Bayesian Belief Networks in R

Abraham Mathew

Carmichael Lynch

Data^3 2015
March 5, 2015


tinder
tinder
25
8

```
> head(dat, 10)
   Hair_Color  Race Text Pictures      Age Miles_Away Shared_Interest Overweight Liked_You
1       Black White    Y        5 22 to 26      Close               0          N         N
2      Blonde White    N        4 22 to 26      Close               1          N         N
3       Black Other    Y        4 27 to 30      Close               4          N         N
4      Blonde White    Y        5 22 to 26      Close               0          N         N
5      Blonde White    N        4 18 to 21      Close               1          N         N
6    Brunette White    Y        6 22 to 26      Close               0          N         N
7       Black White    N        4 22 to 26      Close               1          N         N
8      Blonde White    Y        6 22 to 26      Close               2          Y         N
9    Brunette White    Y        6 18 to 21        Far               0          N         N
10     Blonde White    Y        5      31+        Far               0          N         N
```

Review of Probability Theory

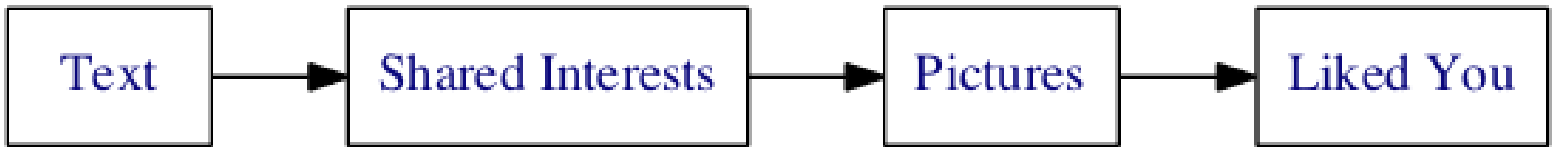
- Sample space: The set of all possible events of an experiment
- Marginal probability: the probability of an event occurring, or $P(A)$
- Conditional probability: the probability of event A occurring given that event B occurs, or $P(A|B)$
- Joint probability: the probability of event A and event B occurring, or $P(AandB)$

Probabilistic Graphical Models

- Probabilistic graphical models represent the conditional dependencies between random variables through a graph structure.
- Nodes correspond to random variables and edges represent statistical dependencies between the variables.
- Bayesian Belief Networks (directed acyclic graphics) and Markov Network (undirected).

Bayesian Belief Networks

- Goal: learn the structure of a network and learn the parameters.



- This is a directed acyclic graphic (DAG) that shows the dependencies between the variables.
- The relationship between shared interests and liked you is dependent on the number of pictures.

Bayesian Belief Networks

- $P(LY|P)$, or what is the probability that a user liked you, given the number of pictures on their profile?
- Each node is associated with a probability function that specifies the probability of that variable conditional on the parent node.

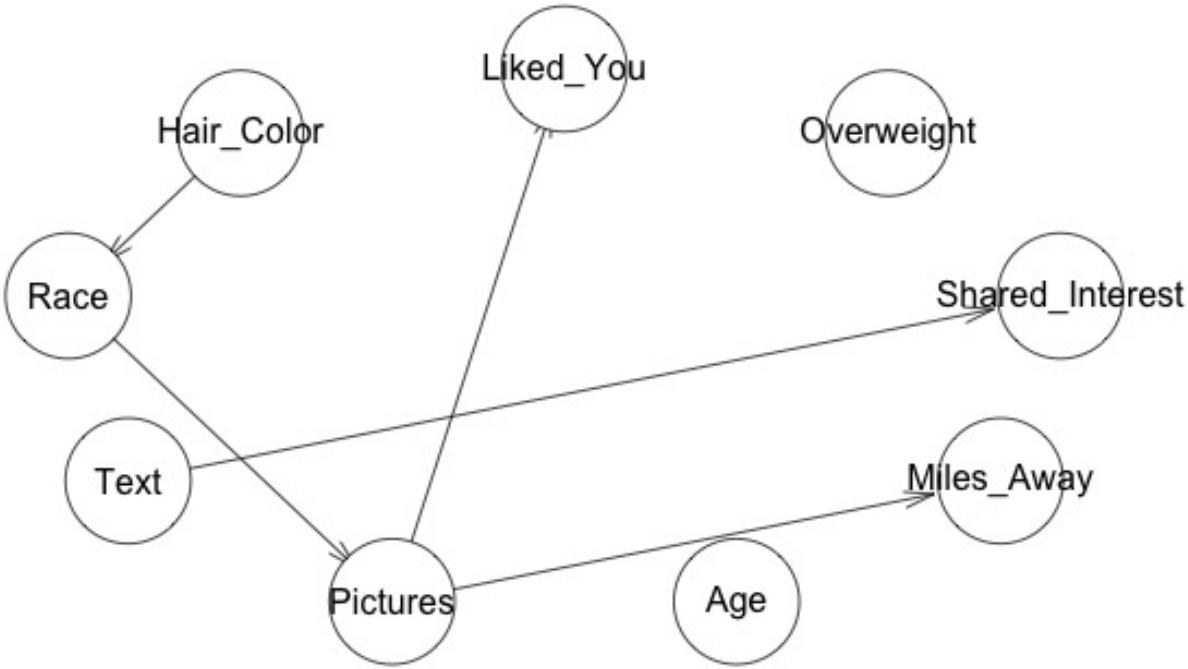
	Pictures			
Liked_You	3	4	5	6
N	0.1666667	0.6363636	0.5384615	0.7500000
Y	0.8333333	0.3636364	0.4615385	0.2500000

BNLEARN Package

- bnlearn package (CRAN)
- The package implements a number of constraint based, pairwise, score-based, and hybrid structure learning algorithms.
- For the Tinder data set, use a score-based algorithm to learn the network structure.

```
library(bnlearn)
net <- hc(dat)
plot(net)
```

BNLEARN Package



BNLEARN Package

```
## HIGHLIGHT NODE AND ITS MARKOV BLANKET
plot(net, highlight = c("Pictures",
                        mb(net, "Pictures")))
## ESTIMATE PARAMETERS
fitted <- bn.fit(net, dat)
fitted
## STRENGTH OF THE ARCS
arc.strength(net, dat)
## CONDITIONAL PROBABILITY QUERIES
cpquery(fitted, (Liked_You=="Y"), TRUE)
cpquery(fitted, (Liked_You=="N"),
                (Pictures=="3" & Text=="N"))
```

Further Reading

- Ross, Sheldon. Introduction to Probability Models
- Scutari, Marco. Bayesian Networks: With Examples in R
- Nagarajan, Radhakrishnan, Scutari, Marco, and Lèbre, Sophie. Bayesian Networks in R: with Applications in Systems Biology