



# Comparing Click Logs and Editorial Labels for Training Query Rewriting

Wei Vivian Zhang and Rosie Jones

Yahoo!

May 8, 2007



# Introduction of Query Rewriting (Jones et al WWW 2006)

Cat canser → cat cancer

Spell correction



Cat cancer → feline cancer

Paraphrasing



Banff flights → Calgary flights

Related/substitutable concept



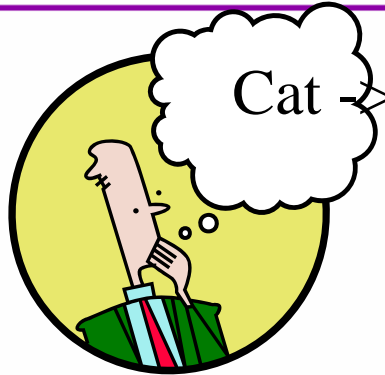
Banff flights → Costa Rica flights

Unrelated concept





# Learning Query Rewrite Relevance Models from Clicks



Cat -> feline?

Editorial evaluation

- Time for reflection
- Label 1000-5000 examples/week
- Synonyms score better?



Cat -> cats?

Users click quickly

- Does this look like what I searched for?
- Result quality affects clicks
- Click on millions of ads/day
- Visually small changes score better?



# Training Data for Machine Learning

- Editorial labeling
  - Small quantities
  - High quality
- Click
  - Large quantities
  - Noisy
  - **Our true target function!**



# Objectives

- Compare editorial labels with user clicks
- Use clicks as training labels
- Evaluate the correlation between clicks and rewriting features using query log analysis
- Compare coefficients when using click data versus human judgments

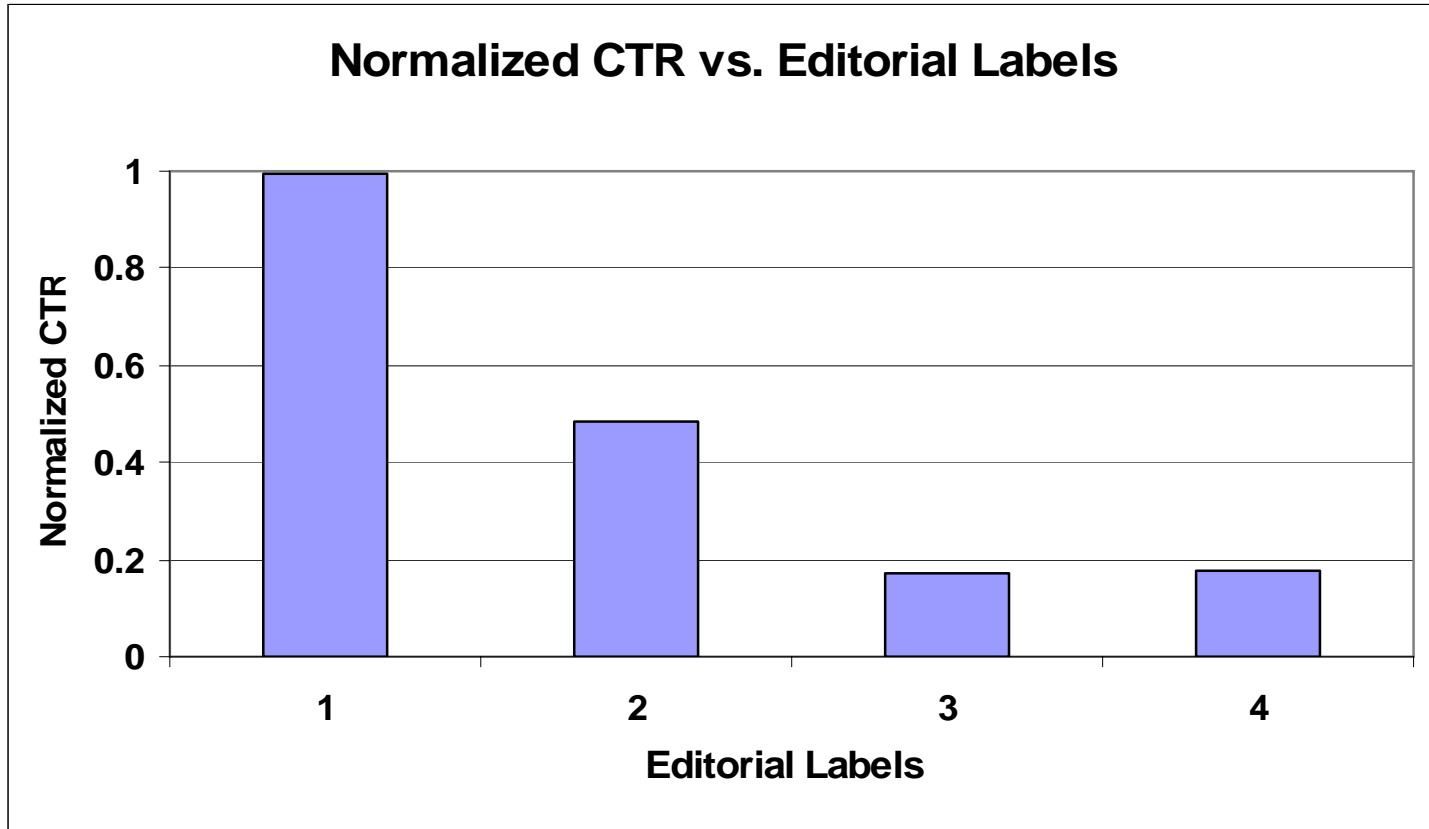


# Editorial Labels

<b>1- Precise Match</b>	A near-certain match. <i>E.g. : automotive insurance - automobile insurance</i>
<b>2- Approximate Match</b>	A probable, but inexact match with user intent. <i>E.g.: apple music player - ipod shuffle</i>
<b>3- Marginal Match</b>	A distant, but plausible match to a related topic. <i>E.g.: glasses - contact lenses</i>
<b>4- Mismatch</b>	A clear mismatch.



# Editorial Labels and Clicks





# Even Bad Rewrites Receive Some Clicks

Original Query	Rewritten Query
wcw	ecw
superpages	white pages
centennial wireless	cingular wireless
helicopter game	game
usda	fda
superpages	yellow pages
airjamaica.com	aa.com
craigslist.com	monster.com
white pages	yellow pages

Marginal matches

Original Query	Rewritten Query
nitric acid	nitric oxide
contender	contender boats
pool	pool tables
girl	cheetah girl
trec	trek
monkeys	sea monkeys
u	us
fight	flights
white pages	zip codes

Mismatches



# Using Clicks as Training Labels

- Binary clicks

$$\log \frac{p}{(1-p)} = c + \sum_i w_i f_i$$

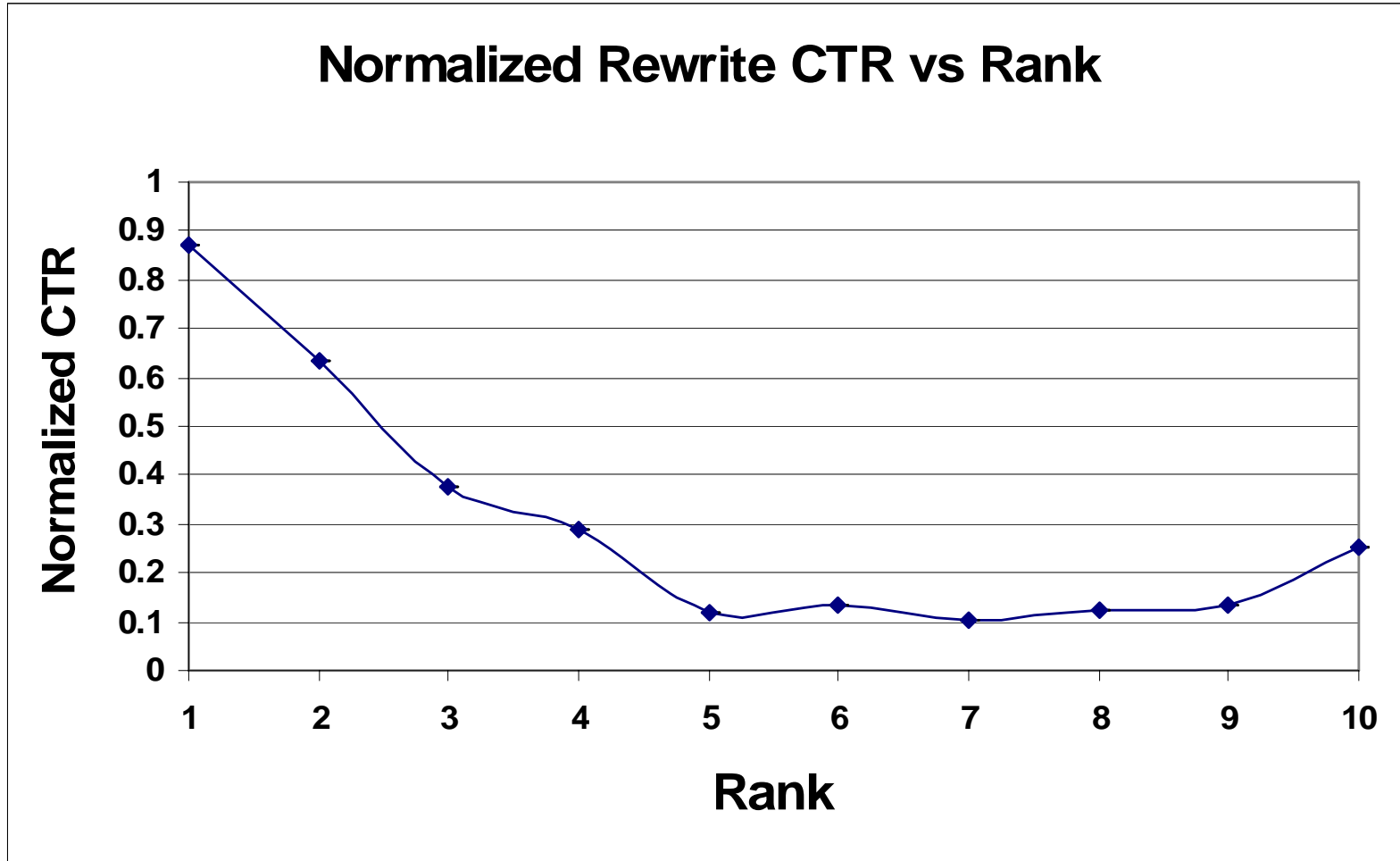
- Clicks Over Expected Clicks (COEC)

$$\frac{\sum_r c(r, m)}{\sum_r ec(r, m)}$$

- Clicks subtract expected clicks (Agichtein et al, 2006)

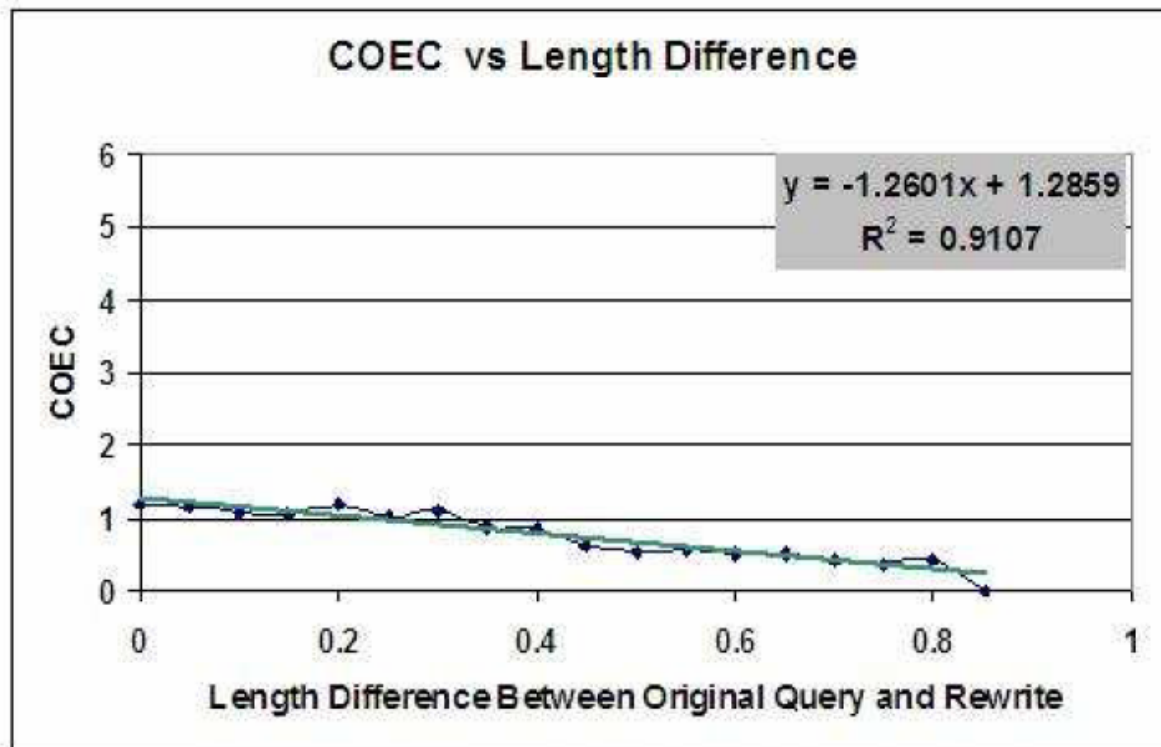


# Rank Effects for Clicks



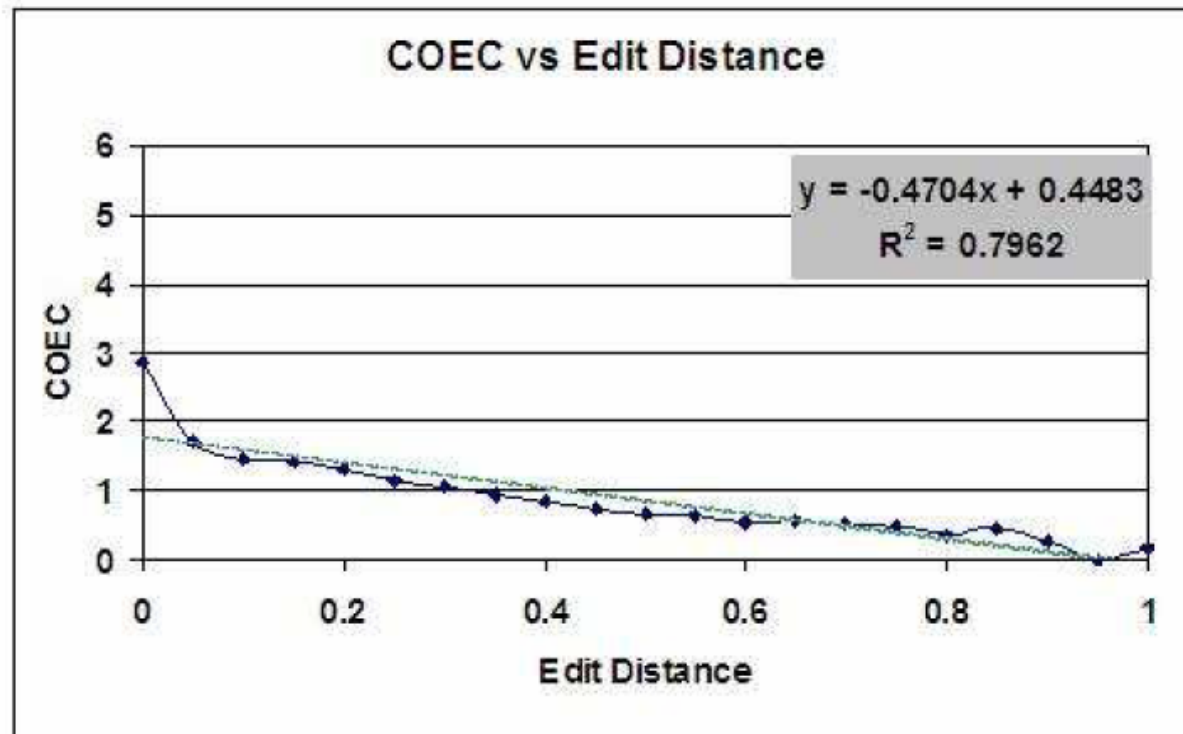


# Correlations Between Clicks and Features using Query Log Analysis



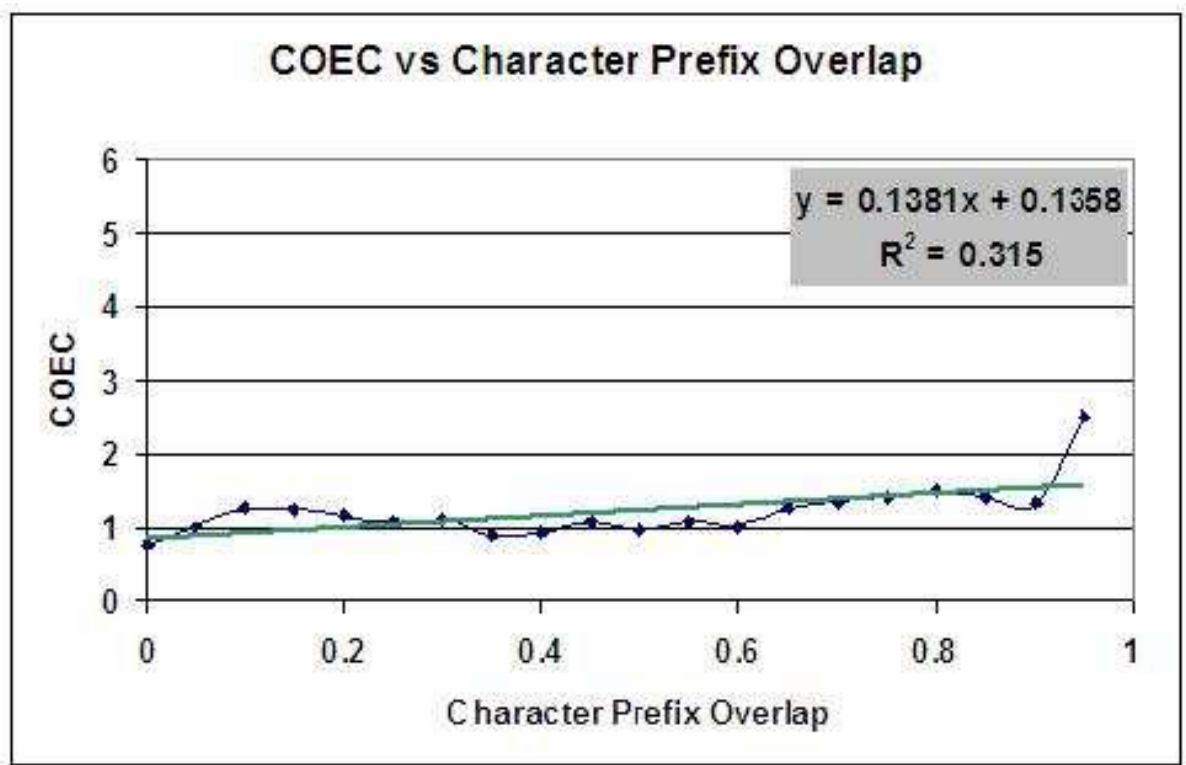


# Correlations Between Clicks and Features using Query Log Analysis (Cont.)





# Correlations Between Clicks and Features using Query Log Analysis (Cont.)





# Model Trained using Binary Clicks

- Query rewritten relevance ranking model trained using editorial labels (Jones et al, WWW '06)

$$f(q1, q2) = 0.74 + 1.88 \text{ editDist}(q1, q2) \\ + 0.71 \text{ wordDist}(q1, q2) \\ + 0.36 \text{ numSubst}(q1, q2)$$

- The model trained with the same set of features using binary clicks

$$\log \frac{p}{(1-p)} = -3.94619 - 1.08 \text{ editDist}(q1, q2) \\ - 0.57 \text{ wordDist}(q1, q2) \\ - 0.08 \text{ numSubst}(q1, q2)$$



# Conclusions

- Generally more relevant rewrites get more clicks.
- Optimizing the relevance for query rewriting may not necessarily lead to optimal clicks.
- Using query log analysis to show the correlation between clicks and the query rewriting features.
- Using clicks to train query rewrites is promising!