Cuisine In and Out:
A Distributed Search Engine
with Attention Neural MT and Doc2Vec

Prof. Matt Doherty
Guangyu Zhang, Rihan Yang
Overview

- System architecture
- Advanced NLP theories behind
- Search engine demo
Interface

FoodSearch

Search for: burger

Rincon Peruano Peruvian Restaurant

... is not acceptable while out on the town. But at the Improv, I figured that my indiscretion could slip by. The daiquiri was great. I recommend everyone order daquiris. I was also pleased with my burger and fries (both were tasty). I have a new rule: whenever I am presented with a menu which includes both entrees which seem slightly pretentious and [pub grub,] with the "pub grub" listed at...

Dieses Unternehmen ist nicht für ihre Arbeit, die ich für einen Preis für die am besten und bei mir erg@ riff@ enen Preis @@ senk@@ ung aus@@ wirkt , und als ich es an der Zeit , die die Kosten für die Umwelt zu ziehen war , war , wie ich es mir , die am wenigsten war , als ich in der Lage war , die am Ende der letzten Zeit nicht zu tun haben , die die meisten gu@@ thei@@ Ben , die die Verbrauch

Huachinango Mexican Grill

... here Friday night, so it seems to be becoming a popular spot. The service was excellent on all counts--a nice surprise since it's a new restaurant. As for the food, it was really good. Great burgers and really good salads--huge, fresh, and super inexpensive for how good they are. Their specials sounded great, too--and someone in our group ordered one, which he said was amazing (meatloaf...

Die Lebensmittel ist immer wirklich gut und wirklich sehr gut , und ich habe mich wirklich sehr nahe , und meine co . die wir hier zumindest zweimal in der
Interface

FoodSearch

cheeseburger  hamburger  turkey_burger  bacon_cheeseburger

Rincon Peruano Peruvian Restaurant
... is not acceptable while out on the town. But at the Improv, I figured that the Indiscretion could slip by. The daiquiri was great. I recommend everyone...

Huauchinango Mexican Grill
... here Friday night, so it seems to be becoming a popular spot. The service was excellent on all counts—a nice surprise since it’s a new restaurant. As for the food, it was really good. Great burgers and really good salads—huge, fresh, and super inexpensive for how good they are. Their specials sounded great, too—and someone in our group ordered one, which he said was amazing (meatloaf...
Neural Machine Translation [Wu et al 2016]
RNN Encoder-Decoder with Attention [Britz et al. 2017]
tf-seq2seq
RNN Encoder-Decoder
Attention mechanism
128-unit GRU cell
Adam optimizer
Bidirectional
doc2vec, word2vec
gensim

Data Challenge R9
>4M reviews
>140K business
WMT’16 en-de
~GB paired bilingual sentences
~32,000 subword units
CIMS GPU servers
>770K steps, 2 weeks
PCA embeddings from 128-D (37007 points)

770K training steps vs. Google: 250,000 GPU hours
Future Works

• Model improvements

Google
BLEU: 24.61
days on GPUs, distributed training
BiRNN with attention
8-layer
512-unit LSTM
TensorFlow, Python
Wu et al. 2016

Facebook
BLEU: 25.16
One night on a single GPU
CNN with attention
6-layer
Hidden size 256
Torch, Lua
Gehring et al. 2017 May 9 (yesterday!)
Future Works

• Scalability for large corpus

• Seq2seq as a general purpose model
  • Image caption, news summarization…
  • Direct business utilizations
Search Engine Demo

http://linserv2.cims.nyu.edu:14789
References


References


