The max-min-hill-climbing algorithm

Michael Bauer

M.Sc. Comp. Science

May 24, 2014

Algorithm 1 $\overline{MMPC}$ Algorithm

procedure $\overline{MMPC}$ $(T, \mathcal{D})$
    Input: target variable T; data $\mathcal{D}$
    Output: the parents and children of T in any Bayesian network faithfully representing the data distribution
    %Phase I: Forward
    $\mathbf{CPC} = \emptyset$
    repeat
        $\langle F, assocF \rangle = MaxMinHeuristic(T; \mathbf{CPC})$
        if $assocF \neq 0$ **then**
            $\mathbf{CPC} = \mathbf{CPC} \cup F$
        end if
    until CPC has not changed

    %Phase II: Backward
    for all $X \in \mathbf{CPC}$ **do**
        if $\exists \mathbf{S} \subseteq \mathbf{CPC}$, s.t. $Ind(X; T|\mathbf{S})$ **then**
            $\mathbf{CPC} = \mathbf{CPC} \setminus \{X\}$
        end if
    end for

    return CPC
end procedure

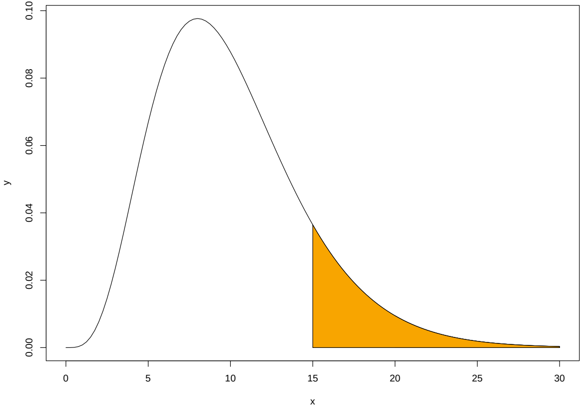
procedure MAXMINHEURISTIC(T,**CPC**)
    Input: target variable T; subset of variables **CPC**
    Output: the maximum over all variables of the minimum association with T relative to **CPC**, and the variable that achieves the maximum
    $assocF = \max_{X \in \mathcal{V}} MinAssoc(X; T|\mathbf{CPC})$
    $F = \arg\max_{X \in \mathcal{V}} MinAssoc(X; T|\mathbf{CPC})$
    return $\langle F, assocF \rangle$
end procedure

Definition

We calculate the G^2 value under the nullhypothesis of the conditional independence of $Ind_P(X_i, X_j|\mathbf{X}_k)$ holding. Then, the G^2 value is defined as:

$$G^2 := 2 * \sum_{a,b,c} S_{ijk}^{ab\mathbf{c}} * ln\left(\frac{S_{ijk}^{ab\mathbf{c}} * S_k^{\mathbf{c}}}{S_{ik}^{a\mathbf{c}} * S_{jk}^{b\mathbf{c}}}\right), \tag{1}$$

where S^{abc} is the number of times in the data where $X_i = a$, $X_j = b$ and $\mathbf{X}_k = \mathbf{c}$. We define in a similar fashion S^{ac}, S^{bc} and S^c, respectively.


0.00
0.02
0.04
0.06
0.08
0.10
y
0
5
10
15
20
25
30
x

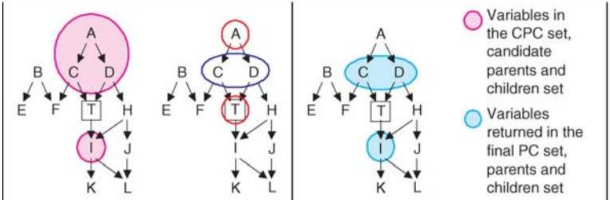
Algorithm 1 $\overline{MMPC}$ Algorithm

procedure $\overline{MMPC}$ $(T,\mathcal{D})$
  Input: target variable T; data $\mathcal{D}$
  Output: the parents and children of T in any Bayesian network faithfully representing the data distribution
  %Phase I: Forward
  $\mathbf{CPC} = \emptyset$
  repeat
    $\langle F, assocF\rangle = MaxMinHeuristic(T; \mathbf{CPC})$
    if $assocF \neq 0$ **then**
      $\mathbf{CPC} = \mathbf{CPC} \cup F$
    end if
  until $\mathbf{CPC}$ has not changed

