Aalto University
School of Science

# Algorithmic methods for mining large graphs
# Lecure 4 : Maximization of submodular functions


Aristides Gionis

Aalto University


Bertinoro International Spring School 2016
March 7–11, 2016

# submodular set functions

- consider a ground set $U$
- a function $f : 2^U \to \mathbb{R}$ is submodular if

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$$

  for all $A, B \subseteq U$

- equivalently ("diminishing returns")

$$f(A \cup \{x\}) - f(A) \geq f(B \cup \{x\}) - f(B)$$

  for all $A \subseteq B \subseteq U$ and $x \in U \setminus B$

# submodular set functions

may or not satisfy the following properties

- non-negative : $f(A) \geq 0$ for all $A \subseteq U$
- monotone : $f(A) \leq f(B)$ for all $A \subseteq B \subseteq U$
- symmetric : $f(A) = f(U \setminus A)$ for all $A \subseteq U$

# examples

- coverage in set systems

  $\Rightarrow$ monotone and non-negative

- cut functions in undirected graphs and hypergraphs

  $\Rightarrow$ symmetric and non-negative

- cut functions in directed graphs

  $\Rightarrow$ non-negative

# example: coverage in set systems

- $S_1, \ldots, S_n$ subsets of $U$
- function $f : 2^{\{1,\ldots,n\}} \to \mathbb{R}_+$
- coverage :

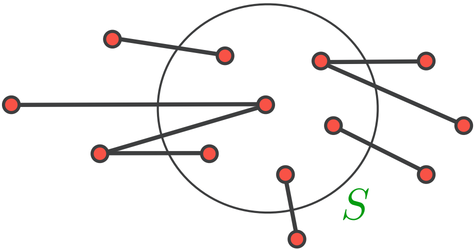
$$f(A) = | \cup_{i \in A} S_i |$$

- weighted coverage :

$$w : U \to \mathbb{R}_+ \text{ and } f(A) = \sum_{x \in \cup_{i \in A} S_i} w(x)$$

# example: cut in graphs

- consider undirected graph $G = (V, E)$
- cut function $f : 2^V \to R_+$ defined as $f(S) = |E(S, V \setminus S)|$



# the maximization problem

- given submodular function $f : 2^U \to \mathbb{R}$

  find $S \subseteq X$ to maximize $f(S)$

  subject to constraints

- value-oracle model
- generalizes many interesting problems **NP**-hard problems
- minimization problem is polynomial (e.g., min-cut)

# monotone functions

- $f(U)$ trivial maximizer

- more interesting to maximize under cardinality constraints
- find $S \subseteq U$ subject to $|S| \leq k$ that maximizes $f(S)$

- MAX $k$-COVER is a special case
- greedy gives $(1 - 1/e)$ approximation
  [Nemhauser et al., 1978]
- no better approximation unless **P**=**NP**

# the greedy algorithm

1. $S \leftarrow \emptyset$
2. while $|S| < k$
3. $\quad i \leftarrow \arg\max_j f(S \cup \{j\})$
4. $\quad S \leftarrow S \cup \{i\}$
5. return $S$

# analysis of the greedy

$S_j$ : first $j$ elements picked by the greedy

$f(S) = \delta_1 + \ldots + \delta_k$

$\delta_j \geq (f(S^*) - f(S_{j-1}))/k$ (monotonicity and submodularity)

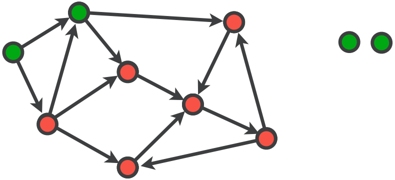
$f(S^*) - f(S_j) \leq (1 - 1/k)^j f(S^*)$ (by induction)

$f(S^*) - f(S) \leq (1 - 1/k)^k f(S^*) \leq (1 - \frac{1}{e}) f(S^*)$

$f(S) \geq (1 - (1 - 1/k)^k) f(S^*)$

# widely applicable in data mining

- example : maximize the spread of influence in social networks [Kempe et al., 2003]
- assume that an action is spread in a social network
- assume a spreading model such as independent cascade
- find a set of *k* initial seeds to maximize the spread
- spreading model is randomized, so we want to maximize expected spread



# non-monotone functions

- unconstrainted version becomes interesting
- find $S \subseteq X$ to maximize $f(S)$

