About Greedoids

Michel Habib,
habib@liafa.univ-paris-diderot.fr
http://www.liafa.univ-paris-diderot.fr/~habib

21 septembre 2016

Schedule

A tour in greediness,

Interval property and greedoid rank function

Rank functions for greedoids

Greedoids as languages,

Branchings and Prim, Dijkstra algorithms

Other examples

The objective is to understand algorithmic greediness, and in particular why it works so well in applications.

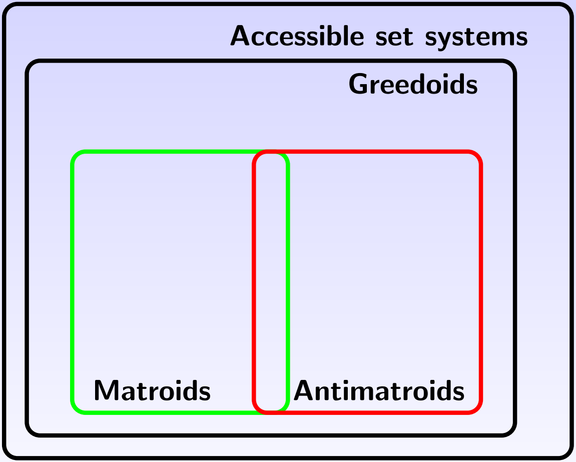

Figure: Greediness Landscape

En français le Glouton ou Carcajou

Accessible systems

$(V, \mathcal{F})$ is an accessible system if it satisfies the following condition : a finite set V, $\mathcal{F}$ a family of subsets of V that satisfies :

1. $\emptyset \in \mathcal{F}$
2. If $X \in \mathcal{F}$ such that $X \neq \emptyset$ then $\exists x \in X$ such that $X \setminus x \in \mathcal{F}$

Remarks :

- If $\mathcal{F}$ is not empty then axiom 2 implies that $\emptyset \in \mathcal{F}$. Axiom 2 is just a weakening of the hereditary property of matroids.
- For every accessible set, there exists a way to build it from the empty set by adding at each step a new element.

Accessible exchange systems also called gredoids

$GR = (V, \mathcal{F})$ an accessible exchange system or greedoid :
for a finite set V, is $\mathcal{F}$ a family of subsets of V that satisfies :

1. $\emptyset \in \mathcal{F}$
2. If $X \in \mathcal{F}$ such that $X \neq \emptyset$ then $\exists x \in X$ such that $X \setminus x \in \mathcal{F}$
3. If $X, Y \in \mathcal{F}$ et $|X| = |Y| + 1$ then $\exists x \in X$-Y such that :
   $Y \cup \{x\} \in \mathcal{F}$.
4. Or equivalently
   If $X, Y \in \mathcal{F}$ et $|X| > |Y|$ then $\exists x \in X$-Y such that :
   $Y \cup \{x\} \in \mathcal{F}$.

Notation

An element of $\mathcal{F}$ is called a feasible element.

As a consequence :
All maximal (under inclusion) elements of $\mathcal{F}$ have the same size are called basis. This common size is called the rank of the greedoid.

A matroid by its independents

Withney [1935] : Let $M = (X, \mathcal{I})$ be a set family where X is a finite ground set and $\mathcal{I}$ a family of subsets of X.
M is a matroid if it satisfies the 3 following axioms :

1. $\emptyset \in \mathcal{I}$
2. If $I \in \mathcal{I}$ alors $\forall J \subseteq I$, $J \in \mathcal{I}$
3. If $I, J \in \mathcal{I}$ et $|I| = |J| + 1$ then $\exists x \in I$-J such that : $J \cup \{x\} \in \mathcal{I}$.

Variations around axiom 3

- 3' If $I, J \in \mathcal{I}$ and $|I| > |J|$ then $\exists x \in I$-J such that : $J \cup \{x\} \in \mathcal{I}$.
- 3" $\forall A \subseteq X$ and $\forall I, J \subseteq A$, $I, J \in \mathcal{I}$ and maximal then $|I| = |J|$
- It is easy to be convinced that 3, 3' and 3" are equivalent.

