

Efficient Stepwise Selection in Decomposable Models

Amol Deshpande (UC Berkeley)

Minos Garofalakis (Bell Labs)

Michael I. Jordan (UC Berkeley)

Decomposable Models

👉 Undirected graphical models (Markov Networks)

- 👉 Applications: Image processing, translation models, word-sense disambiguation, multi-dimensional histograms

👉 Decomposable Models :

- 👉 Subclass with underlying graphical structure equivalent to chordal (triangulated) graphs

👉 Advantages :

- 👉 MLE can be computed analytically
- 👉 Closed form expressions for test statistics
- 👉 Can be represented by means of both directed and undirected graphs

👉 Mixtures of Decomposable Models can be used for more representational power

Searching for Optimal Decomposable Model

👉 Intractable in general

👉 Heuristic Search Techniques

👉 Need to make sure we stay in the space of Decomposable Models

👉 Backward Selection: Remove edges from the model incrementally

⇨ Theorem [Wermuth 76]: An edge can be removed from a decomposable model iff it belongs to exactly one maximal clique of the graph.

👉 Forward Selection: Add edges to a model incrementally

⇨ Naive Algorithm : $O(n^4)$ per step

👉 We will focus on Forward Selection in this talk

Forward Selection Problem Definition

- ➡ Add an edge to a given decomposable model to get the decomposable model that has the highest *score* amongst possible such models.
 - ➡ Step 1: Identify the edges that can be added without violating chordality of the graph.
 - ➡ Step 2: Choose the best edge given a score function and add that edge to the model.
- ➡ Example of a Score Function
 - ➡ Kullback-Leibler Divergence

Overview of the Talk

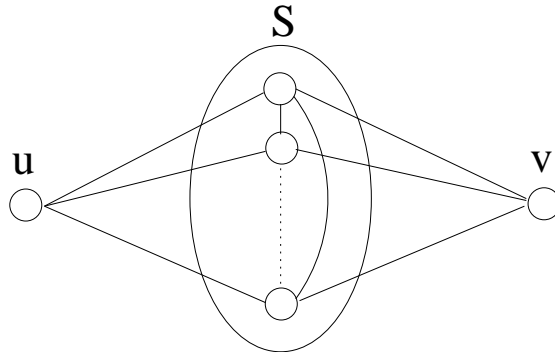
- ➡ Characterization of edges eligible for Forward Selection
- ➡ Auxiliary structures required
- ➡ Updating the auxiliary structures
- ➡ Complexity of the complete forward selection procedure
- ➡ Extensions

Characterization

👉 An edge (u, v) can be added to a chordal graph (V, E) while maintaining its chordality iff :

👉 $\exists S \subseteq V - \{u, v\}$ s.t. $(u, x), (x, v) \in E \quad \forall x \in S$

👉 S is the *minimal separator* for u and v

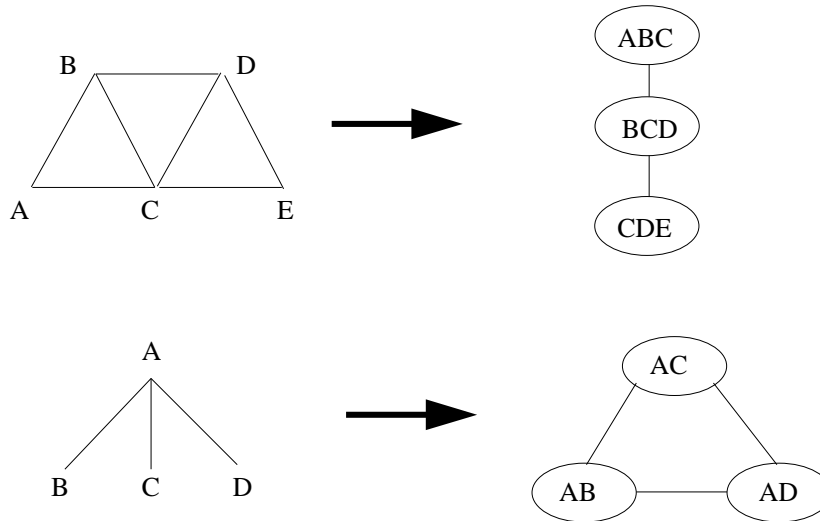


Clique Graph CG_G of a Chordal Graph G

👉 $CG_G = (V_{CG}, E_{CG})$ where :

👉 $V_{CG} =$ maximal cliques of G

👉 $(C_1, C_2) \in E_{CG}$ iff $S = C_1 \cap C_2$ separates $C_1 - S$ and $C_2 - S$



Properties of Clique Graph

- ➡ A maximum spanning tree of CG_G is a junction tree of G
- ➡ CG_G is union of all junction trees of G
- ➡ Equivalent to Almond Trees in representational power
- ➡ Can be used for inference as well
- ➡ Two nodes u and v in G satisfy the forward selection criterion iff $\exists (C_1, C_2) \in E_{CG}$ s.t. $u \in C_1$ and $v \in C_2$
 - ➡ Can be used to enumerate edges eligible for forward selection
 - ➡ $O(n^4)$ worst case

