

Filtering Personal Queries From Mixed-use Query Logs

May 8th, 2014

Canadian AI conference

Ary Fagundes Bressane Neto,
Philippe Desaulniers,
Pablo Ariel Duboue,
Alexis Smirnov

Radialpoint SafeCare Inc.

Setting

Technical Support Call centres

Google Queries

- can not open links in windows live mail
- cannot open hyperlinks windows live mail
file cannot be found
- live mail prohibited file type
- live mail prohibited file type how to
allow
- windows live mail cannot open links
- windows live mail cannot open links xp
- windows live mail cannot open pdf files

Our Problem

Mine Google queries
report information needs

Good

“We are missing information on
Shiny New Toy 3000”

Bad

“We are missing information on
How to cheat on drug screening tests”



Web

Images

Maps

Videos

More ▾

Search tools



About 9,330,000 results (0.48 seconds)

Ad related to **problem connecting my LG tv to wifi** ⓘ[LG Smart Televisions - Experience True HD Picture Quality - lg.com](http://www.lg.com/smart-tvs)www.lg.com/smart-tvs ▾

Bring Your Home To Life With LG TVs

LG Electronics Canada has 577 followers on Google+

LG's TV Accessories

LG's Cinema 3D TVs

Stunning LG Home Audio

LG's New Ultra HD TVs



2

[LG Smart TV - Wireless Network Setup \(Built in WI-FI\) - YouTube](https://www.youtube.com/watch?v=vRfv4tl8gfl)www.youtube.com/watch?v=vRfv4tl8gfl ▾

Jan 4, 2013 - Uploaded by ExperienceLG

Enjoy your LG Smart TV to the fullest by **connecting** to **wireless network** with built-in Wi-Fi technology. No ...[Problems with LG smart TV connecting to wifi? - Telstra Crowdsupport](http://crowdsupport.telstra.com.au)crowdsupport.telstra.com.au ▸ [CrowdSupport](https://crowdsupport.com) ▾I have recently bought an LG smart TV - LG 32LS575T Full HD 32" LED TV however whenever I try to **connect** to **my** home broadband it won't. It has found the ...

1

[How to connect your Smart TV to the Internet - LG Blog](http://www.lgblog.co.uk/2012/.../how-to-connect-your-smart-tv-to-the-internet...)www.lgblog.co.uk/2012/.../how-to-connect-your-smart-tv-to-the-internet... ▾Oct 26, 2012 - 1 in 4 had never used their Smart TV to **connect** to the internet. I have a **problem**: the TV is **connected** to internet can't see internet icon on home ... Every time I turn my 3d smart lg tv off it disconnects from **my wifi** network.[47LM6700 - LG Electronics Questions and Answers](http://answers.lg.com/answers/7676/product/.../lg-electronics.../questions.htm)answers.lg.com/answers/7676/product/.../lg-electronics.../questions.htm ▾

★★★★★ Rating: 4.6 - 406 votes

I've been having this **problem** as well, while everything else on the TV functions well it seems ... I tired to **connect my** Ipad 2 to **my TV** using Smart Share but it didn't work. ... Will any High Speed HDMI cable work OK, including for **WiFi** internet?[55LM7600 - LG Electronics Questions and Answers](http://answers.lg.com/answers/7676/product/.../lg-electronics.../questions.htm)answers.lg.com/answers/7676/product/.../lg-electronics.../questions.htm ▾

★★★★★ Rating: 4.7 - 614 votes

wifi password issue How can I get my LM7600 TV **connected** via Wi-Fi? i had**Ernie solved it!**

14 days ago

23

My customer wanted to connect his LG TV to WiFi, and I had to reset the Linksys driver

[How to reset Linksys WRT54G router ? | Linksys wrt54g firmware](http://linksys-wrt54g-firmware.blogspot.com/.../reset-links...)linksys-wrt54g-firmware.blogspot.com/.../reset-links...

17 nov. 2012 - This article contains suggestions for resetting your router so that you can get your Internet connection up and running again, I'll outline...

**Gina solved it!**

11 days ago

11

Follow this thread instructions to reset the firmware.

[WRT54G password reset? - Linksys Community](http://community.linksys.com/...Routers/WRT54G...reset/...)community.linksys.com/...Routers/WRT54G...reset/...