To be an hereditary family is stronger than axiom 2 of greedoids. Thus every matroid is a greedoid. Furthermore matroids are exactly the hereditary greedoids.

Examples of greedoid :

Rooted trees systems or undirected branchings

For an undirected connected graph G and a vertex $r \in G$, we define the set systems :
$(E(G), \mathcal{F})$
Where $F \in \mathcal{F}$ iff F is the edge set of a tree containing r.

Rooted trees gredoids

Axiom 1 of greedoids is trivially satisfied, because r itself is a tree rooted in r.
If F defines a tree and has at least one edge, then it contains least two leaves x, y. Say $x \neq r$ and e_x its attached edge.
$F \setminus e_x$ is a tree (eventually empty is $y = r$). Therefore axiom 2 is satisfied.

Axiom 3

Let X, Y be two subtrees rooted in r such that : $|X| > |Y|$.
If we denote by $span(X)$, resp. $span(Y)$ the set of vertices adjacent to at least one edge of X, resp. Y.
$|span(X)| > |span(Y)|$ and therefore if we consider the cut C yielded by $span(Y)$ in $span(X)$ around the component containing r, there exists at least one edge e in this cut and $Y + e$ is a tree containing r.
So $Y + e \in \mathcal{F}$.

Theorem

Undirected rooted branchings are greedoids.
A basis of an undirected branching corresponds to the set of edges of a graph search in G starting at r.

Rooted trees are not matroids

Although forests in a graph yield a matroid, rooted trees does not share the hereditary property of matroids, since deleting an edge could disconnect the rooted tree.

Rooted branchings in directed graphs

Rooted branchings

For a directed connected graph G and a vertex $x_0 \in G$, we define the set systems :
$(A(G), \mathcal{F})$
Where $F \in \mathcal{F}$ iff $F \subseteq A(G)$ is the arc set of an arborescence containing x_0 or equivalently for each vertex $x \in F$ there is a unique path from x_0 to x.

Rooted trees gredoids

Axiom 1 of greedoids is trivially satisfied, because r itself is a tree rooted in r.
If F is an arborescence and has at least one arc, then it contains least two leaves x, y. Say $x \neq x_0$ and a_x its attached arc.
$F \setminus a_x$ is an arborescence (eventually empty is $y = x_0$). Therefore axiom 2 is satisfied.

Axiom 3

Let X, Y be two arborescences rooted in x_0 such that : $|X| > |Y|$.
If we denote by $span(X)$, resp. $span(Y)$ the set of vertices adjacent to at least one arc of X, resp. Y.
$|span(X)| > |span(Y)|$ and therefore if we consider the cut C yielded by $span(Y)$ in $span(X)$ around the component containing x_0, there exists at least one arc a in this cut and $Y + a$ is a tree containing x_0.
So $Y + a \in \mathcal{F}$.

Theorem

Directed rooted branchings are greedoids.
$Br = (V(G), \mathcal{F})$ is a directed branching greedoid if
$\mathcal{F}$ is the set of all branchings rooted in x_0 of G.
A basis of an undirected branching corresponds to the set of arcs of a graph search in G starting at x_0.

Directed branchings are not matroids

Since they does not share the hereditary property of matroids, since deleting an edge could disconnect the branching.

Antimatroids another example of greedoids

$AM = (V, \mathcal{F})$ is an antimatroid :
a finite set V, $\mathcal{F}$ a family of subsets of V that satisfies :

1. $\emptyset \in \mathcal{F}$
2. If $X \in \mathcal{F}$ such that $X \neq \emptyset$ then $\exists x \in X$ such that $X \setminus x \in \mathcal{F}$
3. If $X, Y \in \mathcal{F}$ and $Y \not\subseteq X$ then $\exists x \in Y \setminus X$ such that $X \cup \{x\} \in \mathcal{F}$.

Remarks :

1. Axiom 3 of antimatroids implies the third axiom of greedoids. Because if $|Y| = |X| + 1$ then necessarily $Y \nsubseteq X$.
2. Therefore Antimatroids yield another particular case of greedoids
3. We notice that Axiom 3 implies $X \cup Y \in \mathcal{F}$.