Auxiliary Structures

👉 $n \times n$ matrix, \mathcal{E}_G

👉 Indexed by the nodes of G

👉 $\mathcal{E}_G(u, v) = \text{true}$ iff (u, v) eligible for forward selection

👉 Used to enumerate eligible edges in $O(n^2)$ time

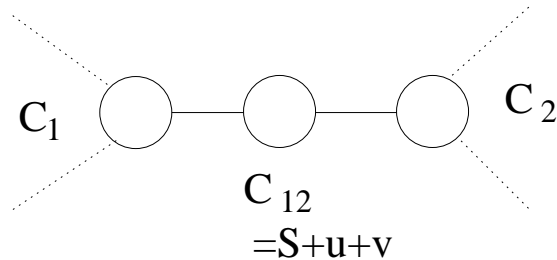
👉 Clique graph CG_G of the model graph

👉 Required to update \mathcal{E}_G in time $O(n^2)$ after adding an edge

👉 Time complexity of enumeration : $O(n^2)$

Updating the Clique Graph

👉 Step 1 : Add the newly generated maximal clique



👉 Step 2 : Remove existing edges if required

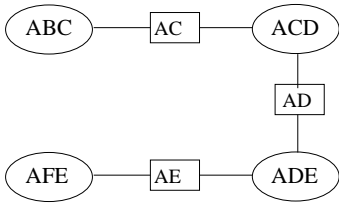
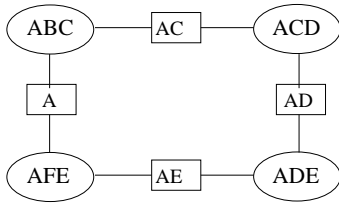
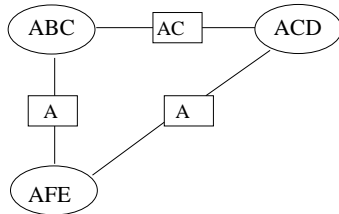
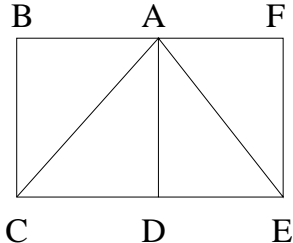
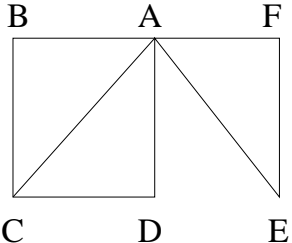
👉 Theorem: An edge (C_i, C_j) may have to be removed only if :

⇒ $C_i \cap C_j = S$.

⇒ u is connected to C_i and v to C_j in $G - S$.

👉 Step 3 : Add edges involving C_{12}

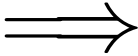
Updating the Clique Graph : Example



Updating \mathcal{E}_G

- Since some edges have been deleted from the clique graph, previously eligible node pairs may not be eligible any more.
- New node pairs of the form (u, x) or (v, x) may become eligible for forward selection.

	A	B	C	D	E	F
A	F	F	F	F	F	F
B	F	F	F	T	T	T
C	F	F	F	F	T	T
D	F	T	F	F	T	T
E	F	T	T	T	F	F
F	F	T	T	T	F	F



	A	B	C	D	E	F
A	F	F	F	F	F	F
B	F	F	F	T	F	F
C	F	F	F	F	T	F
D	F	T	F	F	F	T
E	F	F	T	F	F	F
F	F	F	F	T	F	F

Step 2 of Forward Selection Procedure

👉 KL Divergence from the saturated model (completely connected graph)

👉 Equivalent to minimizing entropy of the model

$$\Rightarrow H(\mathcal{M}) = \sum_{C \in \mathcal{C}} H(C) - \sum_{S \in \mathcal{S}} H(S)$$

where $H(A_1, A_2, \dots, A_k) = \sum_a \frac{f(a)}{N} \log \frac{f(a)}{N}$

👉 Only require the list of maximal cliques (\mathcal{C}) and separators (\mathcal{S}) in a junction tree

👉 Theorem: If \mathcal{M}' is obtained from \mathcal{M} by adding edge (u, v) , then

$$H(\mathcal{M}) - H(\mathcal{M}') = H(S + u) + H(S + v) - H(S + u + v) - H(S)$$

👉 Precompute all the entropies required for the next step

👉 Theorem: Number of new entropies that need to be computed after adding an edge is at most $4n$.

Overall Complexity of Update

👉 $O(n^2 + nE)$, where E is average time required to compute an entropy

👉 Enumerating eligible edges and selecting the *best* edge, $O(n^2)$

👉 Updating the auxiliary structures, $O(n^2)$

👉 Precomputing entropies required for next step, $O(nE)$

👉 Naive computation of entropies using sorting

👉 $E = O(m \log m)$, where m is the number of data points

👉 Time to compute the entropies can be reduced significantly by simultaneous computation

Extensions

👉 Backward Selection

- 👉 Similar time complexity
- 👉 Number of entropies required to be computed at each step at most n

👉 Strong Decomposability

- 👉 *E.g.*, Mixed Graphs, Decision Graphs
- 👉 Mixed Graphs :

➡ Theorem (Lauritzen, 1996): Given a strongly decomposable mixed graph $G = (V, E)$, with $V = V_d \cup V_c$, the graph $G' = (V + *, E + E_*)$, with $E_* = \{(*, v) \mid v \in V_d\}$, is chordal.

Conclusions

- ➡ Proposed a characterization for edges eligible for forward selection
- ➡ Presented an algorithm to do stepwise selection efficiently
- ➡ $O(n^2)$ time complexity per step as opposed to $O(n^4)$ of the naive algorithm

Bibliography

- ➡ Model search among multiplicative models, Nanny Wermuth, Biometrics, 1976
- ➡ Chordal graph and their clique graph, Galinier and Habib and Paul, Workshop on Graph-Theoretic Concepts, 1995
- ➡ Graphical Models, S. L. Lauritzen, Clarendon Press, Oxford, 1996
- ➡ Decomposable graphical Gaussian model determination, Giudici and Green, Biometrika,
- ➡ Word-sense disambiguation using decomposable models, Bruce and Wiebe, ACL, 1994