



# PREDICTING CLICK THROUGH RATE FOR NEW ADS WITH SEMANTICALLY SIMILARITY MEASUREMENT

1

Hamed Sadeghi Neshat

# CONTENTS

- Introduction
- Related works
- Proposed CTR Prediction Method
- Evaluation
  - Data Collection
  - Methodology
  - Result and Comparison
- Conclusion & Future works

# INTRODUCTION

- Internet advertising
- Add Quality
  - Rank factor = Quality factor \* offered price
- PPC model

October-10



credit card

Search

About 333,000,000 results (0.18 seconds)

[Advanced search](#)

Everything

News

Images

More

The web

Pages from Canada

Any time

Latest

Past 2 days

All results

Wonder wheel

Sites with images

Page previews

More search tools

[Capital One® Credit Card](#)

[www.capitalone.ca](http://www.capitalone.ca) Canadian No Hassle Card Selection. Official Site – Apply Online!

[Apply For A Credit Card](#)

[Canada.CreditCards.com](http://Canada.CreditCards.com) Start Comparing Credit Cards Here. Apply Today for a New Credit Card.

[CitiFinancial® Loans](#)

[Loans.CitiFinancial.ca](http://Loans.CitiFinancial.ca) Fast and Efficient Loan Process Is Available To Our Applicants Today!

[Credit card - Wikipedia, the free encyclopedia](#) ☆

A **credit card** is a small plastic card issued to users as a system of payment. It allows its holder to buy goods and services based on the holder's promise ...

[How credit cards work - Features - Security problems and solutions](#)

[en.wikipedia.org/wiki/Credit\\_card](http://en.wikipedia.org/wiki/Credit_card) - Cached - Similar

[Canadian Credit Card - Credit Card Applications Canada](#) ☆

The definitive online resource for Canadian **Credit Cards**. Search and compare the various cards available to you from all major Canadian banking ...

[Compare. - Credit Card FAQ - Glossary](#)

[www.creditcards.ca/](http://www.creditcards.ca/) - Cached - Similar

Sponsored links

Sponsored links

[RBC® Credit Card Offers](#)

Choose from Low Rate, No Annual Fee and RBC Rewards® Credit Cards.

[RBC.com](http://RBC.com)

[Credit card](#)

Get 2% cash back. It's a smarter way to pay. Apply online today.

[www.Scotiabank.com](http://www.Scotiabank.com)

[Apply for PC® MasterCard®](#)

No Annual Fee. Free Groceries with PC points. Apply Today & Save!

[www.PCFinancial.ca/PC-Mastercard](http://www.PCFinancial.ca/PC-Mastercard)

[Canadian Tire MasterCard](#)

Earn Rewards Everywhere You Shop. Fill out an Application Now!

[ctfs.com/OptionsMasterCard](http://ctfs.com/OptionsMasterCard)

## RELATED WORKS

- Query Content Model, University of Waterloo, 2009
- MLE Model, Poland University & Microsoft research, 2008
- LR Model, Microsoft research, 2007

# PROPOSED CTR PREDICTION METHOD

- High level Idea:
  - We are working with similar ads to predict CTR
  - If we have two similar ads (based on content's meanings and general features), it is reasonable to expect same CTR for both of them
  - We are working with contents' semantic as well their general features