- Antimatroid is a bad name, since an antimatroid is not the inverse of some matroid.
  Anti comes from anti-exchange property which we will see soon.
- Matroids are closed by intersection
- Antimatroids are closed by union
- But both are greedoids !
- The intersection of 2 matroids is not a matroid.

The intersection (i.e. ; common parts) of a matroïd and an antimatroïd does not lead to a greedoïd.
But there exist set families which are both matroïd and antimatroïd. As for example the trivial complete set family 2^V.

A large class of greedoids share the interval property

Interval property

A greedoid $(E, \mathcal{F})$ satisfies the interval property if :
$\forall A, B, C \in \mathcal{F}$, with $A \subseteq B \subseteq C$ and $x \in E \setminus C$,
$A \cup x, C \cup x \in \mathcal{F}$ implies $B \cup x \in \mathcal{F}$.

First consequences

Matroids and Antimatroids have both this interval property.
Directed branchings also satisfy this property, but not every greedoid does.

Interval property without lower bounds

Matroids also satisfy a stronger property :
$\forall A, B \in \mathcal{F}$, with $A \subseteq B$ and $x \in E \setminus C$,
$B \cup x \in \mathcal{F}$ implies $A \cup x \in \mathcal{F}$.

Interval property without upper bounds

Antimatroids also satisfy a stronger property :
$\forall A, B \in \mathcal{F}$, with $A \subseteq B$ and $x \in E \setminus C$,
$A \cup x \in \mathcal{F}$ implies $B \cup x \in \mathcal{F}$.

Rank function for greedoids

Let $GR = (V, \mathcal{F})$ be a greedoid.
We define :
$r : 2^V \rightarrow N$ such that :
$\forall A \subseteq V$, $r(A) = max\{|F| \text{ such that } F \in A \cap \mathcal{F}\}$

Characterization theorem

A function $r : 2^V \to N$ is a rank function of a greedoid iff r satisfies the following properties :

1. $r(\emptyset) = 0$
2. $\forall A \in 2^V$, $r(a) \leq |A|$
3. $\forall A, B \in 2^V$, $A \subseteq B$ implies $r(A) \leq r(B)$
4. [a] $\forall A \in 2^V$ and $x, y \in V$
   $r(A) = r(A + x) = r(A + y)$ implies $r(A) = r(A + x + y)$

---

a. This property is called local submodularity

Some usual definitions

Given a finite alphabet E, we denote by E^* the free monoid of all words over the alphabet E.

For a word ω we denote by $|\omega|$ its length (number of letters) and ω_s its support i.e., the set of letters in ω

A word is simple if it does not contain any letter more than once (i.e., $|\omega| = |\omega_s|$), and language $L \subset E^*$ is simple if it is made up with simple words.

Every simple language over a finite alphabet is always finite. Let us denote by E_s^* the set of simple words over the alphabet E.

A simple language $L \subseteq E_s^*$ over a finite alphabet E is a greedoid language if it satisfies :

1. If $\omega = \mu\nu \in L$ then $\mu \in L$.
2. If $\omega, \mu \in L$ and $|\omega| > |\mu|$, then ω contains a letter x such that $\mu x \in L$

Maximal words in L are called basic words.
This presentation of greedoids as languages captures the construction process of feasible words.

Greedoids and greedoid languages are somehow equivalent

1. If L is a greedoid language over an alphabet E then L_s is a greedoid.
2. If $(E, \mathcal{F})$ in which E is a finite set, $\mathcal{F}$ a family of subsets of E is a greedoid
   then $L(F) = \{x_1 \ldots x_k \in E_s^*$ such that $\{x_1, \ldots x_i\} \in F$ for every i, $1 \leq i \leq k\}$
   is a greedoid language.

