Neural IR

Challenges:

- Vocabulary mismatch

Q: How many people live in Sydney?

- *Sydney's population is 4.9 million*
  [relevant, but missing 'people' and 'live']
- *Hundreds of people queuing for live music in Sydney*
  [irrelevant, and matching 'people' and 'live']

- Need to interpret words based on context (e.g., temporal)
  Q: uk prime minister


Today


Recent


1990s

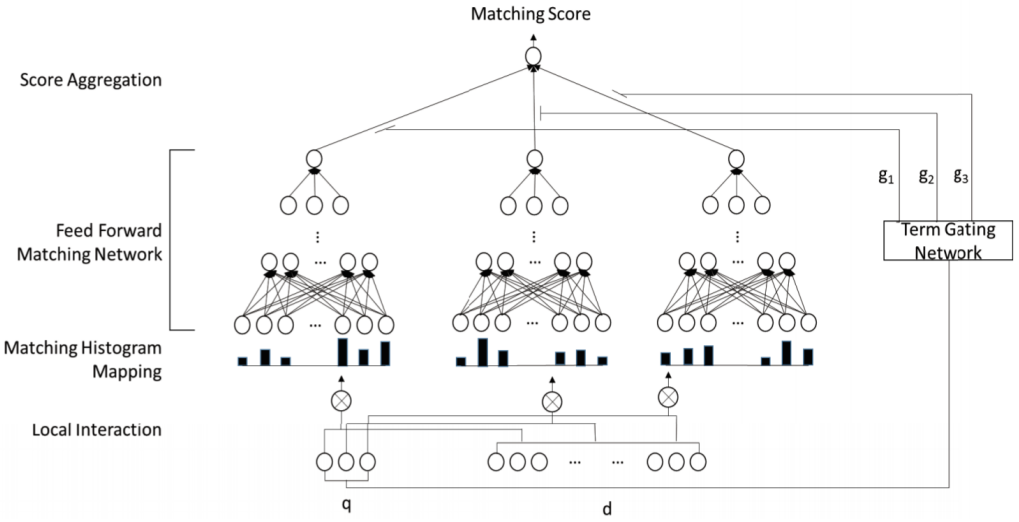
Techniques:

- Word embeddings for IR: expand query using embeddings, IR models that work in the embedding space...

global	local
cutting	tax
squeeze	deficit
reduce	vote
slash	budget
reduction	reduction
spend	house
lower	bill
halve	plan
soften	spend
freeze	billion

Terms similar to "cut" for a word2vec model trained on:
- global: a general news corpus
- local: documents related to "gasoline tax"

- Deep neural networks for IR:
  to generate query representation,
  to generate document representation,
  estiamte relevance



Profile: Computer Scientist
Team: Arantxa, Eneko