Networks for machine learning

Esteban Bautista, Patrice Abry, Paulo Gonçalves

ENS de Lyon, INRIA, CNRS
esteban.bautista-ruiz@ens-lyon.fr

October 10, 2019
The big data era

- Modern systems generate massive amounts of data
  - Sensor systems
  - Internet of things
  - Digital documents

- Technological progress permits to store and effortlessly access it

- Valuable source of information to better solve real world problems

- To get insights from all these data, we need to first organize it
  - Separate emails that are spam from those that are not
  - Organize documents by topic
  - Identify bank transactions that are fraud

- Numerous classification techniques have been proposed over the years

- Classifiers need to learn from annotated data.

- Issue: annotated data do not follow the big data trend
The scarcity of labelled data

Raw data

Linear classifier

Labelled examples are insufficient to learn something about the data
Q: How to learn from limited amounts of labelled data?

A: Use a similarity graph to learn from both labelled and unlabelled data
Graph node classification

Graph from data

Label propagation

Graphs are powerful objects to represent data
Q: How to propagate the labelled data in the graph?

A: Google’s PageRank!
PageRank for data classification

The PageRank assigns a score function to graph vertices according to

\[
\arg \min_{f} \left\{ \sum_{u,v \in G} W_{uv} \left( \frac{f_u}{D_{uu}} - \frac{f_v}{D_{vv}} \right)^2 + \mu \sum_{u \in G} \frac{1}{D_{uu}} (f_u - y_u)^2 \right\}
\]

- \(y\): indicator function of annotated nodes (1 if annotated, 0 otherwise)
- \(\mu\): regularization parameter
- \(f\): PageRank vector
- Smoothness: similar vertices should have similar values in \(f\)
- Fitting: \(f\) should be consistent with labelled data

The solution of this problem coincides with the equilibrium state of a random walk process
PageRank as a diffusion process

The solution of the PageRank problem is a random walk process:

$$f^T = \sum_{k=0}^{\infty} (1 - \alpha) \alpha^k y^T P^k,$$

where $$\alpha = 1/(1 + \mu)$$

- $$k = 0$$: walker at a label with probability one
- $$k + 1$$: walkers decides to jump to a neighbor with probability $$\alpha$$, or to restart to the labels with probability $$(1 - \alpha)$$
- $$f_u \propto$$ probability of finding a walker, at equilibrium, at node $$u$$
PageRank vectors $u \in S_k : \arg\max_k F_{uk}$
Reducing the number of labelled points

In practice, unfeasible to collect labelled data for hundreds or thousands of classes. Can we identify classes individually?
The sweep-cut algorithm

Sweep-cut

A sweep-cut is a procedure to identify individual classes from a PageRank vector.

- Let $v_1, \ldots, v_N$ be a rearrangement of the vertices in descending order, so that the permutation vector $q$ satisfies $q_{v_i} = f_{v_i}/d_{v_i} \geq q_{v_{i+1}} = f_{v_{i+1}}/d_{v_{i+1}}$.
- Let $S_j = \{v_1, \ldots, v_j\}$ be the set of vertices indexed by the first $j$ elements of $q$.
- Let $\tau(f) = \min_j h_{S_j}$, where $h_{S_j}$ is the ratio of external and internal links of $S_j$.
- Retrieve $\hat{S}_{gt} = S_j$ for the set $S_j$ achieving $\tau(f)$.
Drawbacks?

- Unbalanced data settings lead to classification biases
- RW properties (e.g. Mixing rate, mean passage times, etc) highly sensitive to non-trivial network structure
- Reliable classifications with few labelled data only under simple data settings
- Curse of dimensionality issue causes flat functions
- Avoid sweep-cuts to retrieve sub-classes conforming one bigger class
Sensitivity to non-trivial network structures

Balanced setting: $w_l = w_r = 1$
Sensitivity to non-trivial network structures

Skewed setting: $w_l = 10; w_r = 1$
How to solve this issue?

Tweaking random walk dynamics: Lévy flights?
Lévy flights induced by $L^\gamma$ for $0 < \gamma < 1$ (Riascos, Phys. Rev. E 90, 2014)

$$L^\gamma = (D - W)^\gamma = Q \Lambda^\gamma Q^T = D^\gamma - W^\gamma$$

$$(P^\gamma)_{uv} = (D^{-1} W^\gamma)_{uv} \propto \Delta_{uv} - \gamma.$$
Proposition: extending PageRank to Lévy flights

\[
\arg \min_{f} \left\{ \sum_{u,v \in G} (W_{\gamma})_{uv} \left( \frac{f_u}{(D_{\gamma})_{uu}} - \frac{f_v}{(D_{\gamma})_{vv}} \right)^2 + \mu \sum_{u \in G} \frac{1}{(D_{\gamma})_{uu}} (f_u - y_u)^2 \right\}
\]

Solution: equilibrium state of Lévy flight process

\[
f^T = \sum_{k=0}^{\infty} (1 - \alpha) \alpha^k y^T P_{\gamma}^k,
\]

where \( \alpha = 1/(1 + \mu) \)
Revisiting the skewed graph via Lévy flights

Skewed setting: $w_l = 10; w_r = 1$ and $\gamma = 0.01$
Recap

- Graphs are useful objects to represent data.
- Graphs allow to learn from labelled and unlabelled data to improve classifiers.
- Random walkers are a simple yet effective approach to propagate information in the graph (PageRank algorithm).
- Random walkers can be sensitive to trapping regions in the graph.
- Anomalous diffusion processes, like Lévy flights, may carry a better alternative in certain applications.