First examples

- Ideals (resp. filters) of a partial order yield an antimatroid.
- Shelling of a forest. $T = (X, E)$ a forest.
  $AM = (E, \mathcal{S})$ such that $\mathcal{S} = \{F \subseteq E$ such that $E \setminus F$ has the same number of non trivial connected components as $T\}$

Example of the most famous antimatroid

Let G be a chordal graph.
$\mathcal{F}= \{X \subseteq V(G) | X$ is the beginning of a simplicial elimination scheme of $G\}$
$AM = (V, \mathcal{F})$ is an antimatroid.

proof

Axiomes 1 and 2 are trivially satisfied by $\mathcal{F}$.
Let us consider the third one. First we use the Y as the starting of a simplicial elimination of G. Consider the ordering τ of X which is the starting of a simplicial elimination scheme of G. Let σ be an elemination scheme starting with Y. Let x_1 the first element of $X \setminus Y$ with respect to σ.
Necessarily just after Y we can use x_1 as a simplicial vertex in the remaining graph, since all necessary vertices as been eliminated before (in $Y \cap X$) and x_1 was simplicial in τ.
Therefore $Y \cup \{x\}$ is the beginning of a simplicial elimination scheme of G.

Natural question

Which are the conditions that must satisfies an elimination scheme for a particular class of graphes to yield an antimatroid ?
(false or true twins for cographs, pending vertex and twin for distance hereditary graphs . . .)

The problem :
Given a greedoid language L over an alphabet E, an objective function $c : L \to \mathcal{R}$, find a basic word α which maximizes $c(\alpha)$.

The greedy algorithm

$\omega \leftarrow \epsilon$; % ϵ the empty word %
tant que $|A| \neq \emptyset$ **faire**
  $A \leftarrow \{x \in E$ tel que $\omega x \in L\}$;
  Choose $z \in A$ such that $c(\omega z) \geq c(\omega y)$ for all $y \in A$;
  $\omega \leftarrow \omega z$;
fin

An objective function $c : L \to \mathcal{R}$ is compatible with a greedoid language if it satisfies the following conditions :
For $\alpha x \in L$ such that $c(\alpha x) \geq c(\alpha.y)$ for every $\alpha y \in L$ (i.e., x is the best choice after α).

1. $\alpha\beta x\gamma \in L$ and $\alpha\beta z\gamma \in L$ imply that $c(\alpha\beta x\gamma) \geq c(\alpha\beta z\gamma)$ (i.e., "x" is the best choice at every stage)
2. $\alpha x\beta z\gamma \in L$ and $\alpha z\beta x\gamma \in L$ imply that $c(\alpha x\beta z\gamma) \geq c(\alpha z\beta x\gamma)$

Remark

Of course is $c(\alpha)$ depends only on the support set α_s then the second previous axiom is trivially true.

Main theorem

Characterization

Let $\mathcal{F}$ be an accessible set system, if for all compatible objective functions the greedy algorithm gives the optimum then $\mathcal{F}$ is a greedoid.

Comment

The proof is quite simple and straightforward for interval greedoids.

Applications

We are now going to show that Prim and Dijkstra are just instances of the greedy algorithm applied on their respective (undirected or directed) greedoids with special objective functions. Graphs are equipped with a positive cost function ω on the edges.

Dijkstra's algorithm

Prim and Dijkstra are just instances of the greedy algorithm applied on their respective (undirected or directed) branching greedoids with special objective functions. Graphs are equipped with a positive cost function ω on the edges.

- For Dijkstra the objective function is :
  For every directed branching F :
  $Cost(F) = \Sigma_{x \in F} d_F(x_0, x)$
  where $d_F(x_0, x)$ is the sum of the cost of the arcs of the unique path from x_0 to x in F.
- As a consequence of the main greediness theorem, Dijkstra's algorithm computes an optimum directed branching rooted in x_0 with respect to the function cost.

- Dijkstra's algorithm just maintains a data structure in order to be able to choose an arc with minimum cost. The greediness explains why it is not necessary to reevaluate a vertex already explored.
- Similar argument holds for Layered search LL (and also a BFS) which computes an optimum rooted branching minimizing the unit cost distance from x_0.
  Since at each step a minimal vertex is visited.
- Of course this does not hold for a DFS.
- And what for A* ?

