The max-min-hill-climbing algorithm

Michael Bauer

M.Sc. Comp. Science

June 2, 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$(T, \mathbf{CPC})$
  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

Reminder (Conditional Independence)

Two variables X and Y are conditionally independent given $\mathbf{Z}$ w.r.t a probability distribution P, denoted as $Ind_p(X; Y|\mathbf{Z})$, if $\forall x, y, \mathbf{z}$, where $P(\mathbf{Z} = \mathbf{z}) > 0$,

$$P(X, Y|\mathbf{Z}) = P(X|\mathbf{Z}) * P(Y|\mathbf{Z}), \tag{1}$$

where $P(X, Y|\mathbf{Z}) = P(X \cap Y|\mathbf{Z})$.

It is equivalent

$$Ind(X; T|\mathbf{Z}) \Longleftrightarrow (Assoc(X; T|\mathbf{Z}) = 0), \tag{2}$$

where $Assoc(X; T|\mathbf{Z})$ is the strength association (dependency) of X and T given $\mathbf{Z}$.

Reminder (Conditional Independence)

Two variables X and Y are conditionally independent given $\mathbf{Z}$ w.r.t a probability distribution P, denoted as $Ind_p(X; Y|\mathbf{Z})$, if $\forall x, y, \mathbf{z}$, where $P(\mathbf{Z} = \mathbf{z}) > 0$,

$$P(X, Y|\mathbf{Z}) = P(X|\mathbf{Z}) * P(Y|\mathbf{Z}), \tag{1}$$

where $P(X, Y|\mathbf{Z}) = P(X \cap Y|\mathbf{Z})$.
It is equivalent

$$Ind(X; T|\mathbf{Z}) \Longleftrightarrow (Assoc(X; T|\mathbf{Z}) = 0), \tag{2}$$

where $Assoc(X; T|\mathbf{Z})$ is the strength association (dependency) of X and T given $\mathbf{Z}$.
We define the minimum association of X and T relative to a feature subset $\mathbf{Z}$, as

$$MinAssoc(X; T|\mathbf{Z}) = \min_{\mathbf{S} \subseteq \mathbf{Z}} Assoc(X; T|\mathbf{S}). \tag{3}$$

Reminder (Conditional Independence)

Two variables X and Y are conditionally independent given $\mathbf{Z}$ w.r.t a probability distribution P, denoted as $Ind_p(X;Y|\mathbf{Z})$, if $\forall x, y, \mathbf{z}$, where $P(\mathbf{Z} = \mathbf{z}) > 0$,

$$P(X, Y|\mathbf{Z}) = P(X|\mathbf{Z}) * P(Y|\mathbf{Z}), \tag{1}$$

where $P(X, Y|\mathbf{Z}) = P(X \cap Y|\mathbf{Z})$.
It is equivalent

$$Ind(X; T|\mathbf{Z}) \iff (Assoc(X; T|\mathbf{Z}) = 0), \tag{2}$$

where $Assoc(X; T|\mathbf{Z})$ is the strength association (dependency) of X and T given $\mathbf{Z}$.
We define the minimum association of X and T relative to a feature subset $\mathbf{Z}$, as

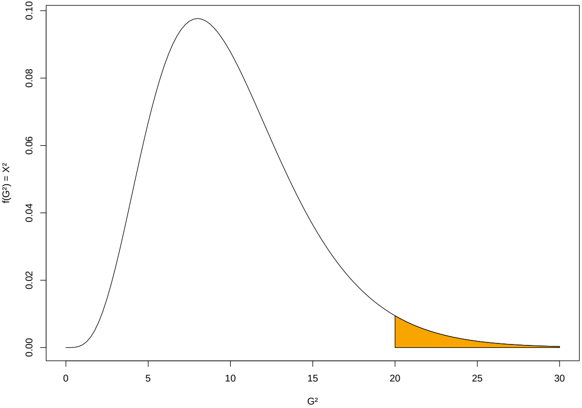
$$MinAssoc(X; T|\mathbf{Z}) = \min_{\mathbf{S} \subseteq \mathbf{Z}} Assoc(X; T|\mathbf{S}). \tag{3}$$