We have a cable modem and I have a Mac G4 and my wife has a Dell laptop. We've been using the WRT54G for about six months now without problems...

How to solve it?

Our solution:

- Hand labelled data
- Bag-of-words
- Text classification

Other techniques:

- Time-based
- Multiple browsers
- Privacy Preserving Data Mining
 - Mixing queries from multiple people
 - Adding fake terms to the queries

Our Solution

Our solution:

- Hand labelled data
- Bag-of-words
 - Keeping capitalization
 - Ignoring numbers

Classification:

- TF/IDF documents are too small (-)
- Singular Value Decomposition not enough data (-)
- Data normalization - remove extra spaces, lowercase, remove stop-words, remove short/long queries (+/-)
- **Maximum Entropy Classification (+)**

Queries

Collection

- 4 months
- 145 participants

Type	Query string
Personal	Amoxicillin alcohol
Business	audio device on high definition audio bus windows xp driver dell
Business	Avaya Voip
Personal	Best poutine store in Montreal

Annotation

Would this query be issued by a technical support agent in the process of answering a customer call? If unsure, annotate it as “personal.”

- Annotated by one author, familiar with the problem
- A sample of 100 queries annotated by a second author
 - Not involved in the original annotation process
- Cohen's kappa: 0.8797

Annotation results

Class	Queries		Unique Queries		Queries		Unique Queries	
	#	%	#	%	#	%	#	%
personal	3336	37.82	1622	31.10	1378	35.40	267	18.80
business	5483	62.18	3597	68.90	2517	64.60	1150	81.20
Total	8819		5219		3895		1417	

