Query Auto-Completion for Rare Prefixes

Bhaskar Mitra
Microsoft
bmitra@microsoft.com

Nick Craswell
Microsoft
nickcr@microsoft.com

Step 1

Mine popular suffixes from historical query logs.

Top suffixes	Top 2-word suffixes	Top 3-word suffixes
com	for sale	federal credit union
org	yahoo com	new york city
net	myspace com	in new york
gov	google com	or no deal
pictures	new york	disney channel com
lyrics	real estate	my space com
edu	of america	in new jersey
sale	high school	homes for sale
games	new jersey	department of corrections
florida	space com	chamber of commerce
for sale	aol com	bath and beyond
us	s com	in las vegas

Step 2

Extract the end-term from an input prefix.

cheapest flight fro	End-term: "fro"
cheapest flight from	End-term: "from"
cheapest flight from	End-term: "from "
cheapest flight from n	End-term: "n"

Step 3

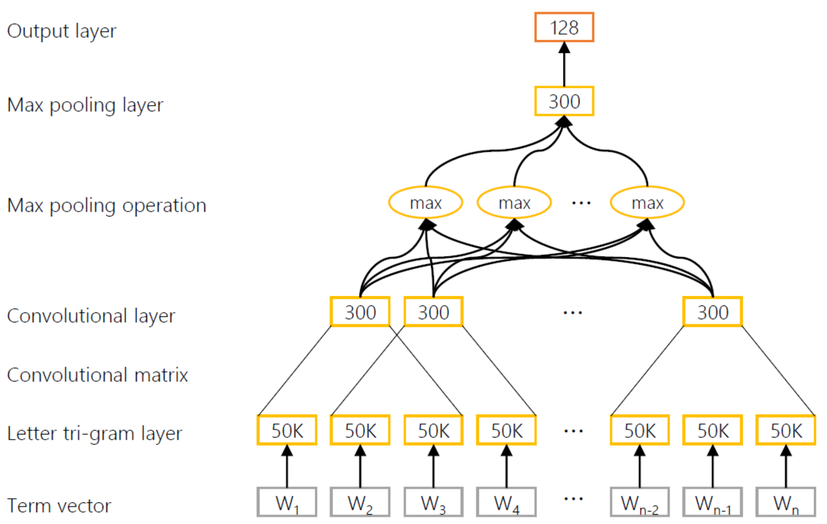
Identify suffixes that begin with the end-term and append them to the prefix to generate synthetic suggestions. Then rank the synthetic suggestions alongside the full-query candidates using a lambdaMART model with *n*-gram and CLSM based features.

cheapest flights from seattle to
cheapest flights from seattle *to dc*
cheapest flights from seattle *to washington dc*
cheapest flights from seattle *to bermuda*
cheapest flights from seattle *to bahamas*
cheapest flights from seattle *to aruba*
cheapest flights from seattle *to punta cana*

Convolutional Latent Semantic Model

We train a novel CLSM based language model on a query prefix-suffix pairs corpus. The training pairs are sampled from the search query logs by splitting each query at every possible word boundary.

The trained CLSM model is then used for generating features for the LambdaMART based ranking model.



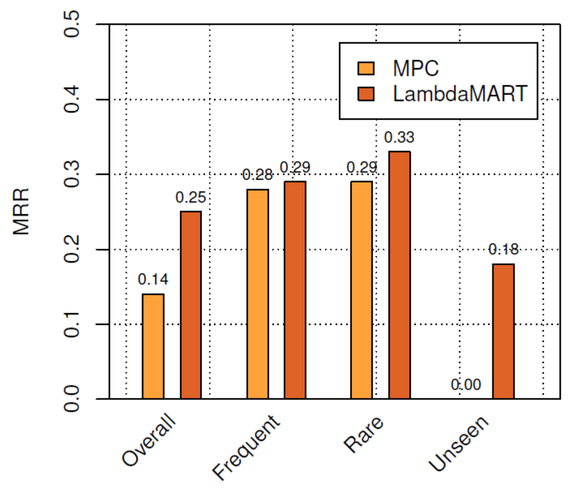
Summary

Generating **synthetic candidates** for query auto-completion by appending popular query suffixes to the input prefix and ranking them alongside traditional full-query candidates using *n*-gram statistics and a novel **deep neural network** model. Unlike next word prediction, this approach always suggests **complete queries**.

Results

Overall, the LambdaMART model with both the *n*-gram and the CLSM features achieves the best MRR. However, compared to the CLSM feature the *n*-gram features show higher improvements.

Most of the improvements in MRR are from the rare and the unseen prefixes as expected.



Models	AOL MRR	AOL % Improv.	Bing % Improv.
Full-query based candidates + Suffix based candidates (Top 100K suffixes)			
MostPopularCompletion	0.1446	-	-
LambdaMART Model (*n*-gram features = no, CLSM feature = no)	0.2105	+45.5*	+39.9*
LambdaMART Model (*n*-gram features = yes, CLSM feature = no)	0.2441	**+68.7***	**+54.2***
LambdaMART Model (*n*-gram features = no, CLSM feature = yes)	0.2248	+55.4*	**+48.9***
LambdaMART Model (*n*-gram features = yes, CLSM feature = yes)	0.2453	**+69.6***	**+55.3***

CIKM2015
19–23 October 2015 Melbourne Australia

Microsoft