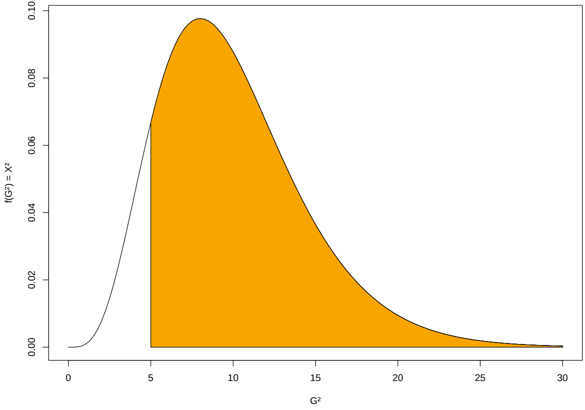
Definition

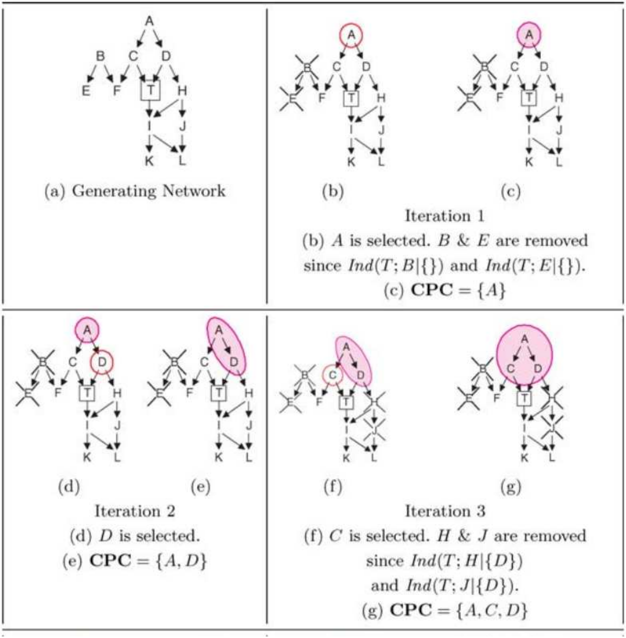
We calculate the G^2 value under the nullhypothesis of the conditional independence of $Ind_P(X_i, X_j|\mathbf{X}_k)$ holding. Then, the G^2 value is defined as:

$$G^2 := 2 * \sum_{a,b,c} S_{ijk}^{ab\mathbf{c}} * ln\left(\frac{S_{ijk}^{ab\mathbf{c}} * S_k^{\mathbf{c}}}{S_{ik}^{a\mathbf{c}} * S_{jk}^{b\mathbf{c}}}\right), \tag{4}$$

where S_{ijk}^{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_{ik}^{ac}, S_{jk}^{bc} and S_k^c, respectively.


0.10
0.08
0.06
0.04
0.02
0.00
f(G²) = X²
0
5
10
15
20
25
30
G²


0.10
0.08
0.06
0.04
0.02
0.00
f(G²) = X²
0
5
10
15
20
25
30
G²



(a) Generating Network

(b) (c)

Iteration 1

(b) A is selected. B & E are removed since $Ind(T; B|\{\})$ and $Ind(T; E|\{\})$.

(c) **CPC** $= \{A\}$

(d) (e)

Iteration 2

(d) D is selected.

(e) **CPC** $= \{A, D\}$

(f) (g)

Iteration 3

(f) C is selected. H & J are removed since $Ind(T; H|\{D\})$ and $Ind(T; J|\{D\})$.

(g) **CPC** $= \{A, C, D\}$

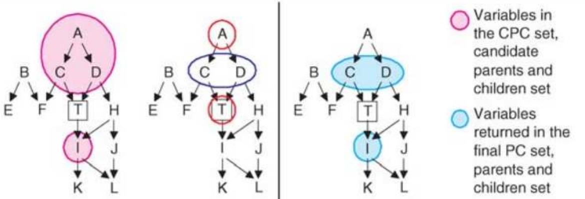
Algorithm 1 $\overline{MMPC}$ Algorithm

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$(T,\mathbf{CPC})$
    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



