Biased-reinforced search on graphs

"On the robustness and optimization of searches"

No free lunch theorem

The no free lunch theorem for search and optimization [2] applies to finite spaces and algorithms that do not resample points. All algorithms that search for an extremum of a cost function perform exactly the same when averaged over all possible cost functions. So, for any search/optimization algorithm, any elevated performance over one class of problems is exactly paid for in performance over another class. *As a NFL class of functions does not change by any permutation on the input space, there is no structure which can be used for search. Hence, all search strategies show the same behavior [1].*

Reinforced search processes

- optimize for "common case"
- always gives results you think you are looking for
- "loss of random walk exploration capabilities"
  - Seach in P2P networks suffer from this. If we find a clever way to find information faster we may loose our ability to explore it somewhere else.
  - But it is also important to find things you were not looking for at the moment.
  - Think of how kids do class research projects. Say, kids have to research about volcanos. In the past one had to talk to uncles and aunts and go through magazines of science to look for information on volcanos (ending up reading about other things). Today, kids can have the same information if they use the same search engine (e.g. all kids will write about the wikipedia entry on volcanos).
  - Loss of serendipity. From wikipedia: *One aspect of Walpole's original definition of serendipity that is often missed in modern discussions of the word is the "sagacity" of being able to link together apparently innocuous facts to come to a valuable conclusion. Thus, while some scientists and inventors are reluctant about reporting accidental discoveries, others openly admit its role; in fact serendipity is a major component of scientific discoveries and inventions. According to M. K. Stoskopf[4] "it should be recognized that serendipitous discoveries are of significant value in the advancement of science and often present the foundation for important intellectual leaps of understanding".*

Examples

Search Engines

Google's reinforcement search algorithm.

What if Google was the only way to find knowledge? "Forget the Web graph".

Academic Knowledge

(DBLP citations over time, see if the process evolves like our model). Compare creativity in the 1800s to creativity now.

P2P file sharing search

Methods to optimize finding information may actually hurt if information location is too dynamic. Random walks work better when information could be anywhere and there is not prior to where it is.

New search engine

- why return a flat list?
- new rules: search results should be directed graphs
- vertices are information (webpages), links capture what one thinks is important to maximize search and serendipity
- Links should be probabilistic between pages and inversely proportional to their distance (a la Kleinberg).
- Power laws are fundamental: links to far away places should exist with non-negligible probability
- distance can be distance on flat list rank, distance in terms of diffusion (random walk distance), distance in terms of content similarity, etc.
- goal is for user to do random walk on a constructed graph.
- constructed graph is better (in some sense) than original hyperlinked graph
- graph structure depends on how confident the user is about the search words:
  - robustness
  - optimization (knob)

References

[1] Stefan Droste, Thomas Jansen, and Ingo Wegener. Optimization with randomized search heuristics: the (a)nfl theorem, realistic scenarios, and difficult functions. *Theor. Comput. Sci.*, 287:131–144, September 2002.

[2] David Wolpert and William G. Macready. No free lunch theorems for optimization. *IEEE Trans. Evolutionary Computation*, 1(1):67–82, 1997.