  %Phase II: Backward
  for all $X \in \mathbf{CPC}$ **do**
    if $\exists \mathbf{S} \subseteq \mathbf{CPC}$, s.t. $Ind(X; T|\mathbf{S})$ **then**
      $\mathbf{CPC} = \mathbf{CPC} \setminus \{X\}$
    end if
  end for

  return $\mathbf{CPC}$
end procedure

procedure MAXMINHEURISTIC$(T,\mathbf{CPC})$
  Input: target variable T; subset of variables $\mathbf{CPC}$
  Output: the maximum over all variables of the minimum association with T relative to $\mathbf{CPC}$, and the variable that achieves the maximum
  $assocF = \max_{X \in \mathcal{V}} MinAssoc(X; T|\mathbf{CPC})$
  $F = \arg\max_{X \in \mathcal{V}} MinAssoc(X; T|\mathbf{CPC})$
  return $\langle F, assocF\rangle$



(k) (l) (m) Legend

Backward Phase

(l) $Ind(A;T|\{C,D\})$

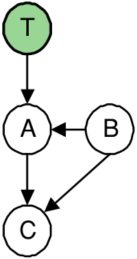
(m) Final Outcome of $\overline{MMPC}$ (T)

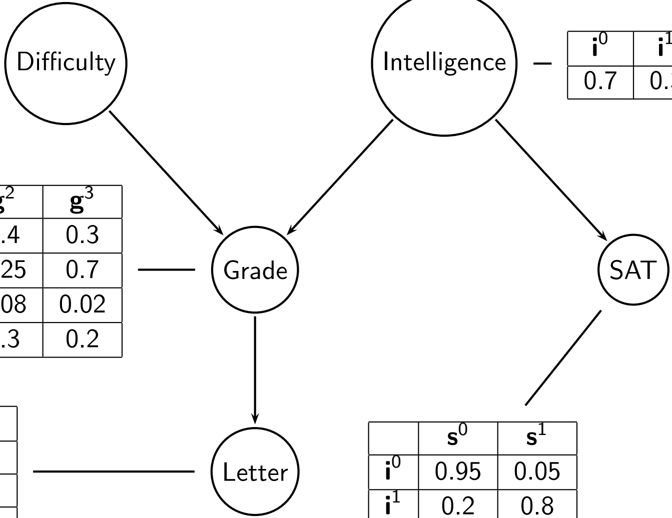
Phase II: Backward Phase

Algorithm 2 Algorithm *MMPC*

procedure MMPC(T,$\mathcal{D}$)
    $\mathbf{CPC} = \overline{MMPC}(T, \mathcal{D})$
    for every variable $X \in \mathbf{CPC}$ **do**
        if $T \notin \overline{MMPC}(X, \mathcal{D})$ **then**
            $\mathbf{CPC} = \mathbf{CPC} \setminus X$
        end if
    end for

    return CPC
end procedure


T
A
B
C



$\mathbf{d}^0$	$\mathbf{d}^1$
0.6	0.4

$\mathbf{i}^0$	$\mathbf{i}^1$
0.7	0.3

	$\mathbf{g}^1$	$\mathbf{g}^2$	$\mathbf{g}^3$
$\mathbf{i}^0,\mathbf{d}^0$	0.3	0.4	0.3
$\mathbf{i}^0,\mathbf{d}^1$	0.05	0.25	0.7
$\mathbf{i}^1,\mathbf{d}^0$	0.9	0.08	0.02
$\mathbf{i}^1,\mathbf{d}^1$	0.5	0.3	0.2

	$\mathbf{s}^0$	$\mathbf{s}^1$
$\mathbf{i}^0$	0.95	0.05
$\mathbf{i}^1$	0.2	0.8

	$\mathbf{l}^0$	$\mathbf{l}^1$
$\mathbf{g}^1$	0.95	0.05
$\mathbf{g}^2$	0.2	0.8
$\mathbf{g}^3$	0.2	0.8

Coming up soon: direct the edges

Thanks for your attention!