# PROPOSED CTR PREDICTION METHOD

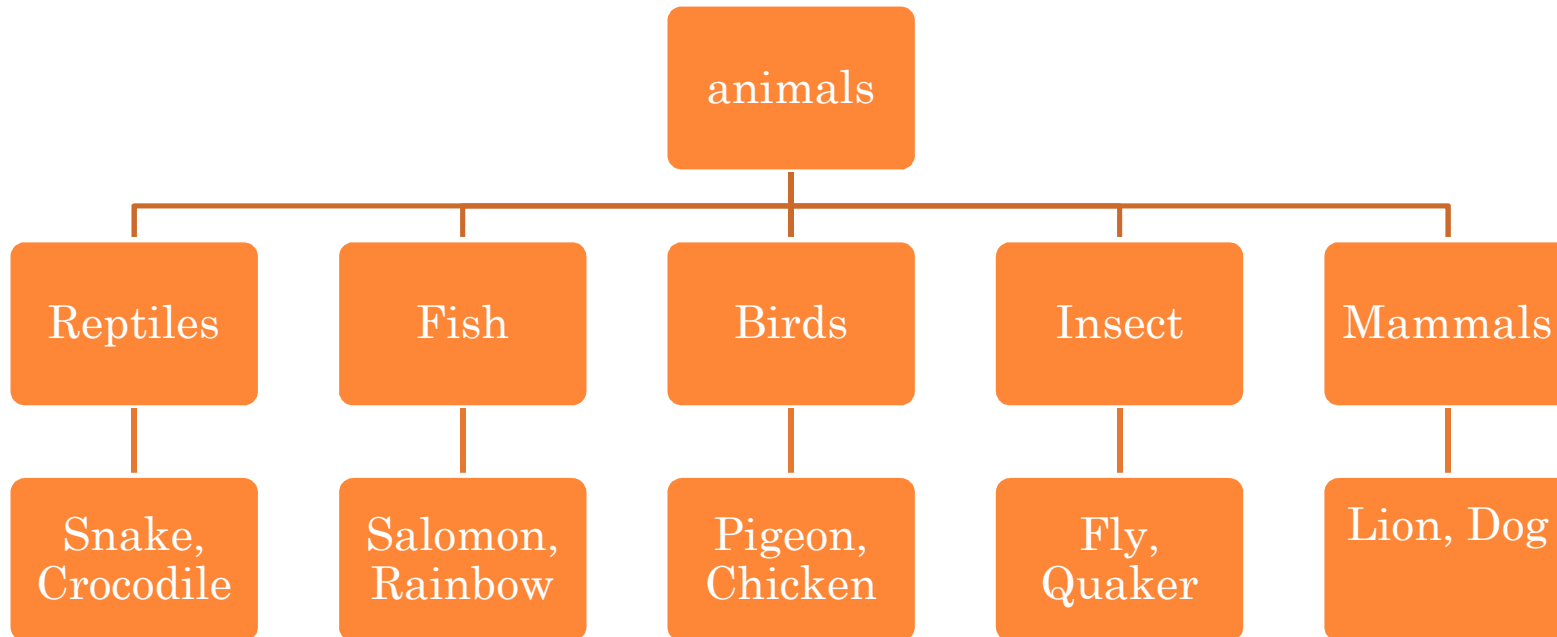
- High level Idea:
  - CTR distribution among different contexts may be depends on different factors. For example in some contexts users may look more at brands, but in another context they should look at specific words in the title or ads' description, so suggesting one general model and feature set probably will not work good for all ads in different context.

# PROPOSED CTR PREDICTION METHOD

1. We used our own dataset of a huge collection of ads and extracted all used terms.
2. Then, we clustered all terms which were used in ads in our dataset, for this reason, we formed 3 sets of clusters:
  1. based on term's features: we used K-Means clustering to form 100 clusters.
  2. based on number of repetition: contains 50 clusters coming from K-Means algorithm based on number of repetition for each term in the all retrieved ads in dataset.
  3. based on terms' concepts (meaning): We categorized all terms in 1150 clusters in hierarchical pattern, from 17 main categories.

