Controlling Information Aggregation for Complex Question Answering

Heeyoung Kwon, Harsh Trivedi, Peter Jansen, Mihai Surdeanu, and Niranjan Balasubramanian
Stony Brook University, The University of Arizona

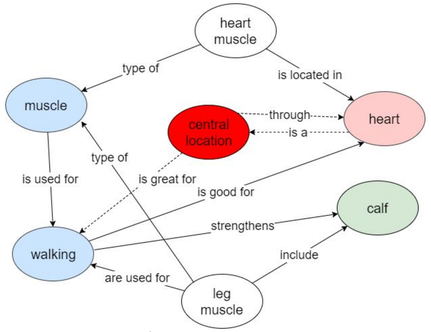
Motivation

We model complex question answering as a task of aggregating related facts in a knowledge graph.

Problem: Aggregating facts is prone to ***"inference drift"***, where long chains of facts quickly drift off topic.

Solution: We develop a drift-sensitive PageRank random walk algorithm for QA graph traversal.

Q. Which muscle is used for walking?
(A) heart (B) calf (C) …

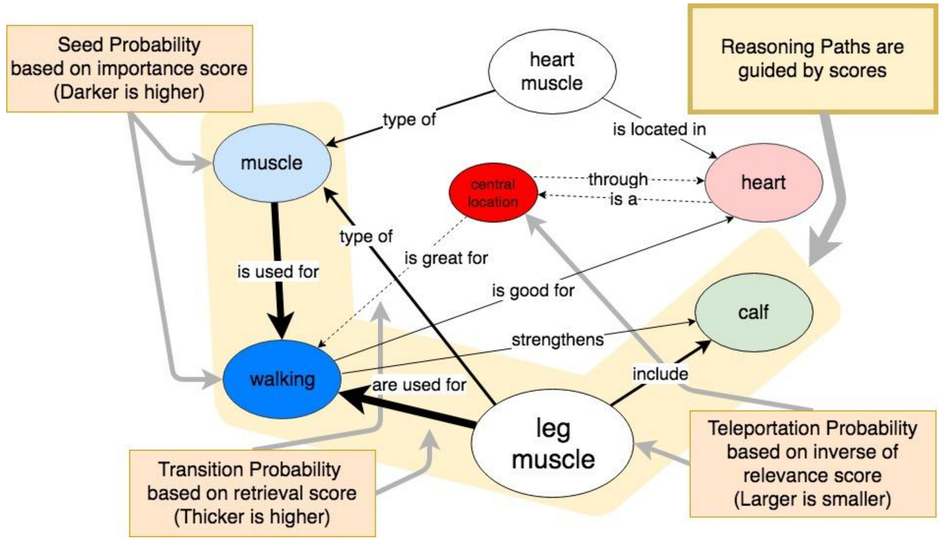


(An example 8th grade science exam question)

Blue	Question Terms
Green	Correct answer choice
Pink	Incorrect answer choice
White	Other concepts from OpenIE (Open Information Extraction)
Red	The node introducing inference drift

Drift-Sensitive PageRank

- Unsupervised Estimation



- Supervised PageRank: Parameterization

$$\pi^{(t+1)} = (1-d)A_\phi \pi^t + dv_\theta$$

- Transition Probability from edge features

$$v_\theta(s) = \frac{f_\theta(x_s)}{\sum_{u \in G} f_\theta(x_u)}$$

- Seed Probability from node features

$$A_\phi(s,t) = \frac{g_\phi(z_{st})}{\sum_{e_{sq} \in G} g_\phi(z_{sq})}$$

Evaluation

- Drift-sensitive PageRank

method	seeds	teleportation	test	reference
page rank	none	uniform	35.51	
TPR	uniform	uniform	38.26	
drift-sensitive	focus	uniform	40.33	(A)
	focus	quest. sim.	41.49	(B)
	sup.	sup.	42.34	Sup.

- Different Graph Sizes

method	top 10	top 20	top 30	top 40	top 50
TPR	39.54	40.63	41.31	38.26	38.68
unsupervised	41.00	41.55	42.46	40.33	39.84
supervised	41.30	42.22	41.80	42.34	42.40

- Utility of Aggregation

method	sent	sent + (A)	sent + (B)	sent + Sup.
accuracy	43.44	44.30	45.58	45.45

Summary

- Drift-sensitive variants of PageRank allow for effective reasoning over large graphs by controlling the random walks
- Drift-sensitive methods achieve substantial gains over standard topic-sensitive PageRank

LUNR language understanding & reasoning