- generalizes MAX-CUT
- what do we know about approximation?
- random set gives $1/2$ ($1/4$ for MAX-DICUT)
- SDP gives 0.878 (0.796 for MAX-DICUT)
  major breakthrough [Goemans and Williamson, 1995]
- 0.53 by spectral approach [Trevisan, 2012]

# unconstrained problem

[Feige et al., 2011]

- first constant-factor approximations for non-negative submodular functions
- simple algorithms: randomized / deterministic, non-adaptive / adaptive
- $1/2$ approx for symmetric functions
- $2/5 = 0.4$ approx for the non-negative functions
- lower bound: better than $1/2$ approx requires exponential number of value queries

# unconstrainted problem

[Feige et al., 2011]

- pick a random set

  $1/4$ for non-negative function (on expectation)

  $1/2$ for symmetric function (on expectation)

- local search
  - initialize $S$ to best singleton
  - $S$ = local optimum (add or delete elements)
  - return the best of $S$ and $U \setminus S$

  $1/3$ approx for non-negative function

  $1/2$ for non-negative symmetric function

# random set analysis

- for $A \subseteq U$, $A(p)$ is a random set where each element of $A$ is selected with prob $p$
- algorithm returns $R = U(1/2)$
- lemma I

$$E[f(A(p))] \geq (1-p)\, f(\emptyset) + p\, f(A)$$

  can prove by induction on the size of $A$
  and using the submodularity property

- lemma II

$$\begin{aligned} E[f(A(p) \cup B(q))] \quad \geq \quad & (1-p)(1-q)\, f(\emptyset) + \\ & p(1-q)\, f(A) + \\ & (1-p)q\, f(B) + \\ & pq\, f(A \cup B) \end{aligned}$$

  to prove use lemma I

# random set analysis

- algorithm returns

$$R = U(1/2) = S^*(1/2) \cup \overline{S^*}(1/2)$$

- by applying lemma II

$$\begin{aligned} E[f(R)] &= E[f(S^*(1/2) \cup \overline{S^*}(1/2))] \\ &= \frac{1}{4}f(\emptyset) + \frac{1}{4}f(S^*) + \frac{1}{4}f(\overline{S^*}) + \frac{1}{4}f(U) \end{aligned}$$

- gives $1/4$ for non-negative and $1/2$ for symmetric function

# unconstrained problem

[Feige et al., 2011]

- local search
  - initialize $S$ to best singleton
  - $S$ = local optimum (add or delete elements)
  - return the best of $S$ and $U \setminus S$

$1/3$ approx for non-negative function

$1/2$ for non-negative symmetric function

# analysis of local search

- lemma if $S$ is a local optimum then
  $f(S) \geq f(T)$ for all $S \subseteq T$ and $T \subseteq S$

- proof
  take $S \subseteq T$ and consider $S = X_0 \subseteq \ldots X_\ell = T$
  by submodularity and local optimality

  $$0 \geq f(S \cup \{x_i\}) - f(S) \geq f(X_i) - f(X_{i-1})$$

  summing up gives $0 \geq f(X_\ell) - (X_0)$ or $f(S) \geq (T)$

- corollary
  for optimum $S^*$ and local optimum $S$ it is
  $f(S) \geq f(S \cup S^*)$ and $f(S) \geq f(S \cap S^*)$

# analysis of local search (cont)

- it is

$$f(S) \geq f(S \cup S^*) \;\; \text{and} \;\; f(S) \geq f(S \cap S^*)$$

- by submodularity and non-negativity

$$f(S \cup S^*) + f(U \setminus S) \geq f(S^* \setminus S) + f(U) \geq f(S^* \setminus S)$$

$$f(S \cap S^*) + f(S^* \setminus S) \geq f(S^*) + f(\emptyset) \geq f(S^*)$$

- combining we get

$$2f(S) + f(U \setminus S) \geq f(S^*)$$

- and so

$$\max\{f(S), f(U \setminus S)\} \geq \frac{1}{3} f(S^*)$$

# unconstrained problem

[Buchbinder et al., 2015]

- tight $1/2$ approximation for general non-negative submodular function
- randomized algorithm, approximation $1/2$
- deterministic algorithm, approximation $1/3$

# deterministic algorithm

[Buchbinder et al., 2015]

---

**Algorithm 1:** DeterministicUSM$(f, \mathcal{N})$

---

$X_0 \leftarrow \emptyset$, $Y_0 \leftarrow \mathcal{N}$.
**for** $i = 1$ *to* $n$ **do**
    $a_i \leftarrow f(X_{i-1} \cup \{u_i\}) - f(X_{i-1})$.
    $b_i \leftarrow f(Y_{i-1} \setminus \{u_i\}) - f(Y_{i-1})$.
    **if** $a_i \geq b_i$ **then** $X_i \leftarrow X_{i-1} \cup \{u_i\}$, $Y_i \leftarrow Y_{i-1}$.
    **else** $X_i \leftarrow X_{i-1}$, $Y_i \leftarrow Y_{i-1} \setminus \{u_i\}$.
**return** $X_n$ (or equivalently $Y_n$).