Data for learning and eval

Training

- 100 tech support agents
- 3895 unique queries
- 36% personal

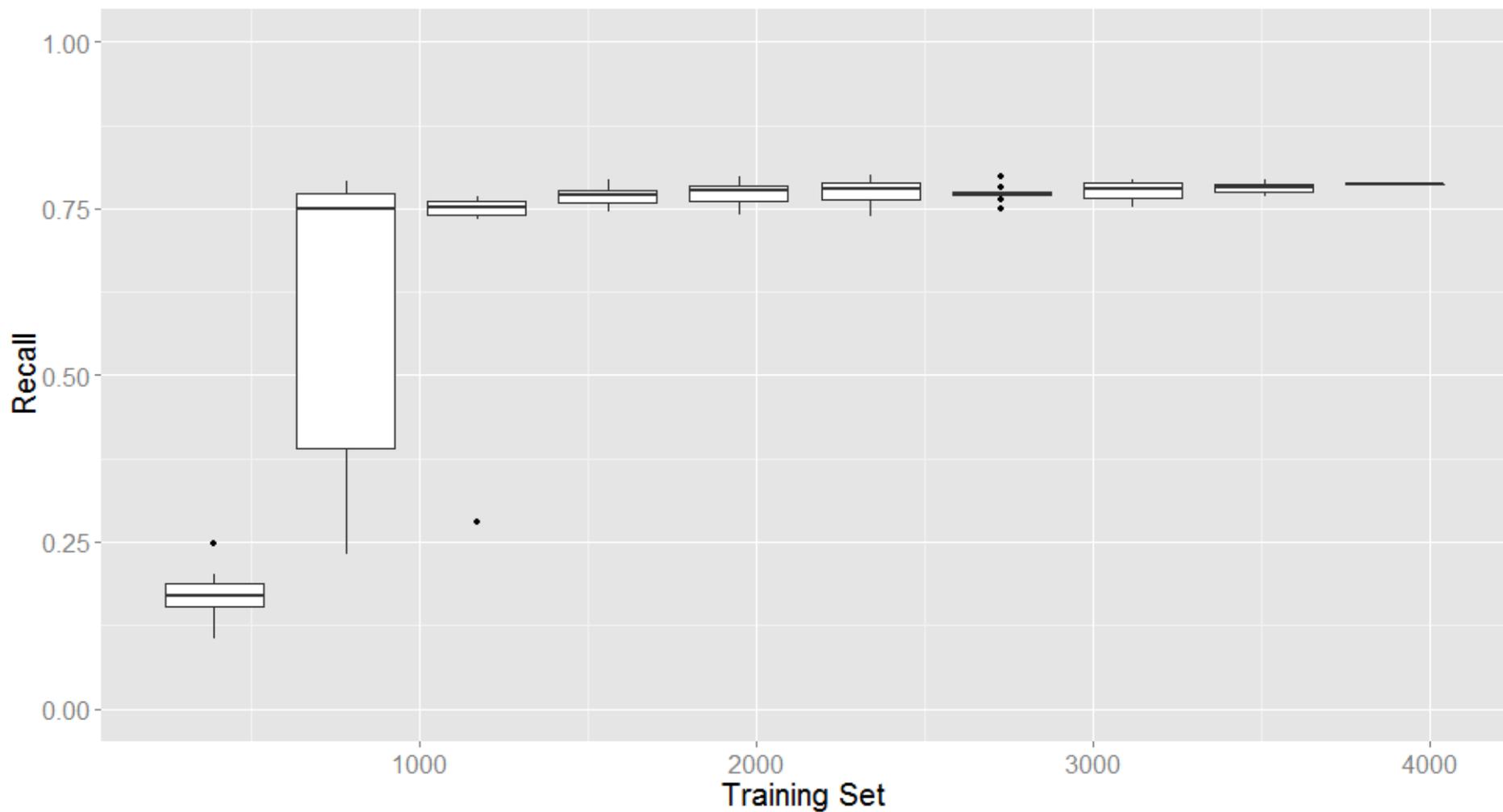
Testing

- 45 tech support agents
- 1417 unique queries
- 18% of them personal

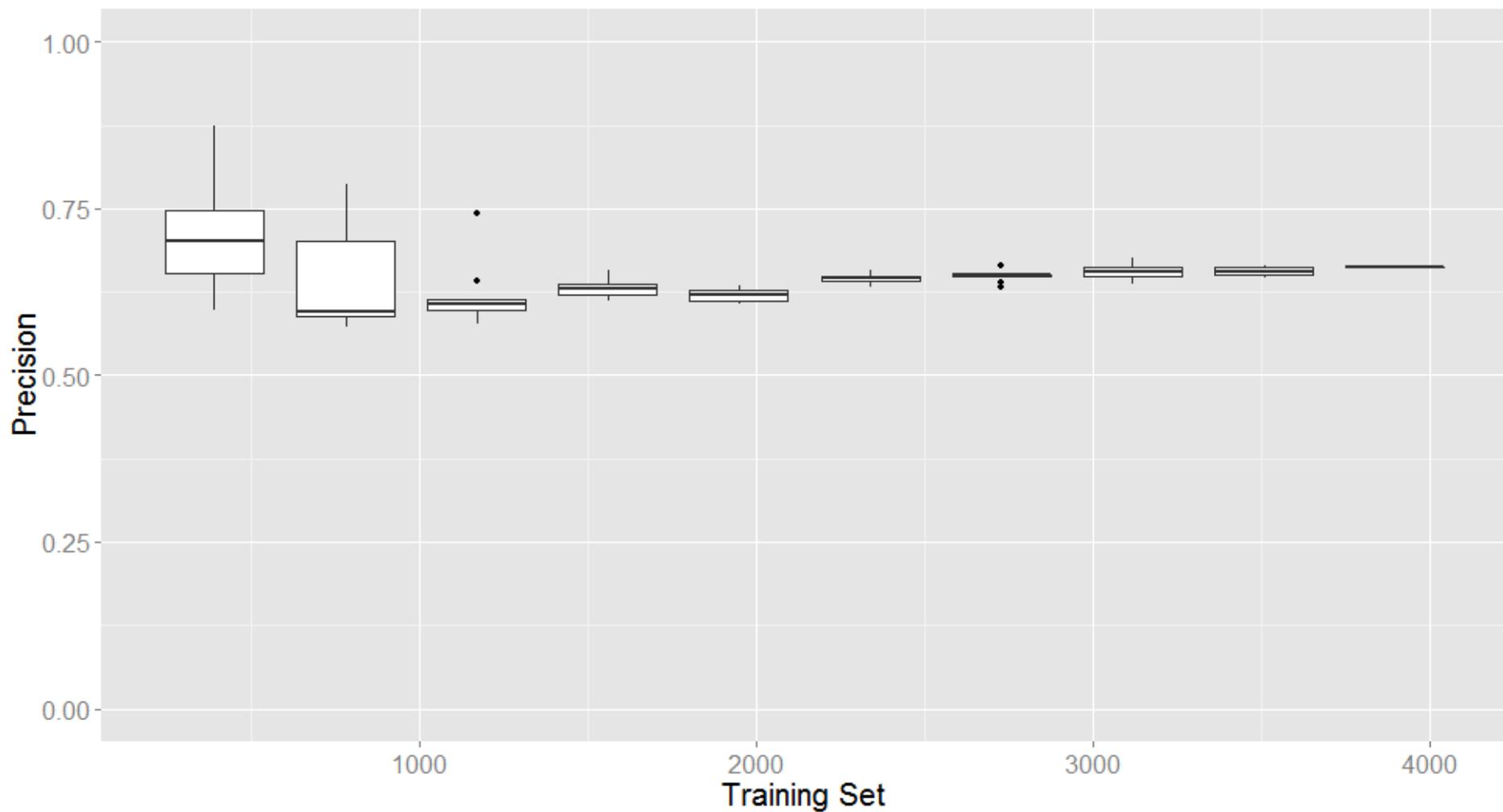
Results

- Correctly identified 210 out of 267 personal queries
 - Recall 78.7%
- Total of 317 queries marked as personal:
 - Precision 66.2%
- We filter out 78.7% of private queries losing only 9.3% of the business queries.

Learning Curves (Rec)



Learning Curves (Prec)



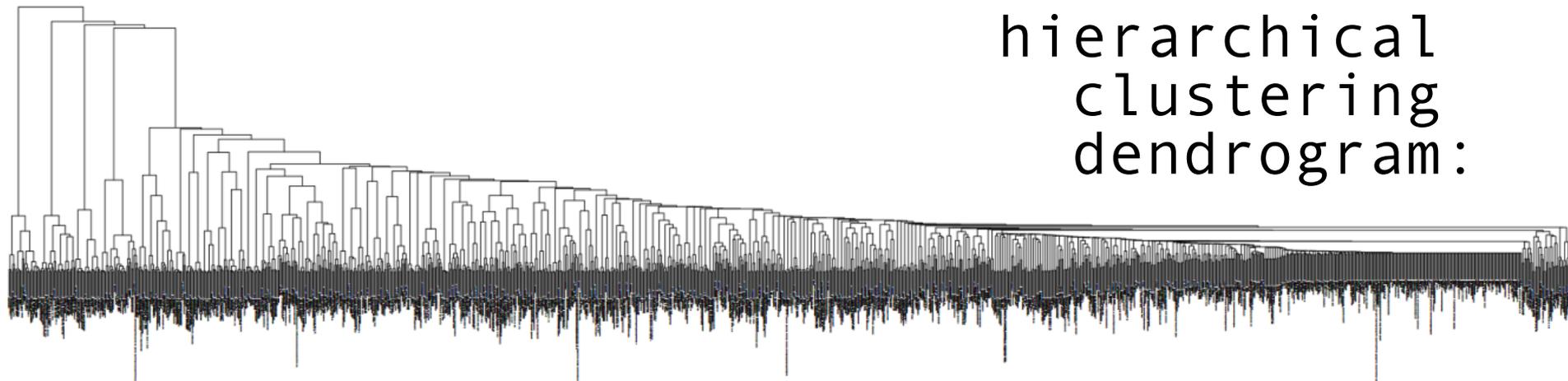
Qualitative error analysis

Type	Erroneously Predicted as	Count	%
Queries with stop-words <i>"how much fiber is in oatmeal"</i>	Business	31	18.9
Personal queries with business terms <i>"best games for gamecube"</i>	Business	20	12.1
Queries with 1 term <i>"igfxtray"</i>	Personal	54	32.9
Only numbers <i>"7035 7036"</i>	Personal	7	4.3
Error in gold standard <i>"hines ward"</i>	Personal	6	3.6
Other	Personal	46	28

Ongoing Work: Clustering

- Clicked URLs
- Top n Google results
- Levenshtein distance on query terms
 - Are “XP install” and “HP install” close?
- Cosine distance on query terms
- Latent Semantic Analysis

hierarchical
clustering
dendrogram:



Working together

Internship program

- www.radialpoint.com/about-radialpoint/career

Contact

- Alexis Smirnov (tomorrow!)
Alexis.Smirnov@radialpoint.com

Of note

- The query data is available
- word2vec lucene query expansion
<https://github.com/radialpoint/word2vec-query-expansion>