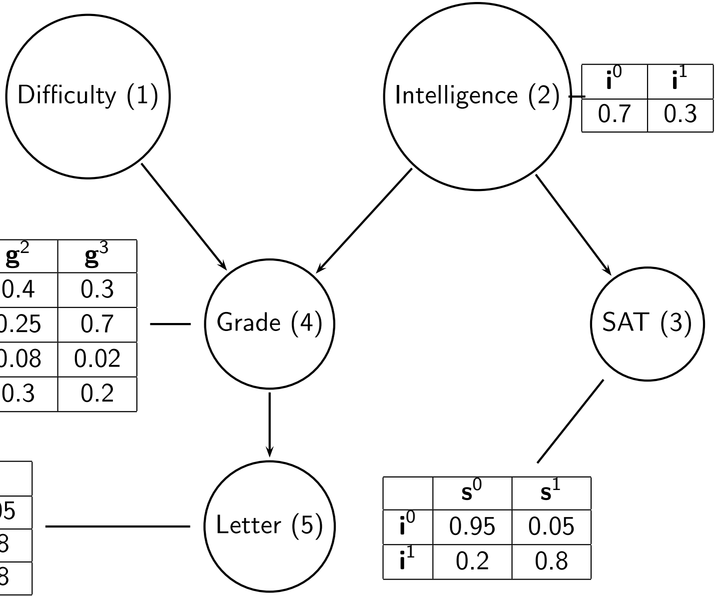
(k) (l) (m) Legend

Backward Phase

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

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

Phase II: Backward Phase



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

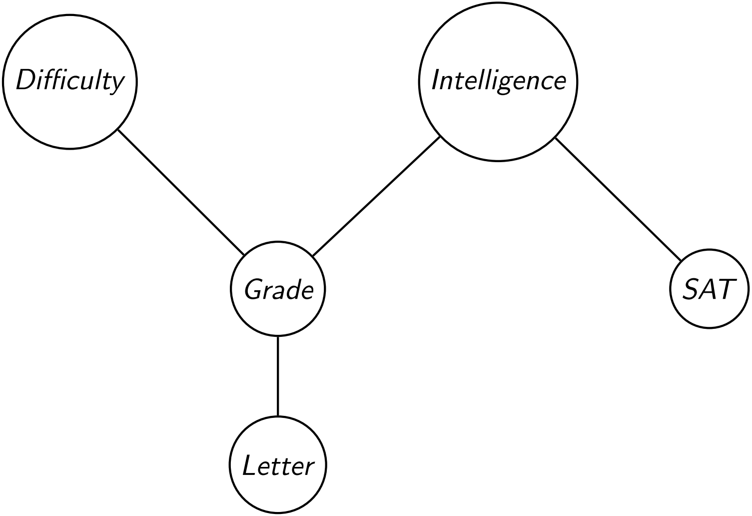
$\mathbf{i}^0$	$\mathbf{i}^1$
0.7	0.3

	$\mathbf{g}^1$	$\mathbf{g}^2$	$\mathbf{g}^3$
$\mathbf{i}^0,\mathbf{d}^0$	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

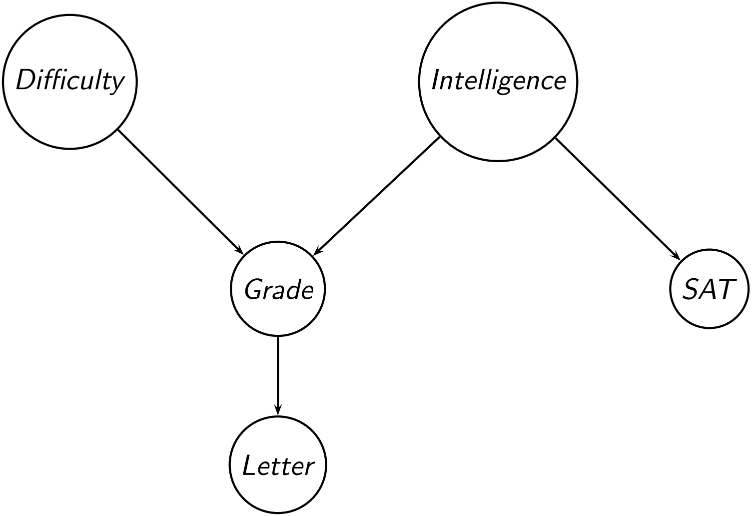
	$\mathbf{s}^0$	$\mathbf{s}^1$
$\mathbf{i}^0$	0.95	0.05
$\mathbf{i}^1$	0.2	0.8

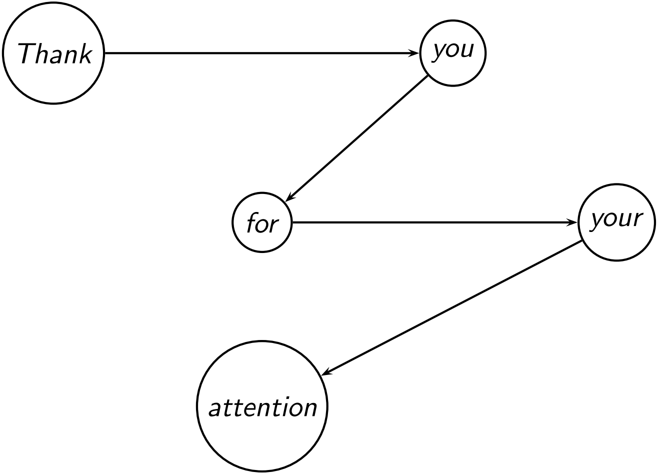
	$\mathbf{l}^0$	$\mathbf{l}^1$
$\mathbf{g}^1$	0.95	0.05
$\mathbf{g}^2$	0.2	0.8
$\mathbf{g}^3$	0.2	0.8

How this is working until now:



The last goal:




Thank
you
for
your
attention