# PROPOSED CTR PREDICTION METHOD

October-10



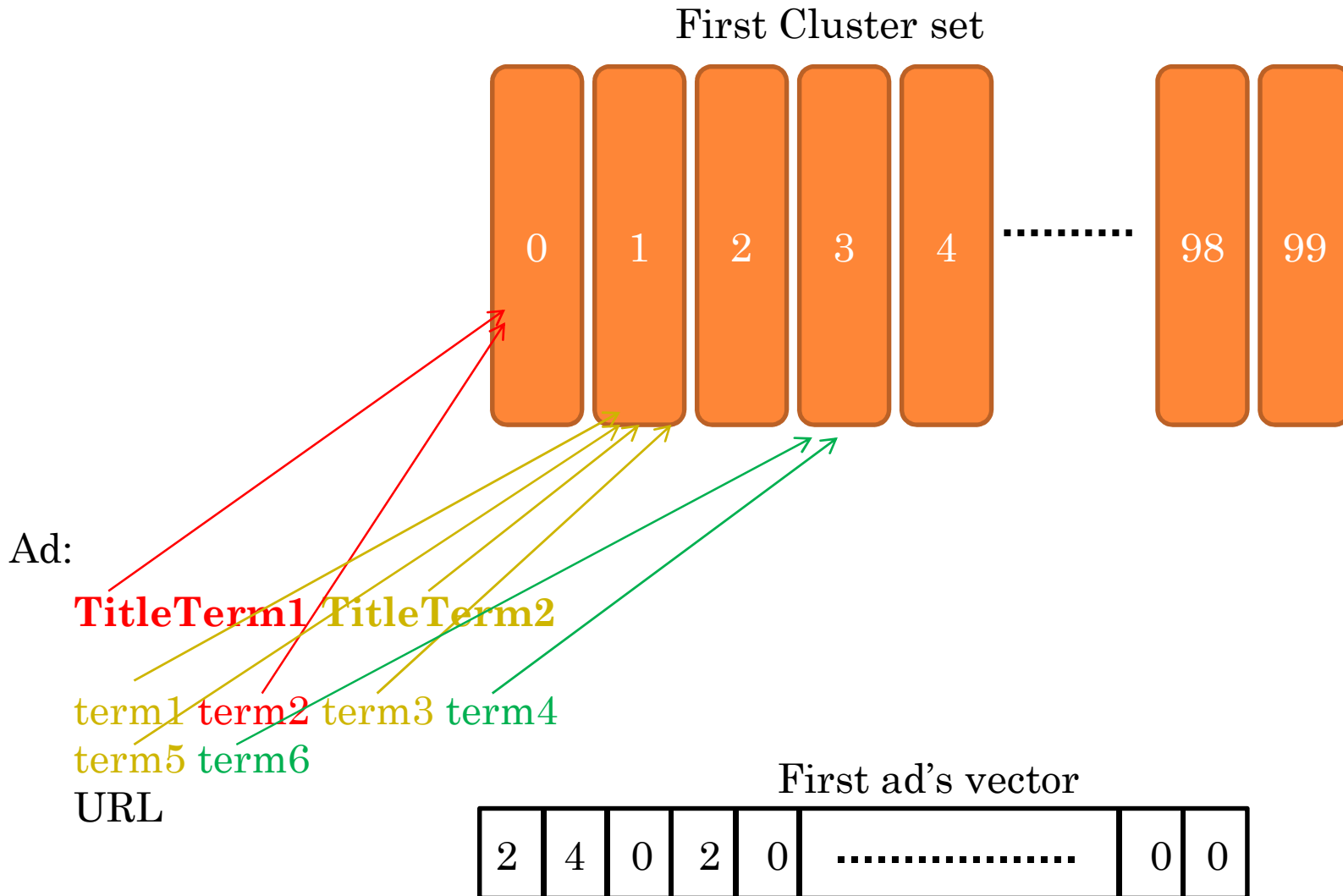


# PROPOSED CTR PREDICTION METHOD

3. For each ad, we formed 3 vectors which each of these vectors was related to one of the cluster's sets, if the ad had a term from n-th cluster , we increase its related vector's entry by 1.

# PROPOSED CTR PREDICTION METHOD

October-10



# PROPOSED CTR PREDICTION METHOD

4. If during the search and selecting ad process, we found a candidate ad which was new and had the same keyword as user entered query ;for predicting its CTR, we will look at all other existing ads with that query in their keywords

## PROPOSED CTR PREDICTION METHOD

5. after collecting enough ads for comparison, we could rank them based on their similarity with new ad by looking at the distance between vectors

$$\text{X2: } D_c(H_i, H_j) = \frac{1}{2} * \sum_{n=1}^V \frac{(h_{in} - h_{jn})^2}{h_{in} + h_{jn}}$$

$$\text{SED: } d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

6. We used rank aggregation method to combine different ranked list together and achieve to one list contains ads which are ordered by their distance from the new ad.

## PROPOSED CTR PREDICTION METHOD

7. By looking at first top ten ads (ten most similar ads) and using their known quality(CTR), we could predict new ads' quality.

# EVALUATION

- Data Collection:
  - We found 800,000 common ads' keywords
  - More than 4,000,000 ads retrieved
  - We extracted all terms from ads and specifying these features:
    - Global Monthly Searches
    - Estimated Daily Impressions
    - Estimated Ad Position
    - Estimated CTR
    - Estimated Daily Clicks
    - Estimated Daily Cost
    - Estimated Avg. CPC
- These works take 30 days non stop process over 10 PCs