Prim's algorithm

Prim and Dijkstra are just instances of the greedy algorithm applied on their respective (undirected or directed) greedoids with special objective functions. Graphs are equipped with a positive cost function ω on the edges.

- For Prim the objective function is :
  For every branching F :
  $Cost(F) = \Sigma_{x \in F} d_F(x_0, x)$
  where $d_F(x_0, x)$ is the maximum of the cost of the edge of the unique path from x_0 to x in F.
- As a consequence of the main greediness theorem, Prim's algorithm computes an optimum branching rooted in x_0 with respect to the function Cost.
- To finish we need to argue that it is a minimum spanning tree.

Another example : convex point set shelling

The shelling sequences of set of points E in the Euclidean space.
Each time a vertex is shelled, it must belong to the convex hull of the set of points.
Partial shelling sequences yield an antimatroid on E.

Proof

Same proof as for the antimatroid of simplicial elemination schemes.

This last antimatroid is interesting since it yields to convex geometries.
Let $AM = (V, \mathcal{F})$ be an antimatroid, we define :
If $\cup_{F \in \mathcal{F}} F = U$
$CG = (V, \mathcal{G})$ where $\mathcal{G} = \{U \setminus F | F \in \mathcal{F}\}$

Convex geometries satisfy an anti-exchange axiom, together with a closure operateur $\mathcal{L}$.
Let K a set and $x, y \notin K$ if $x \in \mathcal{L}(K+y)$ then $y \notin \mathcal{L}(K+x)$
Easy interpretation for convexity in Euclidian spaces where $\mathcal{L}$ means convex hull.

How can we use this antimatroidal structure ?

Optimisation on simplicial elimination schemes

For a chordal graph G :
Let $\tau = x_1, \ldots x_n$ be a simplicial elimination scheme.
$Bump(\tau) = \#$ consecutive pair of elements in τ which are not adjacent in the graph G.
$Bump(G) = min_\tau Bump(\tau)$
Can we compute in a greedy way a simplicial elimination scheme with a minimum number of bump ?
Seems to be related to the leafage problem :
We have $Bump(G) = 0$ if G is an interval graph and more generally :
$Bump(G) \leq Leafage(G) - 2$.

Submodular functions

Definition

E a finite ground set, and $f : 2^E \to \mathcal{R}$. f is submodular if for all $\forall A, B \subseteq E$, $f(A \cup B) + f(A \cap B) \leq f(A) + f(B)$

- For a bipartite graph $G = (X, Y, E)$
  Let us define $\forall A \subseteq X$, $\delta(A) = \#$ the number of edges adjacent to A.
  δ this generalized degree function is modular.
  $\delta(A) + \delta(B) = \delta(A \cup B) + \delta(A \cap B)$
- As a little generalization we can define, cut functions in graphs (resp. hypergraphs) as follows :
  $\forall A \subseteq V(G)$, $c(A) = \#$ the number of edges having exactly one extremity in A
  Then c is submodular (no equality).

Examples

- Number of splitters of a module (a case by case proof)
- Rank functions of matroids
- Any positive linear combination of submodular functions is still submodular.

We similarly define supermodular functions.

Sometimes we have to consider families that are only submodular on intersecting (resp. crossing) elements.

We called them intersecting (resp. crossing) submodular functions.

Minimizing a submodular functions is most of the time polynomial.
Examples (flows , modules, . . .).
Maximizing is more difficult.
Recent uses in Machine Learning Theory.

Polymatroids from Jack Edmonds

A polymatroid is a set systems for which the rank function is a monotone non-decreasing submodular function
verifying :
$f(\emptyset) = 0$
$\forall A, B \subseteq E$, $A \subseteq B$ implies $f(A) \leq f(B)$
Greedy algorithms can be defined in this framework.

Anders Björner and Günter M. Ziegler.
Introduction to greedoids.
Chapitre de livre jamais publié sous cette forme, 1989.

Brenda L. Dietrich.
Matroids and antimatroids - a survey.
Discrete Mathematics, 78 : 223-237, 1989.

Bernhard Korte and László Lovász.
The intersection of matroids and antimatroids.
Discrete Mathematics, 73(1-2) :143–157, 1989.