the models of federated sponsored search auctions described in [2] to the case in which advertisers are allowed to submit a bid to more than one DSSE, and we show how this can be posed as a mechanism design problem. In particular, the main differences with the model proposed in [2] are *(i)* the introduction a *value merge function*, $mv_a(\hat{\mathbf{v}})$, which returns the highest value associated with a certain ad a given the reported value $\hat{\mathbf{v}}$ (we assume that the DSSE that provides the highest value for an ad is the only DSSE that receives money from the advertiser when the ad is clicked), and a *quality merge function*, $mq(\hat{\mathbf{q}})$, which returns the merged quality for all the ads give the reported qualities $\hat{\mathbf{q}}$ (thus, the *click probability function*, α_a, results to be a function of the allocation and the quality merge function); *(ii)* the analysis of two scenarios: one where DSSEs are able to hide their ads and fabricate advertisers if this is in their interest, and one (called *verified*) where this is not possible since an ad verification mechanism is in place.

2) We theoretically prove that in general there exists no execution–contingent mechanism that simultaneously guarantees *allocation efficiency* (AE), *incentive compatibility* (IC), *individual rationality* (IR), and *weak budget balance* (WBB), even considering *ex–post* implementation.

3) Due to the result described above, we focus on mechanisms that guarantee three of these properties. A natural candidate to consider is the unique state–of–the–art mechanism for an FSE model [2]. We observe that, in the general setting in which ads can be shared, the mechanism violates both IR and WBB.

4) We provide a range of different mechanisms that violate at most one property among WBB, IR, and AE.

Mechanisms that violate the WBB property: We propose a mechanism for the general case, *MinRep*, and a mechanism for the verified case, *VerifiedMinRep*. The idea behind them is similar to the one proposed (for a completely different domain) in [5]: in the computation of a DSSE's payment, the mechanism considers virtual qualities (that substitutes the ones actually reported by the agent) such that the social welfare is minimized, and in turn also her payment. The formula of the payment of a DSSE s in expectation w.r.t. the set of clicked ads ω, depending on the reported qualities and values $\hat{\theta}$, given the true ones, θ, is given by:

$$\mathbb{E}_\omega[p_s(\hat{\theta}|\omega)|\theta] = \min_{\mathbf{q}'_s \in Q} sw^*(\hat{\mathbf{v}}, \langle \mathbf{q}'_s, \hat{\mathbf{q}}_{-s} \rangle) - \sum_{a \in f(\hat{\theta}) \setminus A^*_s} \alpha_a(f(\hat{\mathbf{v}}, \hat{\mathbf{q}}), mq(\mathbf{q})) \cdot mv_a(\hat{\mathbf{v}}_{-s}), \quad (1)$$

where Q is the set of virtual qualities, f is the efficient allocation, and sw^* is the social welfare of the efficient allocation (where $\mathbf{x}$ is an allocation):

$$sw^*(\mathbf{v}, \mathbf{q}) = \max \sum_{a \in \mathbf{x}} \alpha_a(\mathbf{x}, mq(\mathbf{q})) \cdot mv_a(\mathbf{v})$$

Note that in Eq. (1), $mq(\mathbf{q})$ is based on the true qualities since the payment is execution–contingent.

The difference between the two mechanisms lies in the ads considered in the set of possible virtual reports, Q. In particular, in the computation of a DSSE's payment, MinRep takes into account virtual qualities for all the ads, while VerifiedMinRep considers only the one corresponding to the DSSE's actual ads.

Mechanisms that violate the IR property: We propose a mechanism for the general case, *MaxRep*, and a mechanism for the verified case, *VerifiedMaxRep*. The basic idea is the same as the one presented for the mechanisms that violate the WBB property, but instead of minimizing the payment, we now maximize it.

Mechanism that violates the AE property: We focus our investigation on strictly randomized mechanisms. We propose a mechanism for the verified case, *VerifiedRand*, that randomly selects with uniform probability which ads to display and, if the ad is provided by multiple DSSEs, it randomly selects with uniform probability which is the one that will receive the money directly from the advertiser if the ad is clicked. Payments are equal to zero. Focusing on the non verified case, we prove that there exists no strictly randomized mechanism that violates only the AE property. Indeed, DSSEs can always influence the randomization by hiding ads that give low expected utility, thus increasing the chance that their 'good' ads are selected.

5) We experimentally evaluate and compare the mechanisms presented in the paper and the state of the art in terms of the FSE's expected revenue and the DSSEs' expected utility. The experimental analysis we propose is based on the *Yahoo! Webscope A3* dataset. Results of the analysis show that, as expected, different mechanisms are appropriate for domains with different requirements. Furthermore, MaxRep and VerifiedMaxRep turn out to be impractical due to the extreme negative utility they provide to the DSSEs. Similarly, MinRep is impractical because the FSE always gets a negative revenue. In contrast, VerifiedMinRep turns out to be suitable for the FSE when the number of DSSEs that share the same ad is not too high (with 5 DSSEs if less that the 60% of them shares the same ads). This pushes us to design a verification mechanism that the FSE can use to define the set of ads that belongs to each DSSE.

4. REFERENCES

[1] D. Braga, S. Ceri, F. Daniel, and D. Martinenghi. Mashing up search services. *IEEE Internet Computing*, 12(5):16–23, 2008.

[2] S. Ceppi, N. Gatti, and E. Gerding. Mechanism design for federated sponsored search auctions. In *AAAI*, 2011.

[3] E. H. Gerding, S. Stein, K. Larson, A. Rogers, and N. R. Jennings. Scalable mechanism design for the procurement of services with uncertain durations. In *AAMAS*, pages 649–656, May 2010.

[4] D. Martinenghi and M. Tagliasacchi. Proximity rank join. *The Proc. of the VLDB Endowment*, 3(1):352–363, 2010.

[5] S. D. Ramchurn, C. Mezzetti, A. Giovannucci, J. A. Rodriguez-aguilar, R. K. Dash, and N. R. Jennings. Trust-based mechanisms for robust and efficient task allocation in the presence of execution uncertainty. *Journal of Artificial Intelligence Research*, pages 119–159, 2009.