# EVALUATION

- Methodology:
  - Removing 100,000 ads
  - Defining a number as their CTR:

$$n_r * \frac{1}{rank+8*p_n}$$

$N_r$ : the number of results for searched query,

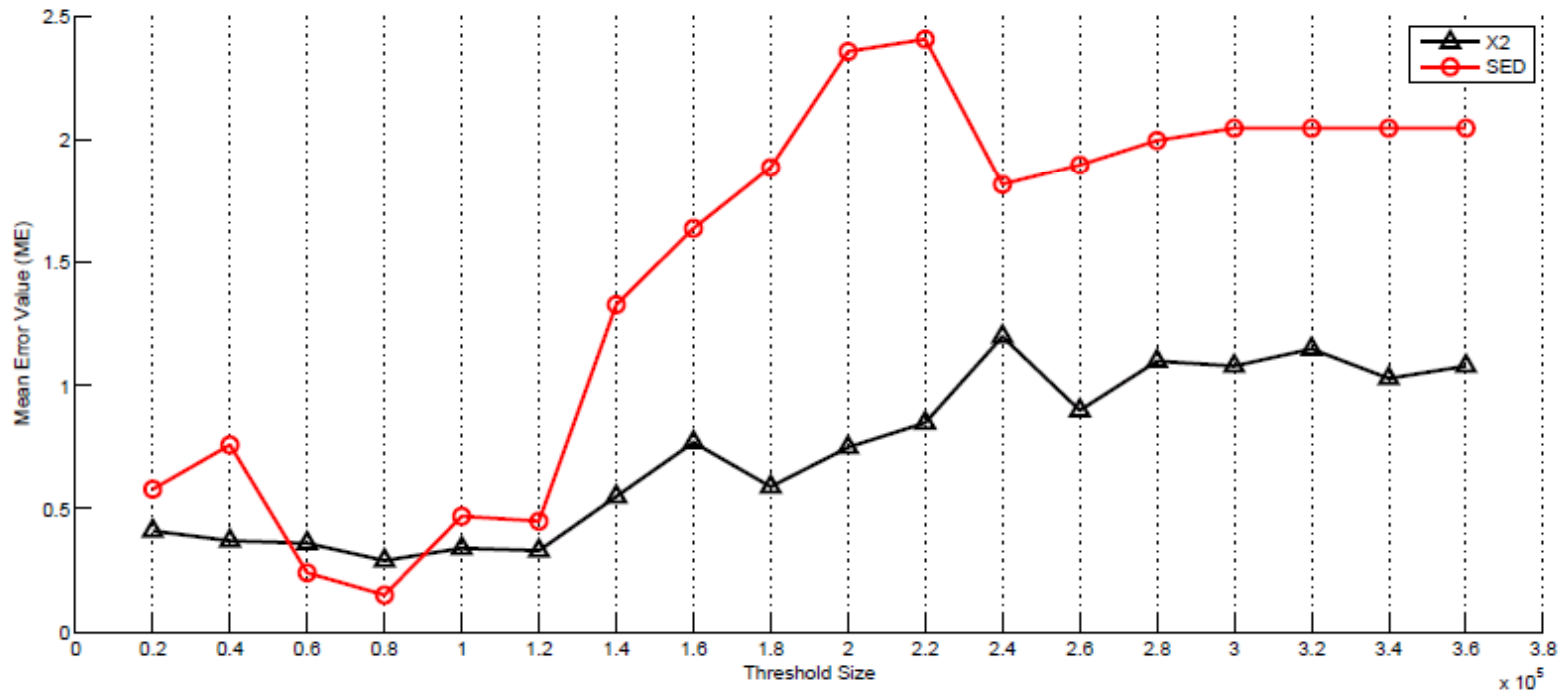
Rank: ad's rank within the search result page

$P_n$ : Page number which ads appeared on

- Accuracy measurement:
  - MSE
  - KL-Divergence

# EVALUATION

## ○ Results:





# EVALUATION

- Results:

|                        | MSE  | KLD  | ME   |
|------------------------|------|------|------|
| SED Similarity         | 2.53 | 2.13 | 0.15 |
| X2 Similarity          | 3.08 | 3.02 | 0.29 |
| Query Intent Model [3] | 3.09 | 3.56 | 0.51 |
| LR Model[4]            | 3.75 | 2.86 | -    |
| MLE Model[5]           | 2.47 | -    | -    |

## CONCLUSION

- For the first time, we used conceptual similarity measurement in real time algorithm for predicting new ads' CTR
- After applying our proposed algorithm, we could estimate new ads' quality and their rank among other ads with 0.15 error value (ME metric) in the reasonable time.

# FUTURE WORKS

- Extending this idea to multimedia contents, like YouTube for more relevant ad and related video representing

