36-350: Data Mining

Handout 5
September 10, 2003

Information, interactions, and 20 questions

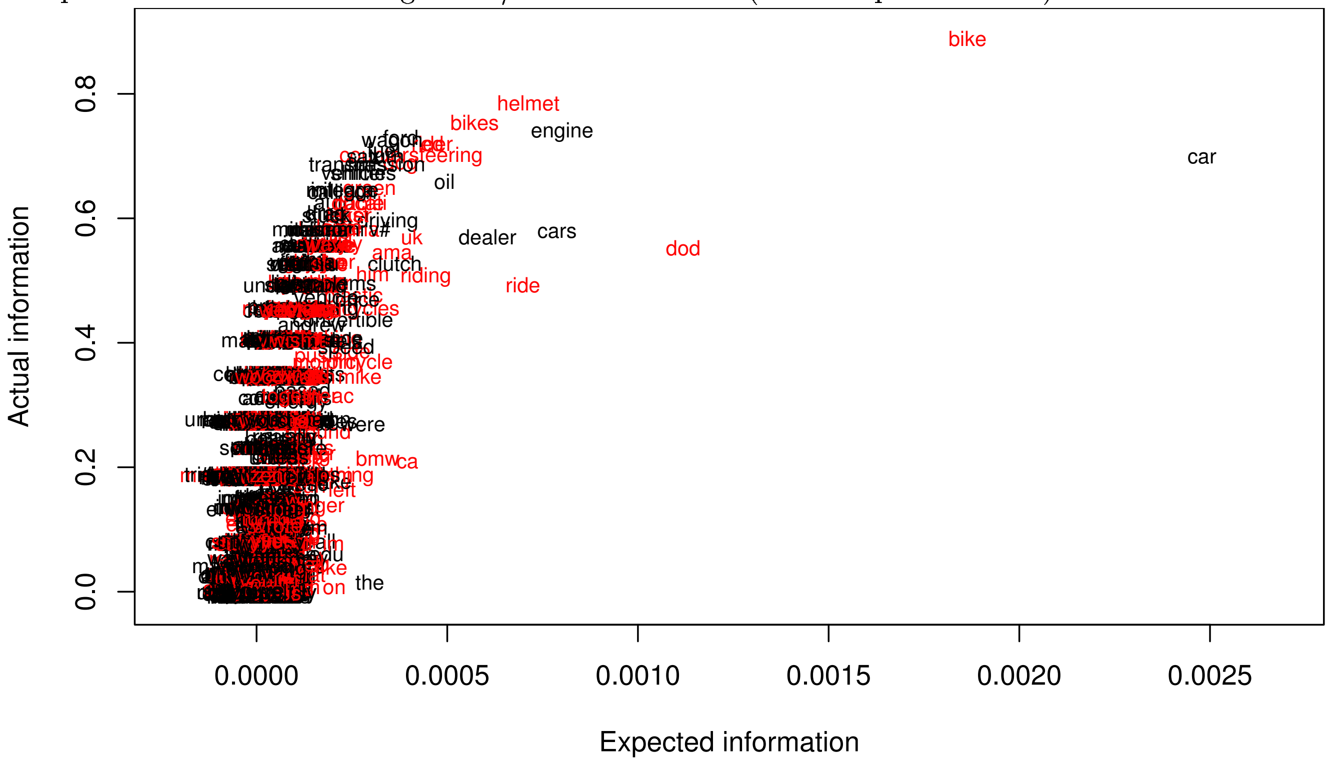
Last time we set up a game in which we went to a random position in the document and tested for a particular word. This tends to give small expected information, since the answer is usually "no". A different question we could ask is "Is this word present anywhere in the document?" The expected information we get from this question can be computed from a table of word-presence counts. For example, this is the table for testing if the word "car" is present (in the larger collection of 200 documents):

```
      word
label  FALSE TRUE
  auto 47    53
  moto 95     5
```