---

# randomized algorithm

[Buchbinder et al., 2015]

**Algorithm 2:** RandomizedUSM$(f, \mathcal{N})$

$X_0 \leftarrow \emptyset$, $Y_0 \leftarrow \mathcal{N}$.
**for** $i$ = 1 *to* $n$ **do**
    $a_i \leftarrow f(X_{i-1} \cup \{u_i\}) - f(X_{i-1})$.
    $b_i \leftarrow f(Y_{i-1} \setminus \{u_i\}) - f(Y_{i-1})$.
    $a'_i \leftarrow \max\{a_i, 0\}$, $b'_i \leftarrow \max\{b_i, 0\}$.
    **with probability** $a'_i/(a'_i + b'_i)^*$ **do**:
    $X_i \leftarrow X_{i-1} \cup \{u_i\}$, $Y_i \leftarrow Y_{i-1}$.
    **else** (with the compliment probability $b'_i/(a'_i + b'_i)$)
    **do**: $X_i \leftarrow X_{i-1}$, $Y_i \leftarrow Y_{i-1} \setminus \{u_i\}$.
**return** $X_n$ (or equivalently $Y_n$).

* If $a'_i = b'_i = 0$, we assume $a'_i/(a'_i + b'_i) = 1$.

# max-sum diversification

[Borodin et al., 2012]

- $U$ is a ground set
- $d : U \times U \to \mathbb{R}$ is a metric distance function on $U$
- $f : 2^U \to \mathbb{R}$ is a submodular function

- we want to find $S \subseteq U$ such that
  $\phi(S) = f(S) + \lambda \sum_{u,v \in S} d(u,v)$ is maximized and
  $|S| \leq k$

# max-sum diversification

[Borodin et al., 2012]

- consider $S \subseteq U$ and $x \in U \setminus S$
- define the following types of marginal gain

  $d_x(S) = \sum_{v \in S} d(x, v)$

  $f_x(S) = f(S \cup \{x\}) - f(S)$

  $\phi_x(S) = \frac{1}{2} f_x(S) + \lambda d_x(S)$

- greedy algorithm on marginal gain $\phi_x(S)$ gives factor 2 approximation

# max-sum diversification – the greedy

[Borodin et al., 2012]

1. $S \leftarrow \emptyset$
2. while $|S| < k$
3. $\quad i \leftarrow \arg\max_{\{j \in U \setminus S\}} \phi_j(S)$
4. $\quad S \leftarrow S \cup \{i\}$
5. return $S$

# conclusions

- maximization of submodular functions
- monotone, constraints, symmetric, ...
- recent developments in theory community
- simple algorithms
- neat analysis
- many applications in data mining

# references


Borodin, A., Lee, H. C., and Ye, Y. (2012).
Max-sum diversification, monotone submodular functions and dynamic updates.
In *Proceedings of the 31st symposium on Principles of Database Systems*, pages 155–166. ACM.

Buchbinder, N., Feldman, M., Seffi, J., and Schwartz, R. (2015).
A tight linear time (1/2)-approximation for unconstrained submodular maximization.
*SIAM Journal on Computing*, 44(5):1384–1402.

Feige, U., Mirrokni, V. S., and Vondrak, J. (2011).
Maximizing non-monotone submodular functions.
*SIAM Journal on Computing*, 40(4):1133–1153.

Goemans, M. X. and Williamson, D. P. (1995).
Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming.
*Journal of the ACM (JACM)*, 42(6):1115–1145.

# references (cont.)


Kempe, D., Kleinberg, J., and Tardos, E. (2003).
Maximizing the spread of influence through a social network.
In *KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM Press.

Nemhauser, G. L., Wolsey, L. A., and Fisher, M. L. (1978).
An analysis of approximations for maximizing submodular set functions I.
*Mathematical Programming*, 14(1):265–294.

Trevisan, L. (2012).
Max cut and the smallest eigenvalue.
*SIAM Journal on Computing*, 41(6):1769–1786.