Translating Relevance Scores to Probabilities for Contextual Advertising

Deepak Agarwal, Evgeniy Gabrilovich, Robert Hall, Vanja Josifovski, Rajiv Khanna

Contextual Advertising

"Publisher" creates page content and wishes to display ads.

"Advertiser" supplies ads.

"Ad network" decides which ads to display.

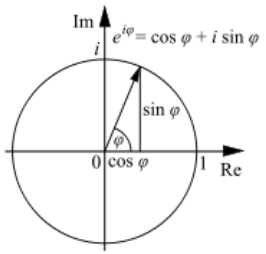
"User" decides whether to click on ads.

JUL 17

Ten Must Read Books about Mathematics

By Antonio Cangiano. Filed under Essential Math, Suggested reading

I love books with the ability to inspire readers. Many non-mathematicians consider mathematics as something abstruse and complicated, suitable only for 'nerds'. Often I highlight the unfounded nature of this prejudice, but nothing is more effective at disproving this stigma than a good book. I was in fact able to quickly change many of my friends' views on the topic, by just giving them a good book which shows the beauty and fascinating nature of mathematics and science in general. The following is a list of great titles, most of which are fairly cheap. Not all of them are suitable



Click Probabilities are Important

The goal of the ad network is to maximize their revenue:

Expected revenue for page.

Price advertiser pays to have ad in position *i*.

Probability that ad is clicked on, given the page content.

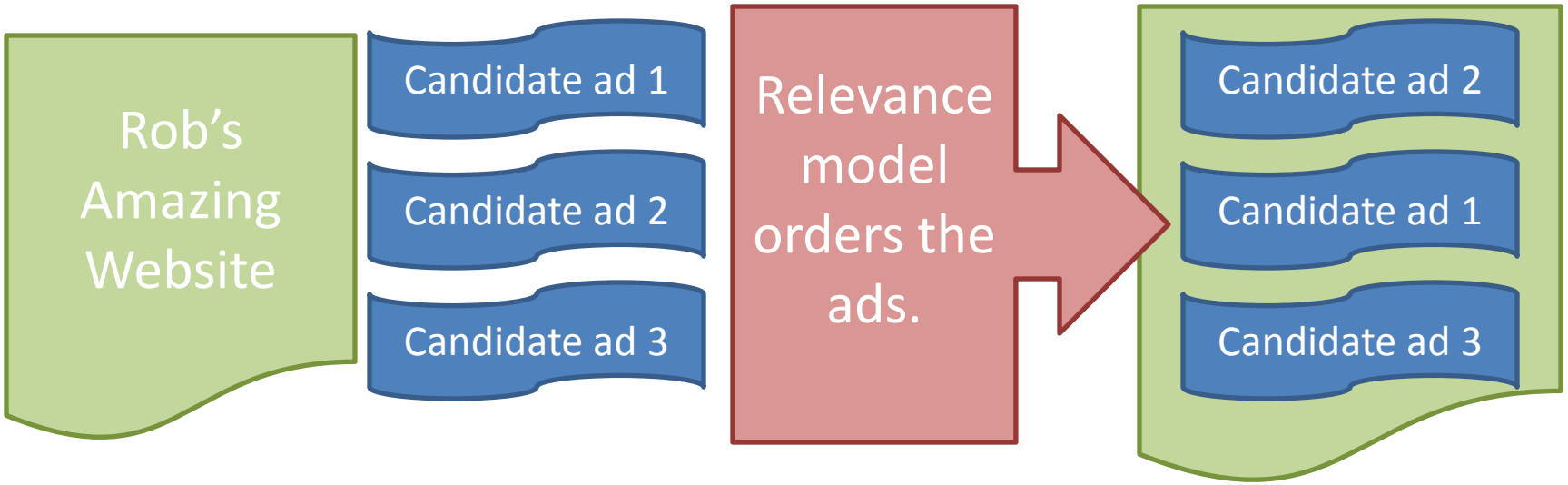
$$ER = \sum_{i=1}^{k} \text{price}(a_i, i) P(\text{click}|\text{page}, a_i)$$

Ad network chooses the ads which maximize this.

A good model of click probability is essential to successfully optimize revenue .

Relevance Models of Ads

"Relevance" (e.g., a vector space model) may be used to select ads for display, and to order them:



To maximize revenue we need good absolute scores as well as a good ordering.

Logistic Regression of Clicks Based on Relevance

$$P(\text{click}|\text{page}, a_i) = p_i$$

$$\text{logit}(p_i) = \phi(\text{page}, a_i)^T \beta + f(r_i)$$

"Features" of the page and ad.

Model parameters

Function of the relevance

Idea: divide space of relevance score into bins.

$$f(r) = \sum_{k=1}^{M} \alpha_k 1(r \in B_k)$$

Model parameters

Indicator of i^{th} bin.

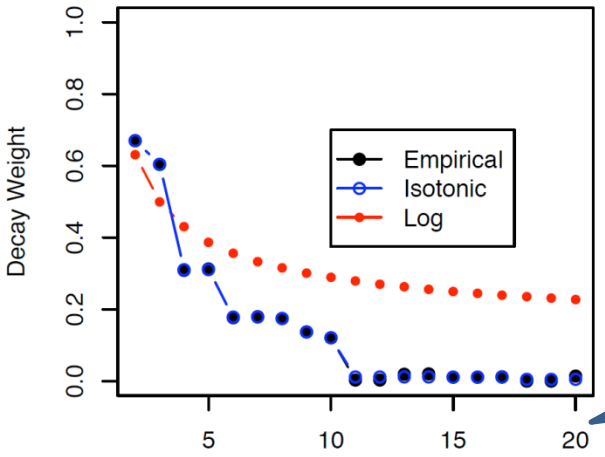
Page and Ad Features

- **Taxonomy**: Label of page and ad in a large hand-made taxonomy of about 6000 classes.
- **Domain**: Domain on which ads are shown (since pages on the same domain have similar structure and ad visibility).
- **Words in common:** words common to both the ad and the page.
- **Ad Position**: Position in the list of ads (since top ranked ads are more often clicked).

Experimental Evaluation

- Task is to rank ads.
- Data are 3M slates of ads taken over 22 days.
- 2M used for estimating parameters, 1M for evaluating model performance.
- Metric is Discounted Cumulative Gain (DCG)

Score you get if ad was clicked.



Position you put ad.

Evaluated Models

- **VSM**: The raw relevance scores.
- **Global**: A single logistic regression fit to all the data.
- **PART**: A logistic regression fit for each of the top 20 publishers (70% of data) and a single logistic regression to the remaining data.
- **EM**: A mixture of logistic regressions to the data.

Results

- PART is the superior model for two reasonable choices of weights:

Model	VSM	PART	Global	GlobalW	EM	EMW
Log-NDCG	.692	.781	.773	.773	.771	.773
Emp-NDCG	.549	.561	.556	.555	.555	.555

- Improvement over VSM is significant: