

Search Log Analysis

Jaime Arguello

INLS 509: Information Retrieval

jarguell@email.unc.edu

April 10, 2013

Search Log Analysis

- Why is search log analysis important?
- What does a search log look like?
- Using search logs to better understand short- and long-term search tasks
- Using search logs to infer document relevance and ranking mistakes

Methods for IR Evaluation

- Test-collection (batch) evaluation
- User studies
- Search log analysis

Test Collection-based Evaluation

advantages

- The experimental set-up is fixed: same queries, same corpus, same judgements
- Evaluations are reproducible: keeps us honest and allows us to easily measure improvement
- Modifying the system and re-evaluating is easy and free!
- Makes error-analysis possible

Test Collection-based Evaluation

disadvantages

- Test-collection-building is time and resource intensive
- Human assessors are not users
- Makes assumptions that do not hold true in “real” life:
 - ▶ relevance is topical
 - ▶ context independent
 - ▶ user independent
 - ▶ stable over time

User Study Evaluation

advantages

- Can collect lots of data about users' reactions to a system
- The experimenter can manipulate or control the search task and the searcher's internal/external context
- Can be used to study unique populations of users

User-Study Evaluation

disadvantages

- Time and resource intensive
- The exact experimental setting cannot be replicated; not a particularly good way to tune parameters
- The laboratory setting is not the user's normal environment
- Study participants know they are being 'observed'

Search-Log Analysis

general idea

- Once a system is deployed, can we reason about how it's performing by analyzing the search log?
- Can we use search-log information to improve its performance?
- Can we use search-log information to provide new services that enhance the user experience?

What is a Search-Log?

- Most search engines save information about every search
 - ▶ the query
 - ▶ a time-stamp
 - ▶ the IP address of the search client
 - ▶ the user id (stored in a cookie)
 - ▶ information about the search client (OS, browser, etc.)
 - ▶ the results that are presented
 - ▶ the results that are clicked
 - ▶ dwell time on a clicked result
 - ▶

What is a Search-Log?

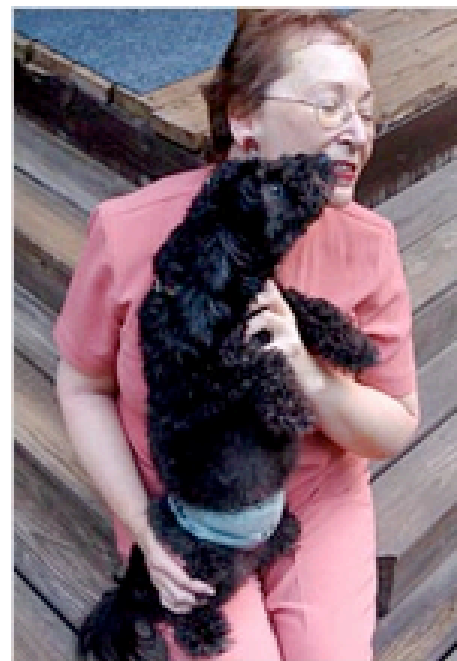
- This information is very sensitive and very valuable
- There are few publicly available Web search query-logs
 - ▶ the Excite Log (1997): ~18K users, ~50K queries
 - ▶ the AOL Log (2006): 650K users, ~20M queries
- Why aren't more search logs publicly available?
 - ▶ competitive reasons
 - ▶ privacy reasons

What is a Search-Log?

A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and TOM ZELLER Jr.
Published: August 9, 2006

Buried in a list of 20 million Web search queries collected by AOL and recently released on the Internet is user No. 4417749. The number was assigned by the company to protect the searcher's anonymity, but it was not much of a shield.



Erik S. Lesser for The New York Times

Thelma Arnold's identity was betrayed by AOL records of her Web searches, like ones for her dog, Dudley, who clearly has a problem.

Multimedia

[Graphic: What Revealing Search Data Reveals](#)

benefit academic researchers.

No. 4417749 conducted hundreds of searches over a three-month period on topics ranging from "numb fingers" to "60 single men" to "dog that urinates on everything."

And search by search, click by click, the identity of AOL user No. 4417749 became easier to discern. There are queries for "landscapers in Lilburn, Ga," several people with the last name Arnold and "homes sold in shadow lake subdivision gwinnett county georgia."

It did not take much investigating to follow that data trail to Thelma Arnold, a 62-year-old widow who lives in Lilburn, Ga., frequently researches her friends' medical ailments and loves her three dogs. "Those are my searches," she said, after a reporter read part of the list to her.

AOL removed the search data from its site over the weekend and apologized for its release, saying it was an unauthorized move by a team that had hoped it would

 SIGN IN TO
E-MAIL THIS

 PRINT

 SINGLE PAGE

 REPRINTS

Descendants
COMING SOON

Search-Logs and Privacy

- It's surprisingly easy to identify a person based on their queries
- Users prefer to remain anonymous
- We issue lots of “interesting” queries:
 - ▶ “how to tell a fake rolaX”
 - ▶ “pictures of stars in the solar system”
 - ▶ “biggest star in the universe”
 - ▶ “effective ways to fish a lizard”
 - ▶ “why does my iguana bob its head”

What does a Search-Log Look Like?

1024071	taraji henson	2006-03-02 00:28:45	4	http://www.tv.com
1024071	taraji henson	2006-03-02 00:28:45	1	http://www.imdb.com
1024071	the flavor of love vh1	2006-03-02 00:31:01	1	http://www.vh1.com
1024071	the flavor of love hoops	2006-03-02 00:38:32	1	http://www.vh1realityworld.com
1024071	beyonce	2006-03-02 22:42:05	1	http://www.beyonceonline.com
1024071	beyonce	2006-03-02 22:42:05	6	http://www.imdb.com
1024071	afc fighting	2006-03-04 22:35:33	2	http://sfuk.tripod.com
1024071	din thomas march 4th	2006-03-05 23:38:54	1	http://www.mmaringreport.com
1024071	mfc march 4th results	2006-03-05 23:45:49	3	http://www.mmaringreport.com
1024071	mfc march 4th results	2006-03-05 23:45:49	9	http://man-magazine.com
1024071	unc basketball roster	2006-03-09 23:45:15	2	http://tarheelblue.collegesports.com
1024071	unc basketball roster	2006-03-09 23:45:15	2	http://tarheelblue.collegesports.com
1024071	nit free picks	2006-03-15 14:02:21	1	http://www.docsports.com
1024071	1490 am radio	2006-03-15 14:48:01	8	http://www.1490wwpr.com
1024071	1490 am radio fl	2006-03-15 14:50:08	2	http://www.ontheradio.net
1024071	benihanas	2006-03-16 17:27:25	1	http://www.benihana.com
1024071	2006 winter music fest miami fl	2006-03-22 00:35:20	1	http://www.wintermusicconference.com
1024071	hotmail	2006-04-01 18:49:02	1	http://www.hotmail.com
1024071	my space	2006-04-02 01:21:41	1	http://www.myspace.com
1024071	my space	2006-04-02 15:59:20	1	http://www.myspace.com
1024071	my space	2006-04-02 22:03:10	1	http://www.myspace.com
1024071	nba jams super nintendo cheats	2006-04-03 21:06:11	2	http://www.elook.org
1024071	my space	2006-04-03 21:16:00	1	http://www.myspace.com
1024071	charlie's dodge fort pierce	2006-05-08 20:06:17	1	http://www.dealernet.com
1024071	charlie's dodge of fort pierce used cars	2006-05-08 20:09:27	2	http://www.automotive.com
1024071	justin timberlake new album	2006-05-12 16:21:36	4	http://www.mtv.com
1024071	mike epps	2006-05-13 19:45:56	6	http://www.hollywood.com
1024071	mike epps bio	2006-05-13 19:51:05	4	http://movies.aol.com
1024071	mike epps bio	2006-05-13 19:51:05	9	http://www.moono.com
1024071	mike epps bio	2006-05-13 19:55:56	14	http://video.barnesandnoble.com
1024071	mike epps bio	2006-05-13 19:55:56	21	http://www.hbo.com
1024071	mike epps bio	2006-05-13 20:01:06	24	http://www.vh1.com
1024071	mind freak	2006-05-14 00:46:18	10	http://video.google.com
1024071	criss angel mind freak	2006-05-14 12:53:35	1	http://www.crissangel.com
1024071	criss angel mind freak	2006-05-14 12:53:35	8	http://www.imdb.com
1024071	06-06-06	2006-05-14 22:29:11	1	http://www.timesonline.co.uk
1024071	show and sell auto fort pierce fl	2006-05-15 16:58:53	1	http://www.traderonline.com
1024071	barry bonds homerun ball 714 for sale	2006-05-25 16:25:41	5	http://www.sportsnet.ca
1024071	ufc 60 live results	2006-05-27 23:00:38	4	http://www.prowrestling.com
1024071	ufc 60 live play by play	2006-05-27 23:07:16	4	http://www.24wrestling.com
1024071	how to tell a fake rolex	2006-05-29 14:53:53	1	http://www.aplusmodel.com
1024071	how to tell a fake rolex	2006-05-29 14:53:53	8	http://www.inc.com
1024071	locating serial number on rolex	2006-05-30 21:51:34	1	http://www.qualitytyme.net

(AOL query-log)

What does a Search-Log Look Like?

1024071	taraji henson	2006-03-02 00:28:45	4	http://www.tv.com
1024071	taraji henson	2006-03-02 00:28:45	1	http://www.imdb.com
1024071	the flavor of love vh1	2006-03-02 00:31:01	1	http://www.vh1.com
1024071	the flavor of love hoops	2006-03-02 00:38:32	1	http://www.vh1realityworld.com
1024071	beyonce	2006-03-02 22:42:05	1	http://www.beyonceonline.com
1024071	beyonce	2006-03-02 22:42:05	6	http://www.imdb.com
1024071	afc fighting	2006-03-04 22:35:33	2	http://sfuk.tripod.com
1024071	din thomas march 4th	2006-03-05 23:38:54	1	http://www.mmaringreport.com
1024071	mfc march 4th results	2006-03-05 23:45:49	3	http://www.mmaringreport.com
1024071	mfc march 4th results	2006-03-05 23:45:49	9	http://man-magazine.com
1024071	unc basketball roster	2006-03-09 23:45:15	2	http://tarheelblue.collegesports.com
1024071	unc basketball roster	2006-03-09 23:45:15	2	http://tarheelblue.collegesports.com
1024071	nit free picks	2006-03-15 14:02:21	1	http://www.docsports.com
1024071	1490 am radio	2006-03-15 14:48:01	8	http://www.1490wwpr.com
1024071	1490 am radio fl	2006-03-15 14:50:08	2	http://www.ontheradio.net
1024071	benihanas	2006-03-16 17:27:25	1	http://www.benihana.com
1024071	2006 winter music fest miami fl	2006-03-22 00:35:20	1	http://www.wintermusicconference.com
1024071	hotmail	2006-04-01 18:49:02	1	http://www.hotmail.com
1024071	my space	2006-04-02 01:21:41	1	http://www.myspace.com
1024071	my space	2006-04-02 15:59:20	1	http://www.myspace.com
1024071	my space	2006-04-02 22:03:10	1	http://www.myspace.com
1024071	nba jams super nintendo cheats	2006-04-03 21:06:11	2	http://www.elook.org
1024071	my space	2006-04-03 21:16:00	1	http://www.myspace.com
1024071	charlie's dodge fort pierce	2006-05-08 20:06:17	1	http://www.dealernet.com
1024071	charlie's dodge of fort pierce used cars	2006-05-08 20:09:27	2	http://www.automotive.com
1024071	justin timberlake new album	2006-05-12 16:21:36	4	http://www.mtv.com
1024071	mike epps	2006-05-13 19:45:56	6	http://www.hollywood.com
1024071	mike epps bio	2006-05-13 19:51:05	4	http://movies.aol.com
1024071	mike epps bio	2006-05-13 19:51:05	9	http://www.moono.com
1024071	mike epps bio	2006-05-13 19:55:56	14	http://video.barnesandnoble.com
1024071	mike epps bio	2006-05-13 19:55:56	21	http://www.hbo.com
1024071	mike epps bio	2006-05-13 20:01:06	24	http://www.vh1.com
1024071	mind freak	2006-05-14 00:46:18	10	http://video.google.com
1024071	criss angel mind freak	2006-05-14 12:53:35	1	http://www.crissangel.com
1024071	criss angel mind freak	2006-05-14 12:53:35	8	http://www.imdb.com
1024071	06-06-06	2006-05-14 22:29:11	1	http://www.timesonline.co.uk
1024071	show and sell auto fort pierce fl	2006-05-15 16:58:53	1	http://www.traderonline.com
1024071	barry bonds homerun ball 714 for sale	2006-05-25 16:25:41	5	http://www.sportsnet.ca
1024071	ufc 60 live results	2006-05-27 23:00:38	4	http://www.prowrestling.com
1024071	ufc 60 live play by play	2006-05-27 23:07:16	4	http://www.24wrestling.com
1024071	how to tell a fake rolex	2006-05-29 14:53:53	1	http://www.aplusmodel.com
1024071	how to tell a fake rolex	2006-05-29 14:53:53	8	http://www.inc.com
1024071	locating serial number on rolex	2006-05-30 21:51:34	1	http://www.qualitytyme.net

what
kinds of
things
could we
do with
this?

(AOL query-log) 14

What does a Search-Log Look Like?

1024071	taraji henson	2006-03-02 00:28:45	4	http://www.tv.com
1024071	taraji henson	2006-03-02 00:28:45	1	http://www.imdb.com
1024071	the flavor of love vh1	2006-03-02 00:31:01	1	http://www.vh1.com
1024071	the flavor of love hoops	2006-03-02 00:38:32	1	http://www.vh1realityworld.com
1024071	beyonce	2006-03-02 22:42:05	1	http://www.beyonceonline.com
1024071	beyonce	2006-03-02 22:42:05	6	http://www.imdb.com
1024071	afc fighting	2006-03-04 22:35:33	2	http://sfuk.tripod.com
1024071	din thomas march 4th	2006-03-05 23:38:54	1	http://www.mmaringreport.com
1024071	mfc march 4th results	2006-03-05 23:45:49	3	http://www.mmaringreport.com
1024071	mfc march 4th results	2006-03-05 23:45:49	9	http://man-magazine.com
1024071	unc basketball roster	2006-03-09 23:45:15	2	http://tarheelblue.collegesports.com
1024071	unc basketball roster	2006-03-09 23:45:15	2	http://tarheelblue.collegesports.com
1024071	nit free picks	2006-03-15 14:02:21	1	http://www.docsports.com
1024071	1490 am radio	2006-03-15 14:48:01	8	http://www.1490wwpr.com
1024071	1490 am radio fl	2006-03-15 14:50:08	2	http://www.ontheradio.net
1024071	benihanas	2006-03-16 17:27:25	1	http://www.benihana.com
1024071	2006 winter music fest miami fl	2006-03-22 00:35:20	1	http://www.wintermusicconference.com
1024071	hotmail	2006-04-01 18:49:02	1	http://www.hotmail.com
1024071	my space	2006-04-02 01:21:41	1	http://www.myspace.com
1024071	my space	2006-04-02 15:59:20	1	http://www.myspace.com
1024071	my space	2006-04-02 22:03:10	1	http://www.myspace.com
1024071	nba jams super nintendo cheats	2006-04-03 21:06:11	2	http://www.elook.org
1024071	my space	2006-04-03 21:16:00	1	http://www.myspace.com
1024071	charlie's dodge fort pierce	2006-05-08 20:06:17	1	http://www.dealernet.com
1024071	charlie's dodge of fort pierce used cars	2006-05-08 20:09:27	2	http://www.automotive.com
1024071	justin timberlake new album	2006-05-12 16:21:36	4	http://www.mtv.com
1024071	mike epps	2006-05-13 19:45:56	6	http://www.hollywood.com
1024071	mike epps bio	2006-05-13 19:51:05	4	http://movies.aol.com
1024071	mike epps bio	2006-05-13 19:51:05	9	http://www.moono.com
1024071	mike epps bio	2006-05-13 19:55:56	14	http://video.barnesandnoble.com
1024071	mike epps bio	2006-05-13 19:55:56	21	http://www.hbo.com
1024071	mike epps bio	2006-05-13 20:01:06	24	http://www.vh1.com
1024071	mind freak	2006-05-14 00:46:18	10	http://video.google.com
1024071	criss angel mind freak	2006-05-14 12:53:35	1	http://www.crissangel.com
1024071	criss angel mind freak	2006-05-14 12:53:35	8	http://www.imdb.com
1024071	06-06-06	2006-05-14 22:29:11	1	http://www.timesonline.co.uk
1024071	show and sell auto fort pierce fl	2006-05-15 16:58:53	1	http://www.traderonline.com
1024071	barry bonds homerun ball 714 for sale	2006-05-25 16:25:41	5	http://www.sportsnet.ca
1024071	ufc 60 live results	2006-05-27 23:00:38	4	http://www.prowrestling.com
1024071	ufc 60 live play by play	2006-05-27 23:07:16	4	http://www.24wrestling.com
1024071	how to tell a fake rolex	2006-05-29 14:53:53	1	http://www.aplusmodel.com
1024071	how to tell a fake rolex	2006-05-29 14:53:53	8	http://www.inc.com
1024071	locating serial number on rolex	2006-05-30 21:51:34	1	http://www.qualitytyme.net

are these
queries
independent?

(AOL query-log) 15

Search Sessions

- Search is a “dialogue” between a user and a search engine
 - ▶ **user:** query
 - ▶ **search engine:** search results
 - ▶ **user:** reformulated query
 - ▶ **search engine:** new search results
- Each “dialogue” is called a search session
- Each dialogue corresponds to an information need (at some level of granularity)
- A dialogue ends when the user is satisfied or gives up

Search Sessions

- **Question:** what proportion of search sessions result in user-satisfaction?
- The answer may be in the search log
- But, first, we have to recover each individual dialogue
- Requires some amount of “detective work”
- The simplest approaches assume that same-dialogue queries are sequential
- In other words, users engage in one dialogue at a time
- Are there environments where this is or is not a valid assumption?

Search Sessions

1024071	taraji henson	2006-03-02 00:28:45	4	http://www.tv.com
1024071	taraji henson	2006-03-02 00:28:45	1	http://www.imdb.com
1024071	the flavor of love vh1	2006-03-02 00:31:01	1	http://www.vh1.com
1024071	the flavor of love hoops	2006-03-02 00:38:32	1	http://www.vh1realityworld.com
1024071	beyonce	2006-03-02 22:42:05	1	http://www.beyonceonline.com
1024071	beyonce	2006-03-02 22:42:05	6	http://www.imdb.com
1024071	afc fighting	2006-03-04 22:35:33	2	http://sfuk.tripod.com
1024071	din thomas march 4th	2006-03-05 23:38:54	1	http://www.mmaringreport.com
1024071	mfc march 4th results	2006-03-05 23:45:49	3	http://www.mmaringreport.com
1024071	mfc march 4th results	2006-03-05 23:45:49	9	http://man-magazine.com
1024071	unc basketball roster	2006-03-09 23:45:15	2	http://tarheelblue.collegesports.com
1024071	unc basketball roster	2006-03-09 23:45:15	2	http://tarheelblue.collegesports.com
1024071	nit free picks	2006-03-15 14:02:21	1	http://www.docsports.com
1024071	1490 am radio	2006-03-15 14:48:01	8	http://www.1490wwpr.com
1024071	1490 am radio fl	2006-03-15 14:50:08	2	http://www.ontheradio.net
1024071	benihanas	2006-03-16 17:27:25	1	http://www.benihana.com
1024071	2006 winter music fest miami fl	2006-03-22 00:35:20	1	http://www.wintermusicconference.com
1024071	hotmail	2006-04-01 18:49:02	1	http://www.hotmail.com
1024071	my space	2006-04-02 01:21:41	1	http://www.myspace.com
1024071	my space	2006-04-02 15:59:20	1	http://www.myspace.com
1024071	my space	2006-04-02 22:03:10	1	http://www.myspace.com
1024071	nba jams super nintendo cheats	2006-04-03 21:06:11	2	http://www.elook.org
1024071	my space	2006-04-03 21:16:00	1	http://www.myspace.com
1024071	charlie's dodge fort pierce	2006-05-08 20:06:17	1	http://www.dealernet.com
1024071	charlie's dodge of fort pierce used cars	2006-05-08 20:09:27	2	http://www.automotive.com
1024071	justin timberlake new album	2006-05-12 16:21:36	4	http://www.mtv.com
1024071	mike epps	2006-05-13 19:45:56	6	http://www.hollywood.com
1024071	mike epps bio	2006-05-13 19:51:05	4	http://movies.aol.com
1024071	mike epps bio	2006-05-13 19:51:05	9	http://www.moono.com
1024071	mike epps bio	2006-05-13 19:55:56	14	http://video.barnesandnoble.com
1024071	mike epps bio	2006-05-13 19:55:56	21	http://www.hbo.com
1024071	mike epps bio	2006-05-13 20:01:06	24	http://www.vh1.com
1024071	mind freak	2006-05-14 00:46:18	10	http://video.google.com
1024071	criss angel mind freak	2006-05-14 12:53:35	1	http://www.crissangel.com
1024071	criss angel mind freak	2006-05-14 12:53:35	8	http://www.imdb.com
1024071	06-06-06	2006-05-14 22:29:11	1	http://www.timesonline.co.uk
1024071	show and sell auto fort pierce fl	2006-05-15 16:58:53	1	http://www.traderonline.com
1024071	barry bonds homerun ball 714 for sale	2006-05-25 16:25:41	5	http://www.sportsnet.ca
1024071	ufc 60 live results	2006-05-27 23:00:38	4	http://www.prowrestling.com
1024071	ufc 60 live play by play	2006-05-27 23:07:16	4	http://www.24wrestling.com
1024071	how to tell a fake rolex	2006-05-29 14:53:53	1	http://www.aplusmodel.com
1024071	how to tell a fake rolex	2006-05-29 14:53:53	8	http://www.inc.com
1024071	locating serial number on rolex	2006-05-30 21:51:34	1	http://www.qualitytyme.net

Search Sessions

1024071	taraji henson	2006-03-02 00:28:45	4	http://www.tv.com
1024071	taraji henson	2006-03-02 00:28:45	1	http://www.imdb.com
1024071	the flavor of love vh1	2006-03-02 00:31:01	1	http://www.vh1.com
1024071	the flavor of love hoops	2006-03-02 00:38:32	1	http://www.vh1realityworld.com
1024071	beyonce	2006-03-02 22:42:05	1	http://www.beyonceonline.com
1024071	beyonce	2006-03-02 22:42:05	6	http://www.imdb.com
1024071	afc fighting	2006-03-04 22:35:33	2	http://sfuk.tripod.com
1024071	din thomas march 4th	2006-03-05 23:38:54	1	http://www.mmaringreport.com
1024071	mfc march 4th results	2006-03-05 23:45:49	3	http://www.mmaringreport.com
1024071	mfc march 4th results	2006-03-05 23:45:49	9	http://man-magazine.com
1024071	unc basketball roster	2006-03-09 23:45:15	2	http://tarheelblue.collegesports.com
1024071	unc basketball roster	2006-03-09 23:45:15	2	http://tarheelblue.collegesports.com
1024071	nit free picks	2006-03-15 14:02:21	1	http://www.docsports.com
1024071	1490 am radio	2006-03-15 14:48:01	8	http://www.1490wwpr.com
1024071	1490 am radio fl	2006-03-15 14:50:08	2	http://www.ontheradio.net
1024071	benihanas	2006-03-16 17:27:25	1	http://www.benihana.com
1024071	2006 winter music fest miami fl	2006-03-22 00:35:20	1	http://www.wintermusicconference.com
1024071	hotmail	2006-04-01 18:49:02	1	http://www.hotmail.com
1024071	my space	2006-04-02 01:21:41	1	http://www.myspace.com
1024071	my space	2006-04-02 15:59:20	1	http://www.myspace.com
1024071	my space	2006-04-02 22:03:10	1	http://www.myspace.com
1024071	nba jams super nintendo cheats	2006-04-03 21:06:11	2	http://www.elook.org
1024071	my space	2006-04-03 21:16:00	1	http://www.myspace.com
1024071	charlie's dodge fort pierce	2006-05-08 20:06:17	1	http://www.dealernet.com
1024071	charlie's dodge of fort pierce used cars	2006-05-08 20:09:27	2	http://www.automotive.com
1024071	justin timberlake new album	2006-05-12 16:21:36	4	http://www.mtv.com

Search Sessions

1024071	mike epps	2006-05-13 19:45:56	6	http://www.hollywood.com
1024071	mike epps bio	2006-05-13 19:51:05	4	http://movies.aol.com
1024071	mike epps bio	2006-05-13 19:51:05	9	http://www.moono.com
1024071	mike epps bio	2006-05-13 19:55:56	14	http://video.barnesandnoble.com
1024071	mike epps bio	2006-05-13 19:55:56	21	http://www.hbo.com
1024071	mike epps bio	2006-05-13 20:01:06	24	http://www.vh1.com
1024071	mind freak	2006-05-14 00:46:18	10	http://video.google.com
1024071	criss angel mind freak	2006-05-14 12:53:35	1	http://www.crissangel.com
1024071	criss angel mind freak	2006-05-14 12:53:35	8	http://www.imdb.com
1024071	06-06-06	2006-05-14 22:29:11	1	http://www.timesonline.co.uk
1024071	show and sell auto fort pierce fl	2006-05-15 16:58:53	1	http://www.traderonline.com
1024071	barry bonds homerun ball 714 for sale	2006-05-25 16:25:41	5	http://www.sportsnet.ca
1024071	ufc 60 live results	2006-05-27 23:00:38	4	http://www.prowrestling.com
1024071	ufc 60 live play by play	2006-05-27 23:07:16	4	http://www.24wrestling.com
1024071	how to tell a fake rolex	2006-05-29 14:53:53	1	http://www.aplusmodel.com
1024071	how to tell a fake rolex	2006-05-29 14:53:53	8	http://www.inc.com
1024071	locating serial number on rolex	2006-05-30 21:51:34	1	http://www.qualitytyme.net

Heuristics for Recovering Search Sessions

- **Time difference:** subsequent queries are part of the same session if the difference between time-stamps is $< t$
 - ▶ 30 minutes works well for library search
 - ▶ no value is better than random on web search!
- **Common term:** subsequent queries are part of the same session if they have at least one common term
 - ▶ high precision, low recall strategy

Heuristics for Recovering Search Sessions

- **Rewrite classes:** subsequent queries are part of the same session if they follow common reformulation patterns
 - ▶ add a term, delete a term, replace a term
 - ▶ Q1: “gop debate”
 - ▶ Q2: “gop debate herman cain”
 - ▶ Q3: “herman cain calls pelosi a princess”
 - ▶ Q1 and Q3 have no terms in common, but are still considered part of the same session.
 - ▶ Q1-Q2 and Q2-Q3 follow common reformulation patterns

Hierarchical Search Sessions

- Search tasks are not “flat”, they’re hierarchical
- Jones and Klinker defined a two-level hierarchy (draw this on the board)
- A similar abstraction is made in discourse theory
- **Goal:** a set of queries with the same criteria for success
- **Mission:** an extended information need that requires satisfying potentially multiple goals

Missions and Goals

example

- **Mission:** plan a hiking trip near San Francisco
 - ▶ **Goal:** find list of nearby parks
 - ▶ **Goal:** driving directions to ano nuevo state reserve
 - ▶ **Goal:** driving directions to china camp state park
 - ▶ **Goal:** find weather information

Missions and Goals

- 312 Yahoo! users tracked over a span of three days
 - ▶ Some missions span more time (e.g., IR lit-review)
- Annotators organized queries into missions and goals
- Some statistics:

	Goals	Missions
Num queries		
min	1	1
max	52	233
median	2	2
Duration		
min	0 mins	0 mins
max	71 hours	71 hours
median	0.42 secs	38 secs

Missions and Goals

- 15% of goals had at least 30 minutes of inactivity
- 16% of goals were interleaved with other goals
- 17% of missions were interleaved with other missions
- Some missions lasted beyond the 3-day window
- Any approach that segments a search-log sequentially is likely to “break-up” queries that have a common objective (i.e., a common mission or a common goal)

Missions and Goals

their approach

- Two possibly non-sequential queries can be predicted to belong to the same goal/mission by combining different types of evidence
 - ▶ time difference between queries
 - ▶ query-terms in common
 - ▶ the frequency with which other users issued the query-pair sequentially (this requires lots of data)
 - ▶ The degree of overlap between the search results from both queries

Missions and Goals

why is this interesting?

- What do long-term missions look like?
- Recognizing long-term missions may provide ways to evaluate search engines by their ability to “point you in the right direction”
 - ▶ Evaluation based on single-query satisfaction misses the “big picture”
- TREC has a Session-Track that explores this issue
- Search engines may be able to customize results based on your “open” missions

What about clicks?

- **Explicit relevance feedback:** asking the user whether a result is relevant/non-relevant to a query
- **Implicit relevance feedback:** predicting relevance based on user interactions
- People don't like to provide explicit feedback
- Can we use clicks to predict relevance?
 - ▶ non-obtrusive
 - ▶ inexpensive
 - ▶ lots of data

Implicit Relevance Feedback

- **Question:** can we use clicks to predict relevance?
- Answering this question requires understanding how users behave
- Are all clicks equally predictive of relevance?
- Are there other “forces” (other than relevance) that motivate us to click on results?

Implicit Relevance Feedback

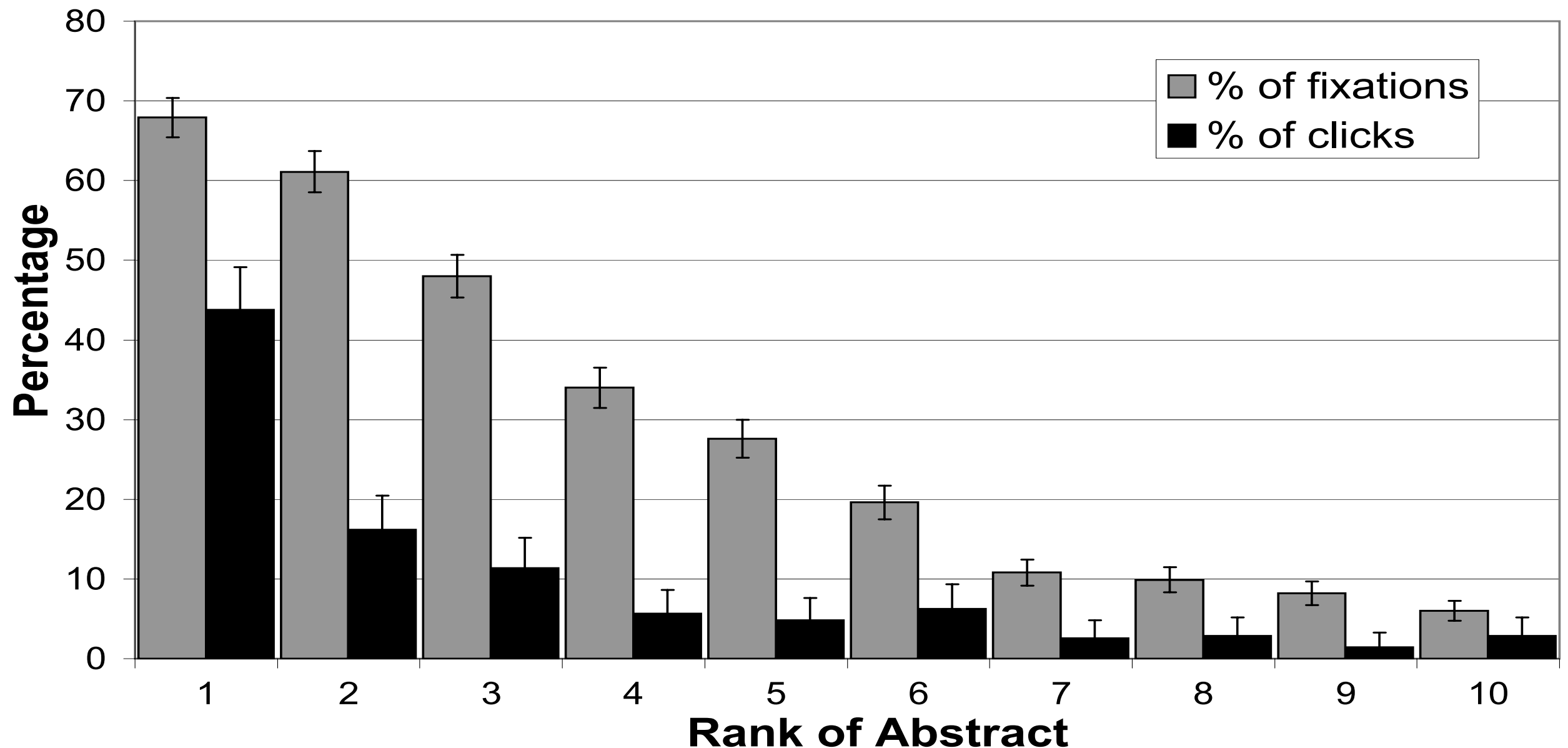
(Joachims *et al.*, 2005)

- First Study
 - ▶ 34 subjects (all Cornell undergrads)
 - ▶ 10 search tasks (5 navigational + 5 informational)
 - ▶ top-10 Google results
 - ▶ Eye tracking + click-logging
 - ▶ Fixation: spatially stable gaze lasting approximately 0.2-0.3 seconds

Implicit Relevance Feedback

(Joachims *et al.*, 2005)

- Which results do users view and click?



- % of searches where user fixated/clicked a result in rank r

Eye Tracking

(Joachims *et al.*, 2005)

- Which results do users view?
- Most people view the first two results (almost equally)
- Fewer than half view the third result!
- Only about 10% scroll down to view results below the fold!
- Views below the fold are fairly evenly distributed. Any ideas why?

Clicks

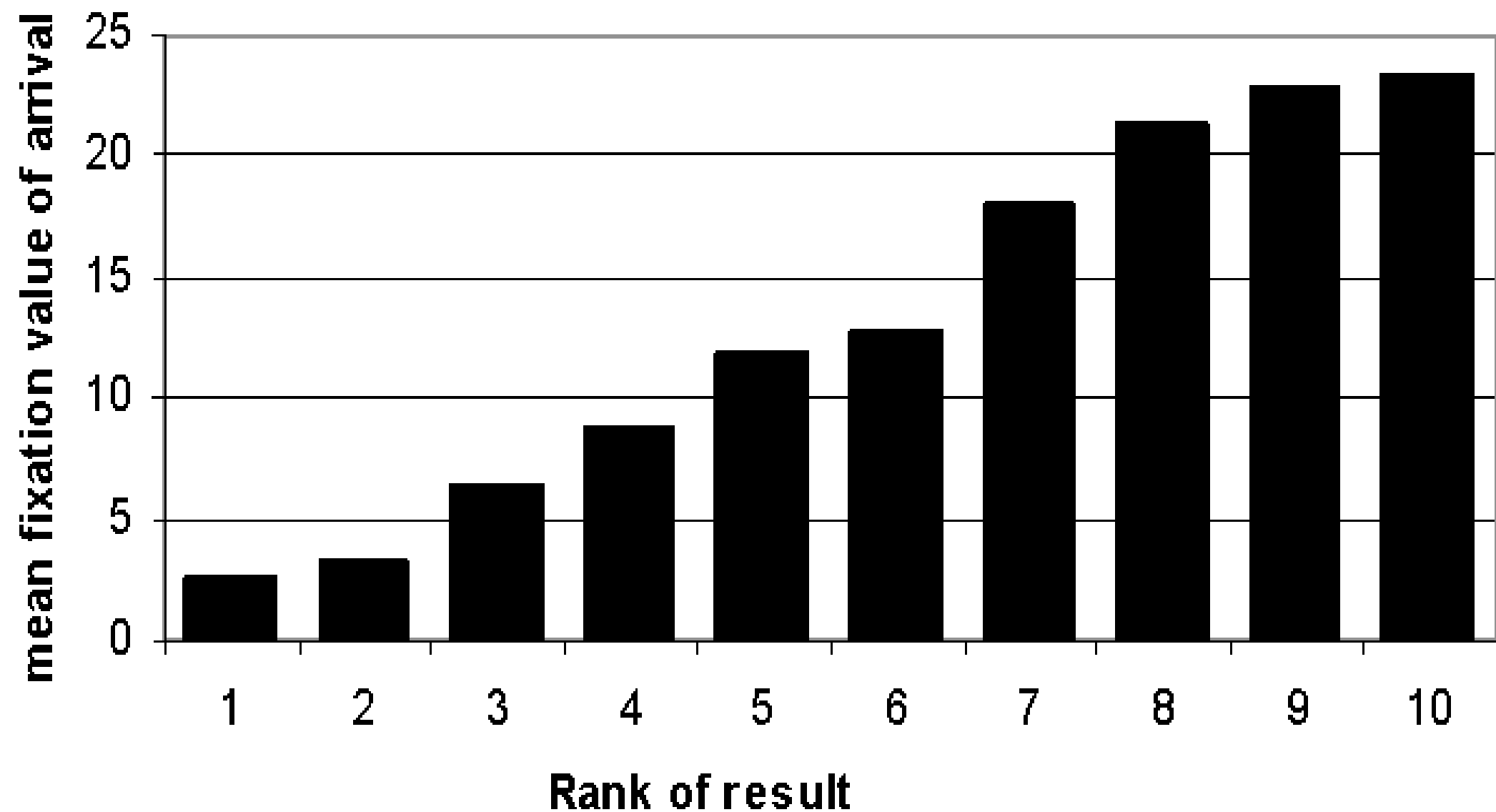
(Joachims *et al.*, 2005)

- Which results do users click?
- While the top-two results are viewed almost equally, the first result is clicked a lot more than the second
 - ▶ Why? Because the first result is better? Because people trust it more?
- Clicks below rank 3 are fairly evenly distributed

Implicit Relevance Feedback

(Joachims *et al.*, 2005)

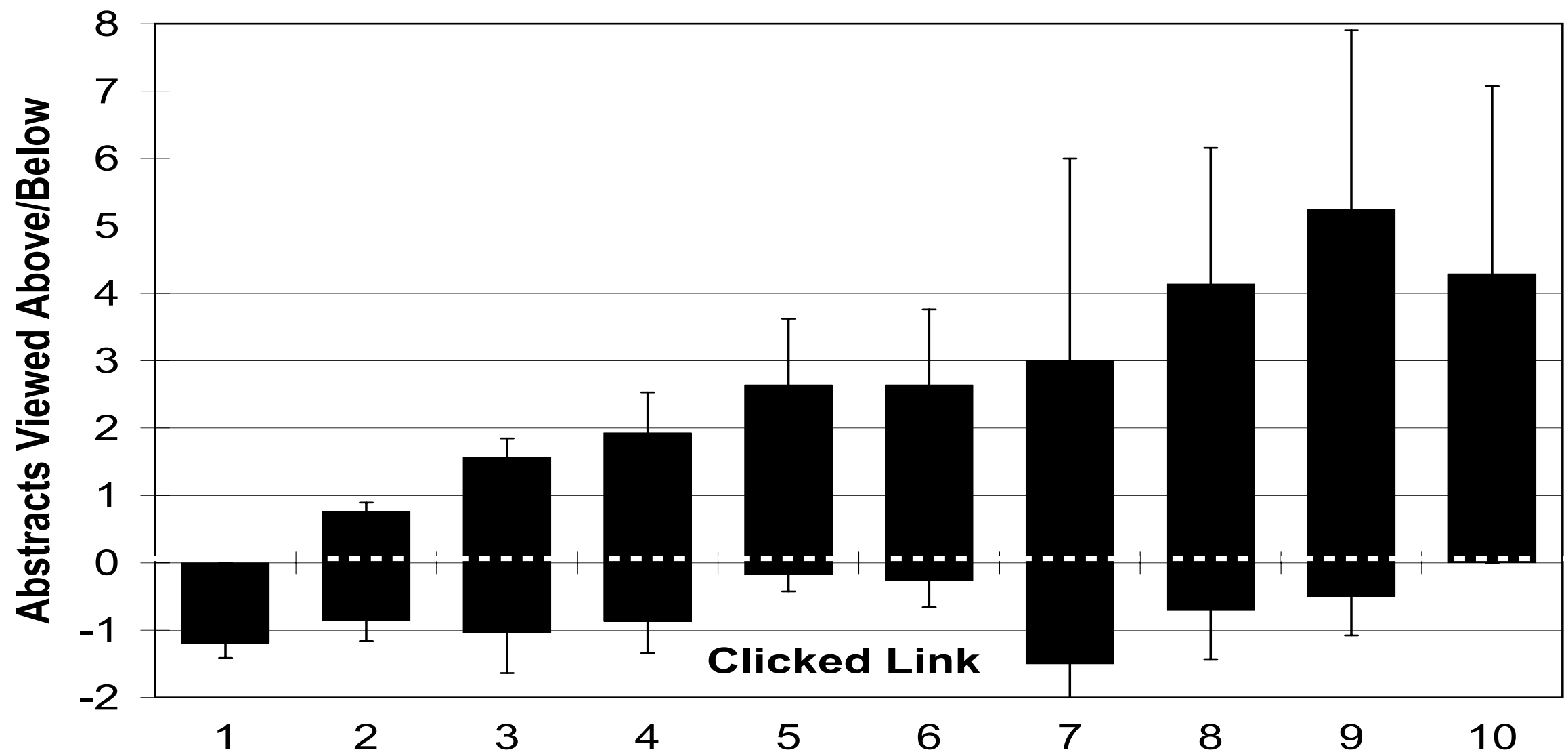
- Users scan results from top to bottom



Implicit Relevance Feedback

(Joachims *et al.*, 2005)

- Which results do users evaluate before clicking?



Implicit Relevance Feedback

(Joachims *et al.*, 2005)

- Which results do users evaluate before clicking?

Viewed Rank	Clicked Rank					
	1	2	3	4	5	6
1	90.6%	76.2%	73.9%	60.0%	54.5%	45.5%
2	56.8%	90.5%	82.6%	53.3%	63.6%	54.5%
3	30.2%	47.6%	95.7%	80.0%	81.8%	45.5%
4	17.3%	19.0%	47.8%	93.3%	63.6%	45.5%
5	8.6%	14.3%	21.7%	53.3%	100.0%	72.7%
6	4.3%	4.8%	8.7%	33.3%	18.2%	81.8%

Implicit Relevance Feedback

(Joachims *et al.*, 2005)

- Users tend to view higher-ranks before clicking on a result
- They look at the one ranked immediately following the clicked result (if there is one)
- This is especially the case for rank 7

Implicit Relevance Feedback

(Joachims *et al.*, 2005)

- Second Study
 - ▶ 34 subjects (all Cornell undergrads)
 - ▶ 10 search tasks (5 navigational + 5 informational)
 - ▶ top-10 Google results (all results judged by assessors)
 - ▶ 3 conditions
 - ▶ normal: Google results 1-10
 - ▶ swapped: Google results 1 and 2 swapped
 - ▶ reversed: results 1-10 reversed

Implicit Relevance Feedback

(Joachims *et al.*, 2005)

- Are clicks influenced by relevance (or just rank)?
- Relevance matters
- In the “reversed” condition (Google results 1-10 reversed), lower-ranked results were clicked more often than expected

Implicit Relevance Feedback

(Joachims *et al.*, 2005)

- So, a click = a relevance judgement?
- Not quite
- Users click on rank 1 more than rank 2 even when rank 2 is more relevant (Trust Bias!)

“normal”	l_1^-, l_2^-	l_1^+, l_2^-	l_1^-, l_2^+	l_1^+, l_2^+	total
$\text{rel}(l_1) > \text{rel}(l_2)$	15	19	1	1	36
$\text{rel}(l_1) < \text{rel}(l_2)$	11	5	2	2	20
$\text{rel}(l_1) = \text{rel}(l_2)$	19	9	1	0	29
total	45	33	4	3	85

“swapped”	l_1^-, l_2^-	l_1^+, l_2^-	l_1^-, l_2^+	l_1^+, l_2^+	total
$\text{rel}(l_1) > \text{rel}(l_2)$	11	15	1	1	28
$\text{rel}(l_1) < \text{rel}(l_2)$	17	10	7	2	36
$\text{rel}(l_1) = \text{rel}(l_2)$	36	11	3	0	50
total	64	36	11	3	114

Implicit Relevance Feedback

(Joachims *et al.*, 2005)

- So, if there's a bias in favor of the top results, how can we use clicks to predict relevance?
- It's difficult to use clicks to predict absolute relevance
- Clicks can be used, however, to predict pairwise preferences!

Implicit Relevance Feedback

(Joachims *et al.*, 2005)

Rank	1	2	3	4	5	6	7	8	9	10
Click	✓		✓				✓	✓		✓

- Click > Skip Above: ???
- Last Click > Skip Above: ???
- Click > Earlier Click: ???
- Click > Skip Previous: ???
- Click > No Click Next: ???

Implicit Relevance Feedback

(Joachims *et al.*, 2005)

Rank	1	2	3	4	5	6	7	8	9	10
Click	✓		✓				✓	✓		✓

- **Click > Skip Above:** $(3 > 2)$, $(7 > 2)$, $(7 > 4)$, $(7 > 5)$, $(7 > 6)$, $(8 > 2)$, $(8 > 4)$, $(8 > 5)$, $(8 > 6)$, $(10 > 2)$, $(10 > 4)$, $(10 > 5)$, $(10 > 6)$, $(10 > 9)$
- **Last Click > Skip Above:** $(10 > 2)$, $(10 > 4)$, $(10 > 5)$, $(10 > 6)$, $(10 > 9)$
- **Click > Earlier Click:** $(3 > 1)$, $(7 > 1)$, $(7 > 3)$, $(8 > 1)$, $(8 > 3)$, $(10 > 1)$, $(10 > 3)$, $(10 > 7)$, $(10 > 8)$
- **Click > Skip Previous:** $(3 > 2)$, $(7 > 6)$, $(10 > 9)$
- **Click > No Click Next:** $(1 > 2)$, $(3 > 4)$, $(8 > 9)$

Implicit Relevance Feedback

(Joachims *et al.*, 2005)

Explicit Feedback Data Strategy	Abstracts					Pages Phase II all
	Phase I “normal”	“normal”	“swapped”	“reversed”	all	
Inter-Judge Agreement	89.5	N/A	N/A	N/A	82.5	86.4
Click > Skip Above	80.8 \pm 3.6	88.0 \pm 9.5	79.6 \pm 8.9	83.0 \pm 6.7	83.1 \pm 4.4	78.2 \pm 5.6
Last Click > Skip Above	83.1 \pm 3.8	89.7 \pm 9.8	77.9 \pm 9.9	84.6 \pm 6.9	83.8 \pm 4.6	80.9 \pm 5.1
Click > Earlier Click	67.2 \pm 12.3	75.0 \pm 25.8	36.8 \pm 22.9	28.6 \pm 27.5	46.9 \pm 13.9	64.3 \pm 15.4
Click > Skip Previous	82.3 \pm 7.3	88.9 \pm 24.1	80.0 \pm 18.0	79.5 \pm 15.4	81.6 \pm 9.5	80.7 \pm 9.6
Click > No Click Next	84.1 \pm 4.9	75.6 \pm 14.5	66.7 \pm 13.1	70.0 \pm 15.7	70.4 \pm 8.0	67.4 \pm 8.2

- % agreement with pairwise preferences derived from relevance judgements from assessors
- **Best strategy:** a clicked result is more relevant than all higher ranked results that were skipped (not clicked)
 - ▶ produces lots of preferences that also happen to agree with explicit judgements

Conclusions and Implications

(Joachims *et al.*, 2005)

- Users' clicking decisions are influenced by relevance
- But, they're also biased in favor of the top results (the first one seen and the ones trusted)
- Clicks should not be treated in 'absolute' terms
- However, they can be used to predict 'relative' (pairwise) relevance

Implicit Feedback

- As we've seen, some clever tricks can be applied to implicit feedback signals to infer relevance
- Now, can we use implicit feedback to improve a search engine's ranking performance?

First, What is Machine Learning?

- Computer scientists design algorithms
- An algorithm is a recipe for solving a problem (e.g., sorting a list of numbers from lowest to greatest)
- (Supervised) machine learning algorithms are algorithms that learn to make predictions from examples
- Machine learning applications include:
 - classifying documents into a topic
 - detecting positive/negative opinions in text
 - learning to rank documents in response to a query

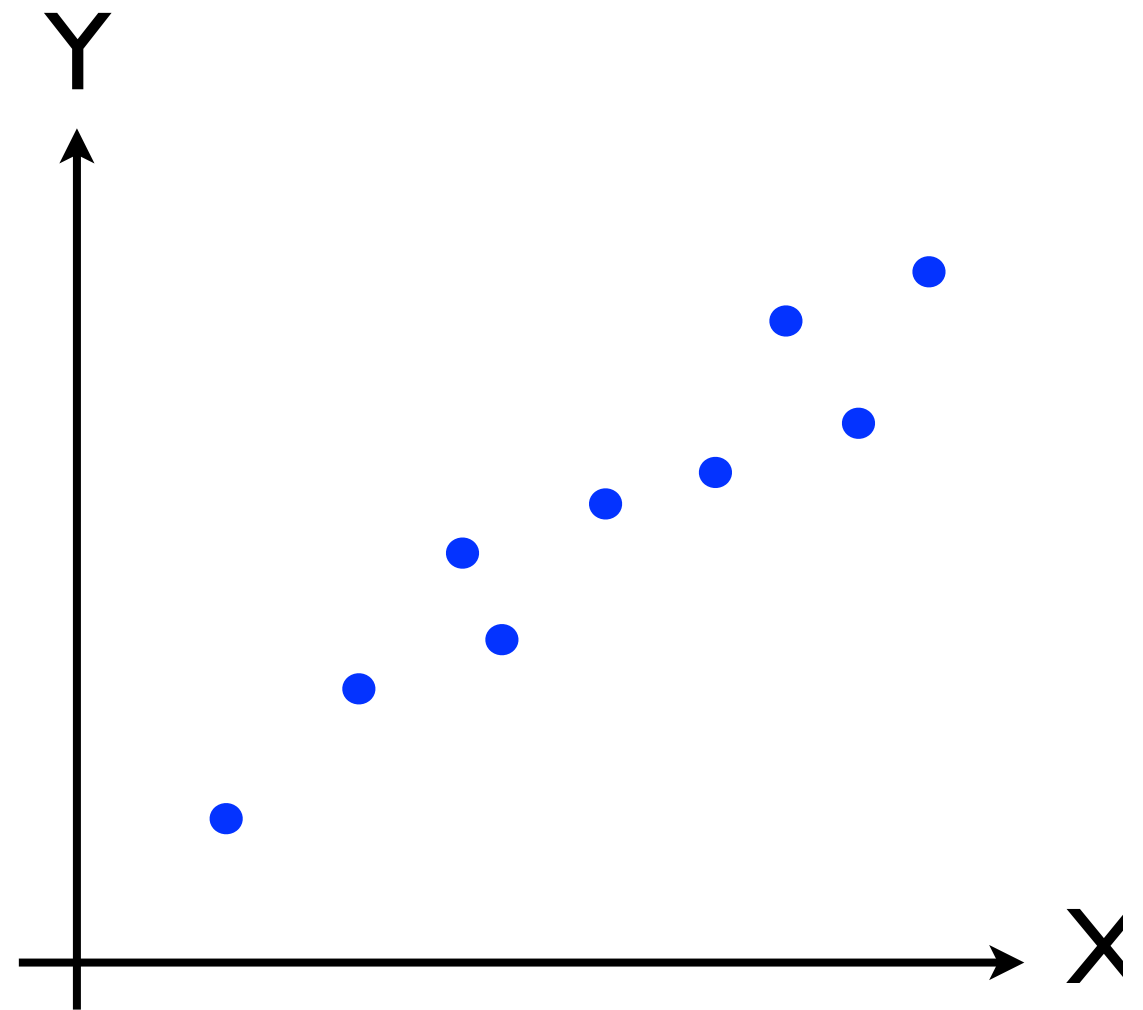
Machine Learning Components

- **Input:** the thing (or things) we observe
 - ▶ we decide what these things are (you don't need a computer science background)
- **Output:** the thing we want to learn to predict
- **Training Data:** a bunch of input-output examples
- **Machine Learning Algorithm:** an algorithm that uses training data to learn a predictive model
- Once trained, the model can predict an output for a previously unseen input
- If things work, the predicted output value will match the true output value

A Simple Machine Learning Algorithm

linear regression

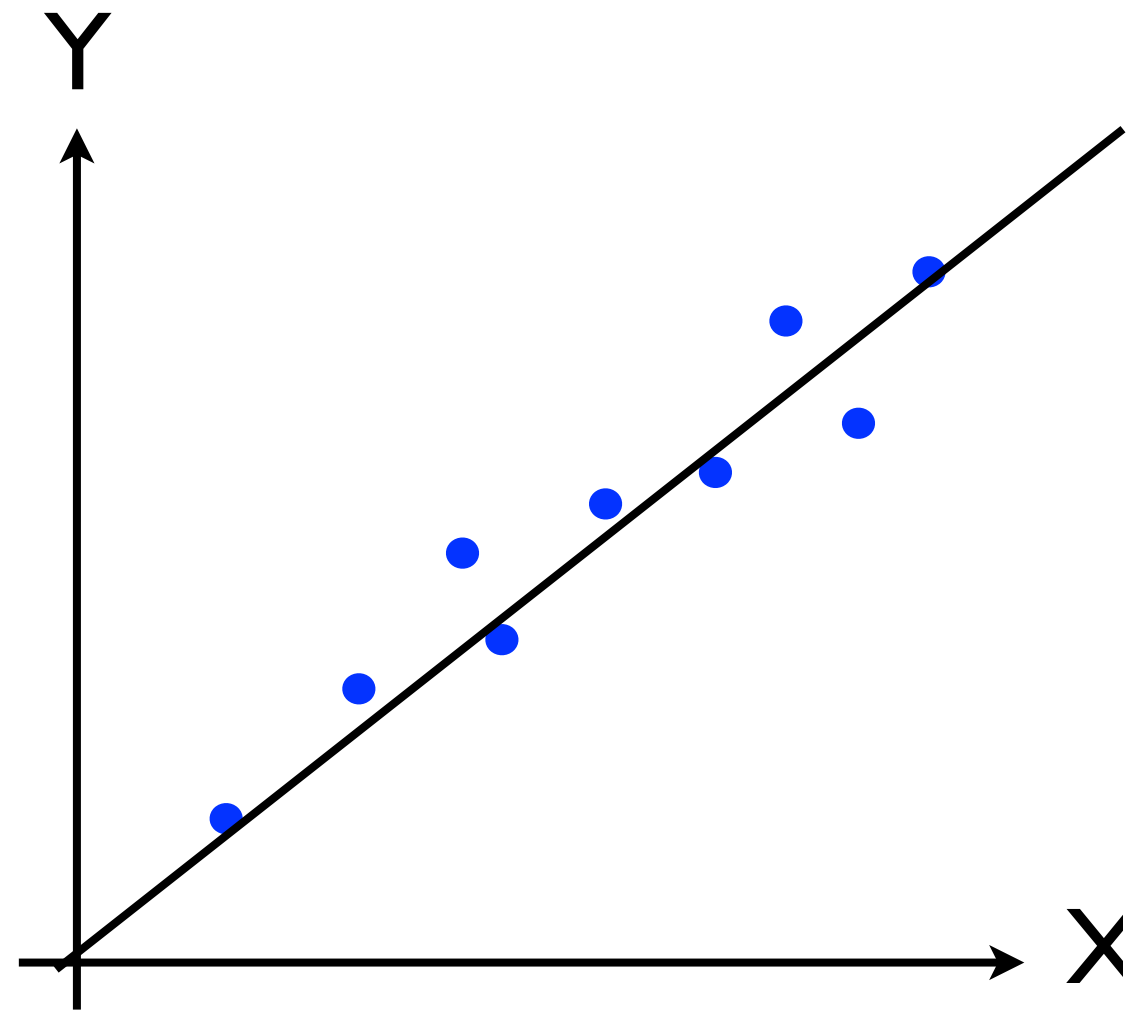
- **Objective:** given a value of X , predict the value of Y
- **Given:** a bunch of X - Y pairs
- **Method:**
 - ▶ assume a linear relation between X and Y
 - ▶ use the given data to “learn” a linear relation
 - ▶ given a previously unseen value of X , predict the value of Y



A Simple Machine Learning Algorithm

linear regression

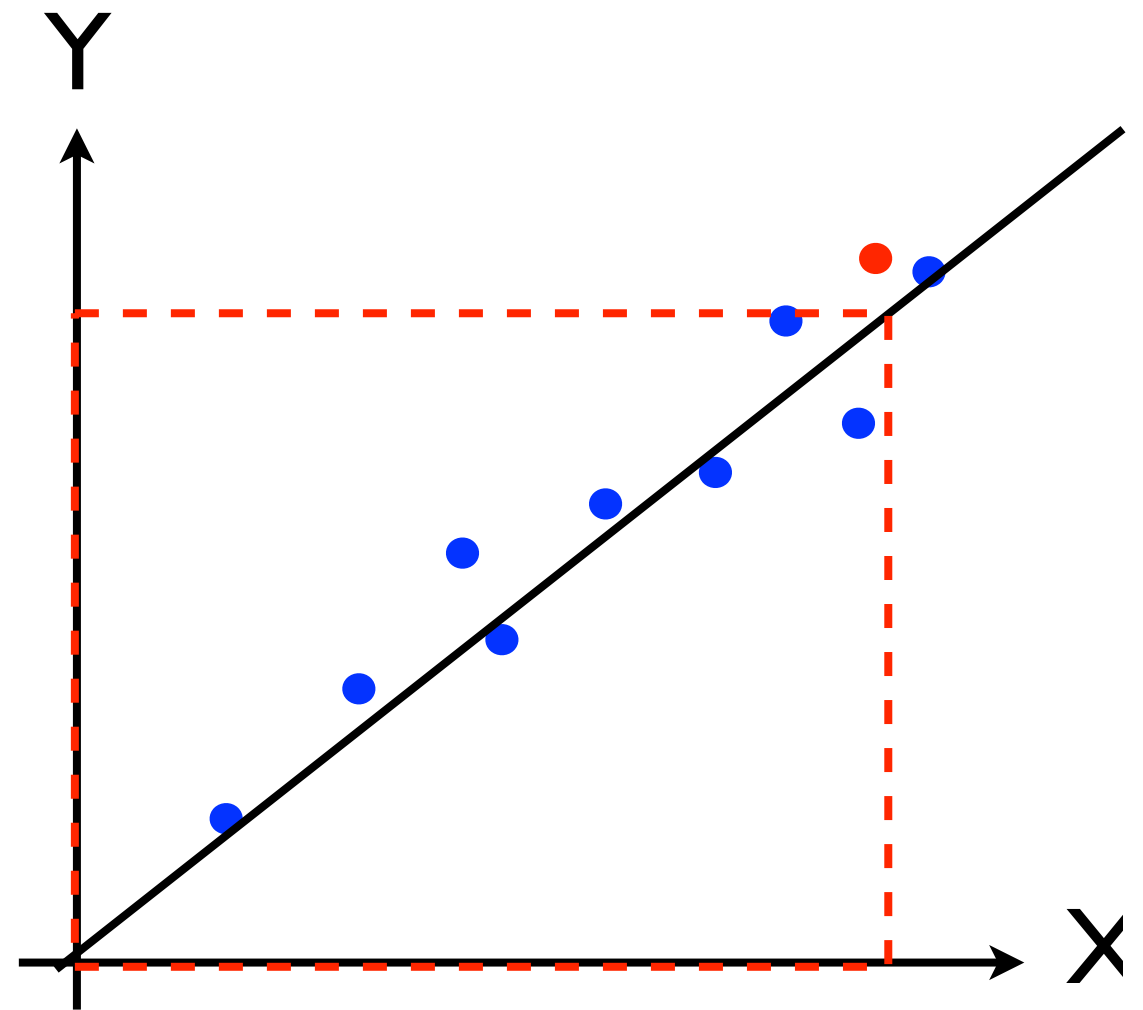
- **Objective:** given a value of X , predict the value of Y
- **Given:** a bunch of X - Y pairs
- **Method:**
 - ▶ assume a linear relation between X and Y
 - ▶ use the given data to “learn” a linear relation
 - ▶ given a previously unseen value of X , predict the value of Y



A Simple Machine Learning Algorithm

linear regression

- **Objective:** given a value of X , predict the value of Y
- **Given:** a bunch of X - Y pairs
- **Method:**
 - ▶ assume a linear relation between X and Y
 - ▶ use the given data to “learn” a linear relation
 - ▶ given a previously unseen value of X , predict the value of Y



Learning to Rank

- “Learning to rank” algorithms learn to predict document preferences for a given query
- **Input:** a document pair $(d1, d2)$ and a query q
- **Output:** the relative relevance between $d1$ and $d2$ given q
- The input is $(d1, d2, q)$ converted into a set of signals (a.k.a., features)
- The output ranking attempts to be maximally consistent with the model’s preference predictions

Learning to Rank

input signals


- Query: **bathing** a **cat**
- The important query terms occur frequently
- Both terms occur
- Terms occur close together
- Terms occur in the title
- Terms occur in the URL

www.wikihow.com/bathe-your-cat

How to **Bathe** a **Cat**
Edited by Kimberly and 107 others

Article Edit Discuss History Tweet 22 Like 50

Even though they mostly keep themselves clean, most **cats** need a **bath** every now and then. Everyone knows that **cats** hate getting wet. **Cats** generally keep themselves clean, and therefore should not be **bathed** any more often than is absolutely necessary. But they sometimes get especially dirty, get bombarded by fleas or ticks, or get into substances that are toxic or otherwise harmful. On such occasions, it is a good idea to **bathe** your **cat**. Here's how to keep your feline fresh and lovely as painlessly as possible!




Ads by Google
[Pet Allergy Info](#) Information On The Symptoms And Triggers For Indoor Allergies. ZYRTEC.com

Steps

1 **Decide, whether your **cat** really needs a **bath**.** There may be other ways to clean your **cat**, such as **brushing the cat**, combing or even rubbing it down with a cloth.

2 **Wear appropriate clothing.** Not only is it important to wash your **cat** but it is important to be safe from any possible scratches, minor or major. A jumper or a long-sleeved shirt must be worn so that your **cat** has no bare skin to scratch. If you don't have one, you could wear long-sleeved gloves. It is also a good idea to wear clothing that isn't new.



3 **Get at least two people involved in washing your **cat**, especially if your **cat** is rather strong and can kick and wriggle its way out of your hands.** One person should hold all four legs and hold the **cat's** jaw so it can't open its mouth to bite you, but be sure you don't hold it really tight so it can't breathe. Hold your cat firmly so it cannot wriggle out from your grip.

⋮

16 **Reward your **cat**.** Give him/her their favorite canned food or catnip or treats, and he/she will come to realize that there is a good side to being **bathed**.

Learning to Rank

input signals

- Terms occur in **anchor-text** that links to the page

How to Raise Angora Cats

Edited by Elizabeth Knudsen and 3 others

- Groom the cat regularly.** The angora cats usually have medium length fur, and need to be groomed at least 3-4 times at week. This does not include bathing, you should only **bathe the cat** if it gets specially dirty, or for a show.

How to Bathe a Cat

Edited by Kimberly and 107 others

Article

Edit

Discuss

History

Tweet 22

Like 50

Even though they mostly keep themselves clean, most **cats** need a **bath** every now and then. Everyone knows that **cats** hate getting wet. **Cats** generally keep themselves clean, and therefore should not be **bathed** any more often than is absolutely necessary. But they sometimes get especially dirty, get bombarded by fleas or ticks, or get into substances that are toxic or otherwise harmful. On such occasions, it is a good idea to **bathe** your **cat**. Here's how to keep your feline fresh and lovely as painlessly as possible!



Ads by Google

Pet Allergy Info Information On The Symptoms And Triggers For Indoor Allergies. ZYRTEC.com

Steps

Edit

- Decide, whether your **cat** really needs a **bath**.** There may be other ways to clean your **cat**, such as **brushing the cat**, combing or even rubbing it down with a cloth.
- Wear appropriate clothing.** Not only is it important to wash your **cat** but it is important to be safe from any possible scratches, minor or major. A jumper or a long-sleeved shirt must be worn so that your **cat** has no bare skin to scratch. If you don't have one, you could wear long-sleeved gloves. It is also a good idea to wear clothing that isn't new.
- Get at least two people involved in washing your **cat**, especially if your **cat** is rather strong and can kick and wriggle its way out of your hands.** One person should hold all four legs and hold the **cat's** jaw so it can't open its mouth to bite you, but be sure you don't hold it really tight so it can't breathe. Hold your cat firmly so it cannot wriggle out from your grip.
- ⋮
- Reward your **cat**.** Give him/her their favorite canned food or catnip or treats, and he/she will come to realize that there is a good side to being **bathed**.



- Same language as query
- High term co-occurrence at the top of the document
- Other terms semantically related to query-terms (e.g., feline, wash)

Learning to Rank

input signals

- The query is similar to other queries associated with clicks on this document
- The document is similar to other documents associated with clicks for this query

How to Bathe a Cat

Edited by Kimberly and 107 others

Article

Edit

Discuss

History

Tweet

22

Like

50

Even though they mostly keep themselves clean, most cats need a bath every now and then. Everyone knows that cats hate getting wet. Cats generally keep themselves clean, and therefore should not be bathed any more often than is absolutely necessary. But they sometimes get especially dirty, get bombarded by fleas or ticks, or get into substances that are toxic or otherwise harmful. On such occasions, it is a good idea to bathe your cat. Here's how to keep your feline fresh and lovely as painlessly as possible!



Ads by Google

[Pet Allergy Info](#) Information On The Symptoms And Triggers For Indoor Allergies.
ZYRTEC.com

Steps

Edit

- 1 **Decide, whether your cat really needs a bath.** There may be other ways to clean your cat, such as [brushing the cat](#), combing or even rubbing it down with a cloth.
- 2 **Wear appropriate clothing.** Not only is it important to wash your cat but it is important to be safe from any possible scratches, minor or major. A jumper or a long-sleeved shirt must be worn so that your cat has no bare skin to scratch. If you don't have one, you could wear long-sleeved gloves. It is also a good idea to wear clothing that isn't new.
- 3 **Get at least two people involved in washing your cat, especially if your cat is rather strong and can kick and wriggle its way out of your hands.** One person should hold all four legs and hold the cat's jaw so it can't open its mouth to bite you, but be sure you don't hold it really tight so it can't breathe. Hold your cat firmly so it cannot wriggle out from your grip.
- ⋮
- 16 **Reward your cat.** Give him/her their favorite canned food or catnip or treats, and he/she will come to realize that there is a good side to being bathed.



Learning to Rank

input signals

- Lots of in-links (endorsements)
- Non-spam properties:
 - ▶ grammatical sentences
 - ▶ no profanity
- Has good formatting

How to Bathe a Cat

Edited by Kimberly and 107 others

Article

Edit

Discuss

History

Tweet

22

Like

50

Even though they mostly keep themselves clean, most cats need a bath every now and then. Everyone knows that cats hate getting wet. Cats generally keep themselves clean, and therefore should not be bathed any more often than is absolutely necessary. But they sometimes get especially dirty, get bombarded by fleas or ticks, or get into substances that are toxic or otherwise harmful. On such occasions, it is a good idea to bathe your cat. Here's how to keep your feline fresh and lovely as painlessly as possible!



Ads by Google

[Pet Allergy Info](#) Information On The Symptoms And Triggers For Indoor Allergies.
ZYRTEC.com

Steps

Edit

- 1 **Decide, whether your cat really needs a bath.** There may be other ways to clean your cat, such as [brushing the cat](#), combing or even rubbing it down with a cloth.
- 2 **Wear appropriate clothing.** Not only is it important to wash your cat but it is important to be safe from any possible scratches, minor or major. A jumper or a long-sleeved shirt must be worn so that your cat has no bare skin to scratch. If you don't have one, you could wear long-sleeved gloves. It is also a good idea to wear clothing that isn't new.
- 3 **Get at least two people involved in washing your cat, especially if your cat is rather strong and can kick and wriggle its way out of your hands.** One person should hold all four legs and hold the cat's jaw so it can't open its mouth to bite you, but be sure you don't hold it really tight so it can't breathe. Hold your cat firmly so it cannot wriggle out from your grip.
- ⋮
- 16 **Reward your cat.** Give him/her their favorite canned food or catnip or treats, and he/she will come to realize that there is a good side to being bathed.



Learning to Rank

input signals

- Author attributes
- Peer-reviewed by many
- Reading-level appropriate for user community
- Has pictures
- Recently modified (fresh)
- Normal length
- From domain with other high-quality documents
- Ads by google ;-)

How to Bathe a Cat

Edited by Kimberly and 107 others

Article

Edit

Discuss

History

Tweet

22

Like

50

Even though they mostly keep themselves clean, most cats need a bath every now and then. Everyone knows that cats hate getting wet. Cats generally keep themselves clean, and therefore should not be bathed any more often than is absolutely necessary. But they sometimes get especially dirty, get bombarded by fleas or ticks, or get into substances that are toxic or otherwise harmful. On such occasions, it is a good idea to bathe your cat. Here's how to keep your feline fresh and lovely as painlessly as possible!



Ads by Google

[Pet Allergy Info](#) Information On The Symptoms And Triggers For Indoor Allergies.
ZYRTEC.com

Steps

Edit

- 1 **Decide, whether your cat really needs a bath.** There may be other ways to clean your cat, such as [brushing the cat](#), combing or even rubbing it down with a cloth.
- 2 **Wear appropriate clothing.** Not only is it important to wash your cat but it is important to be safe from any possible scratches, minor or major. A jumper or a long-sleeved shirt must be worn so that your cat has no bare skin to scratch. If you don't have one, you could wear long-sleeved gloves. It is also a good idea to wear clothing that isn't new.
- 3 **Get at least two people involved in washing your cat, especially if your cat is rather strong and can kick and wriggle its way out of your hands.** One person should hold all four legs and hold the cat's jaw so it can't open its mouth to bite you, but be sure you don't hold it really tight so it can't breathe. Hold your cat firmly so it cannot wriggle out from your grip.
- ⋮
- 16 **Reward your cat.** Give him/her their favorite canned food or catnip or treats, and he/she will come to realize that there is a good side to being bathed.



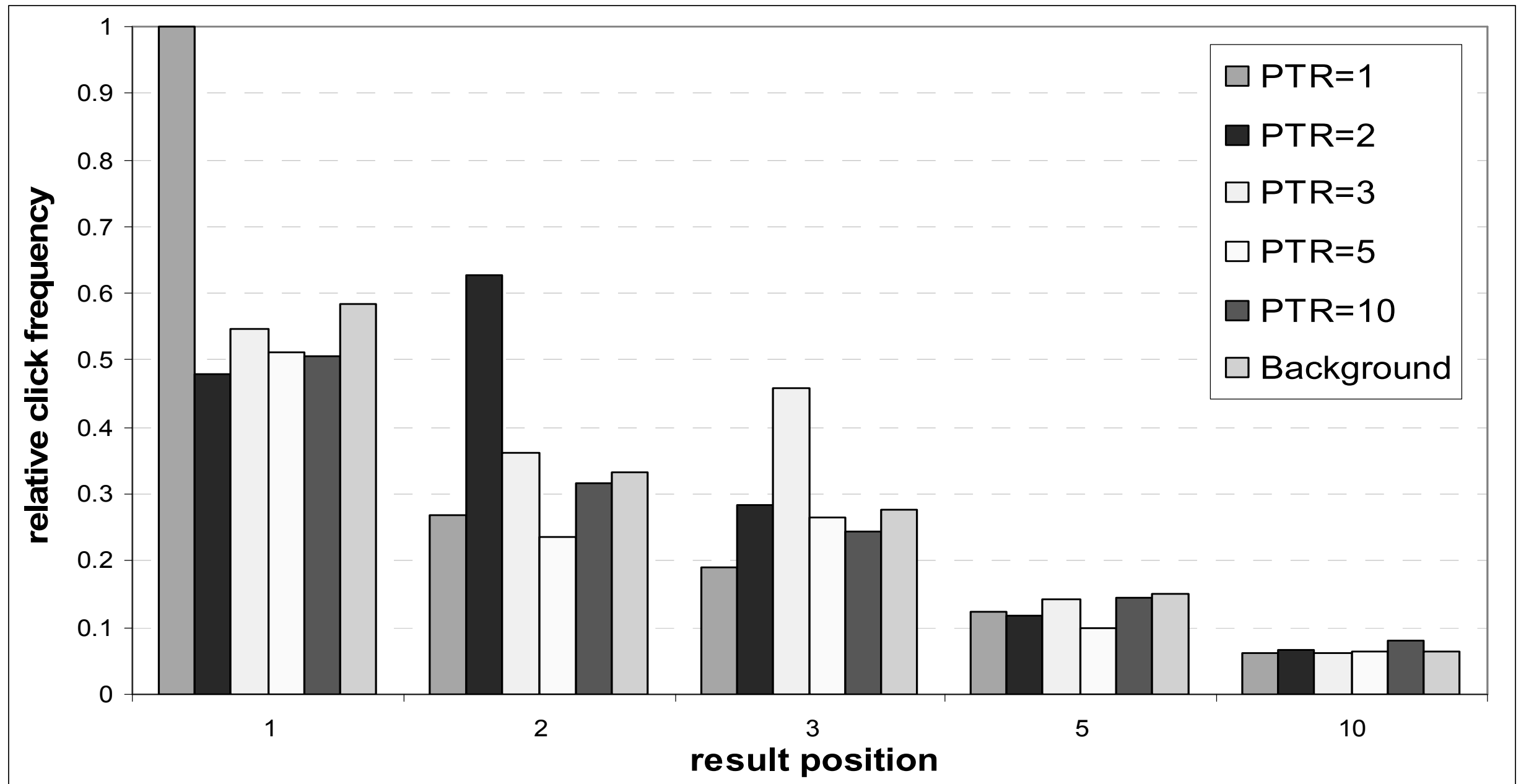
Using Implicit Feedback to Improve Retrieval

- Suppose someone issues the query “bathing my cat”
- The system has seen this query 20 times before and logged all user interactions with the search results
- **Question:** can we use the interactions logged during those 20 searches to improve the ranking for this search?
- **Problem:** implicit feedback signals have a bias!

Implicit Feedback

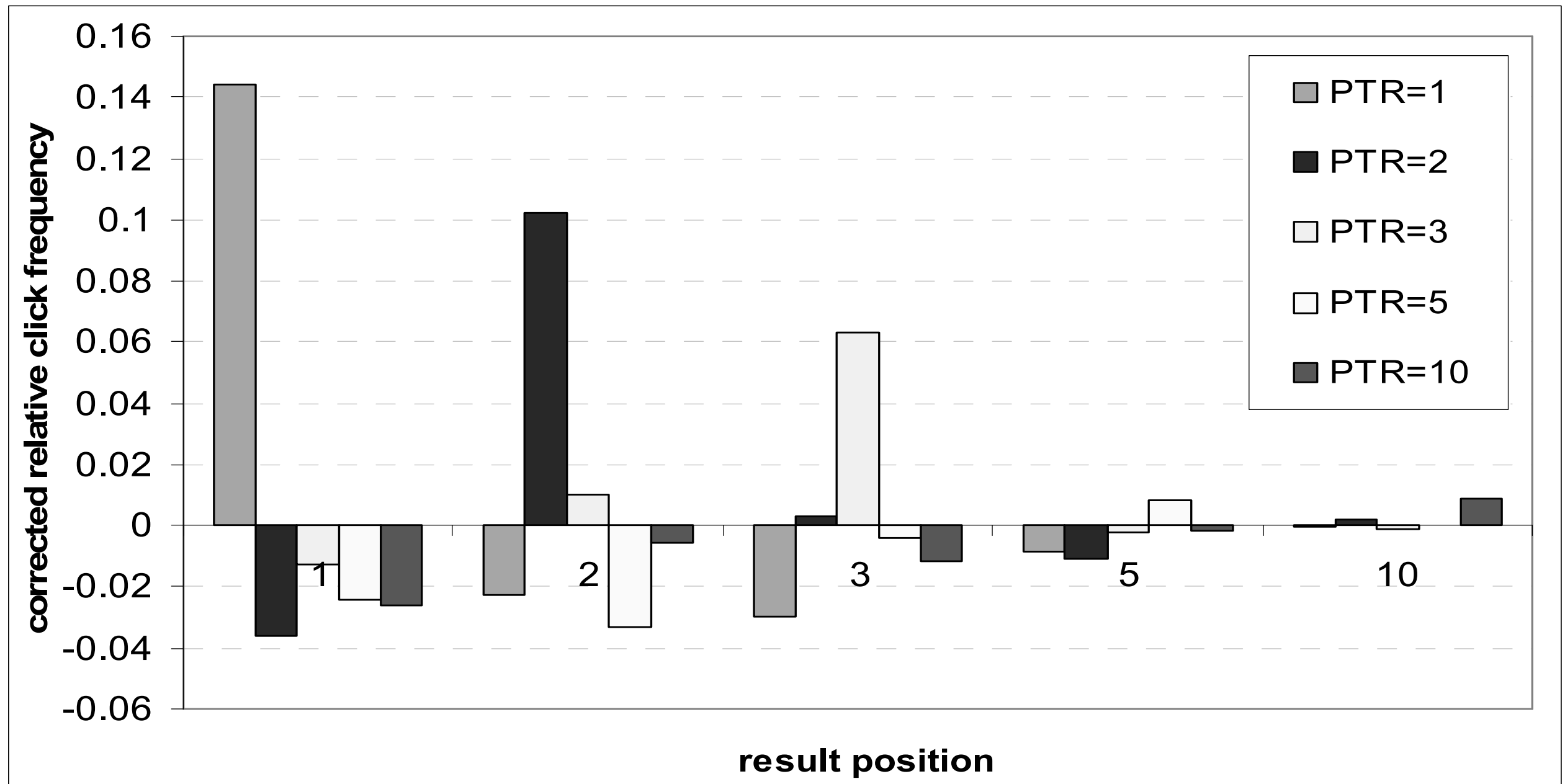
- Previously, we saw that implicit feedback (i.e., clicks) cannot be interpreted as 'absolute' judgements
- They can, however, be interpreted as 'relative' judgements
- There is another solution

Implicit Feedback



- PTR = position of top relevant document

Implicit Feedback



- We can calibrate implicit feedback by “subtracting” the expected amount of noisy feedback (feedback that is irrespective of relevance)

(Agichtein et al., 2006) 62

Calibration Analogy



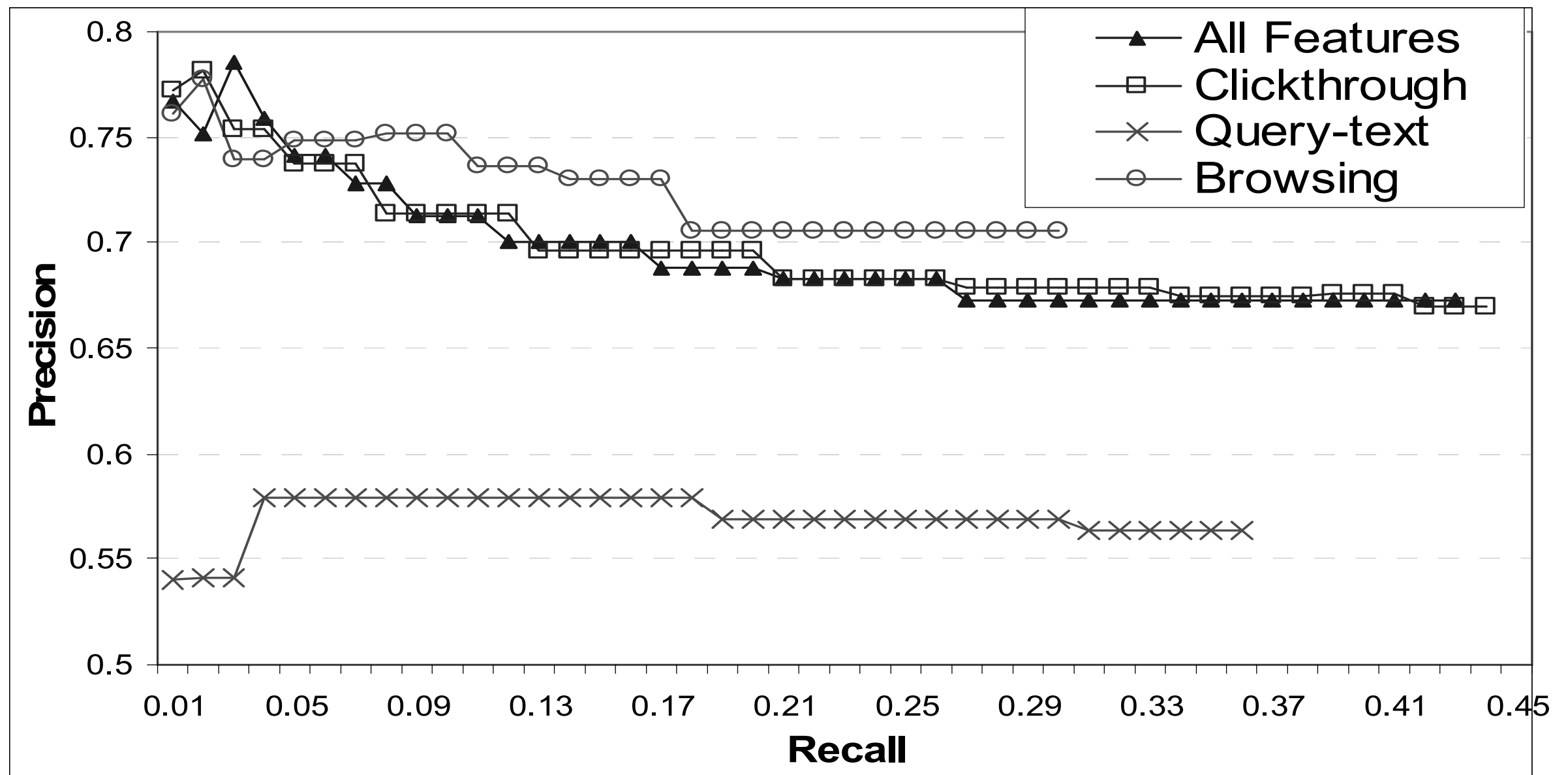
- “Before I weight you, press this button and let me weight the air around here so that I can give you a more accurate prediction!”

Input Features for Learning to Rank

- Query-text features:
 - ▶ Term-overlap between the query and the document title
 - ▶ Term-overlap between the query and the document summary
- Click-based features: calibrated clicks on the page for previous impressions on the query
- Browsing features: calibrated dwell time (time spent on the page) for previous impressions of the query

Do implicit feedback features help?

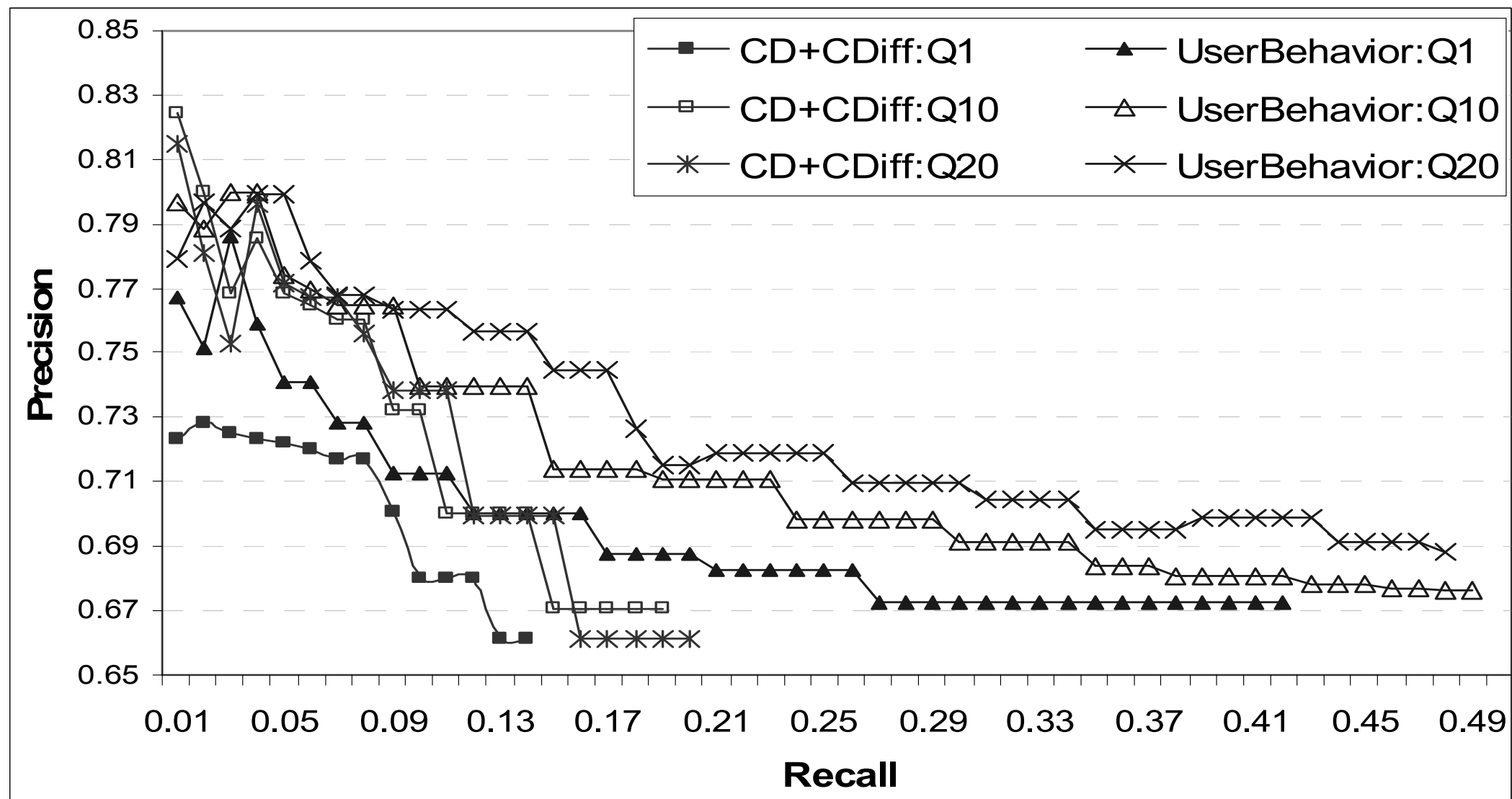
results



- Implicit feedback features alone achieve high precision for low levels of recall
- In combination with query-text features, they achieve high precision for higher level of recall

Do implicit feedback features help?

results



- QX = query seen X times before
- Performance improves as we collect more user-behavior information about the query

Conclusions and Implications

- We spent a lot of time talking about “theoretical” retrieval models (vector space, language model)
- These are still very important
- But, the field is shifting towards machine-learning based solutions
- ML solutions can incorporate various different types of evidence, including implicit feedback evidence
- Document scores produced by “theoretical” retrieval models are one important source of evidence ML-based rankers use to predict relevance