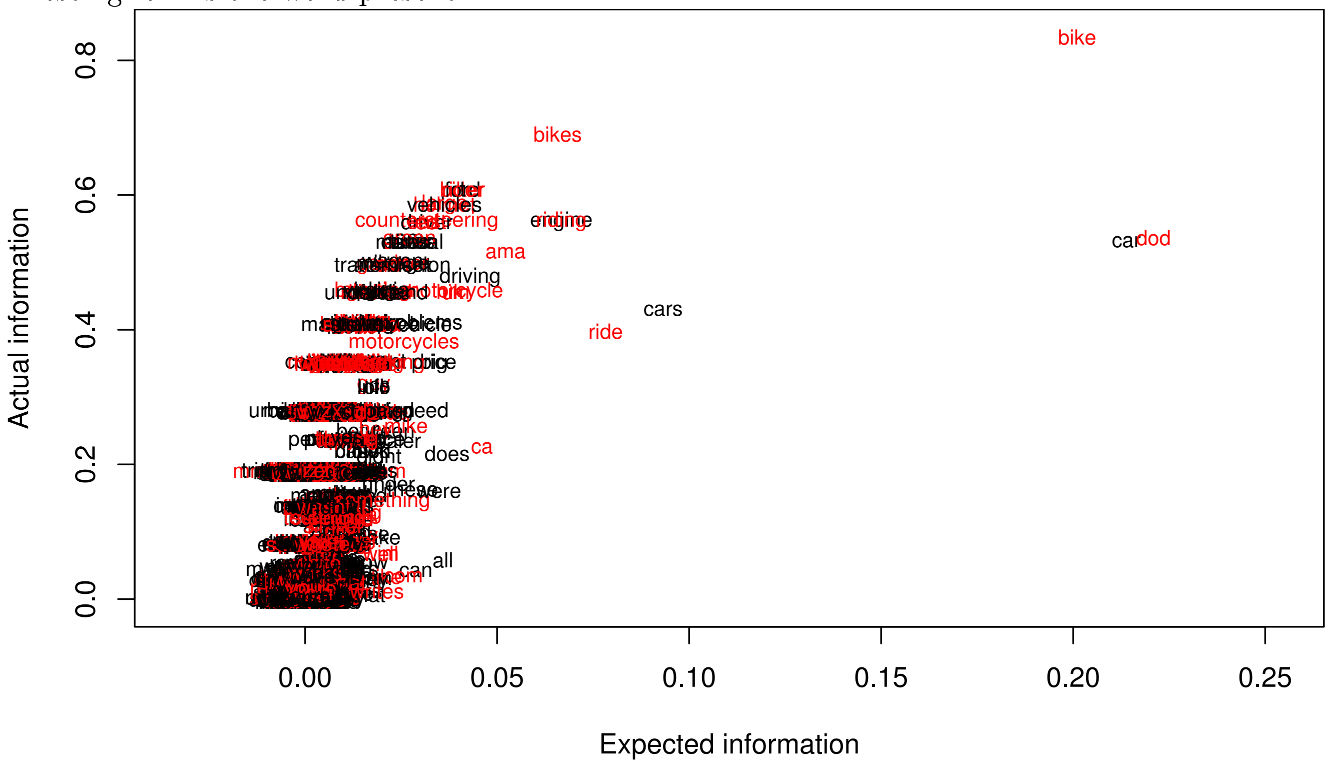
This table tells you that "car" is present in 58 documents, 53 of which are in the "auto" group, and 5 of which are in the "moto" group. The other column describes the cases where "car" is not present. The expected information is 0.21 bits, compared to 0.0025 for the random-position test.

We can do this computation for all words and sort them, as shown on the next page. Interestingly, "car" is not the most informative. "DoD" is the name of an on-line motorcycle club.

Important words for the large auto/moto collection: (random-position test)



Testing for “is the word present?”:



Black means the word favors “auto”. Red means the word favors “moto”.

So far we've been computing the information content in a single word. Suppose we are allowed to ask another question about the document, after getting the answer to the first question. The best second question is not necessarily the second-best first question. For example, if you know "car" is present, it doesn't help as much to know that "cars" is also present. This effect is called **interaction**. Whereas correlation and information are properties between two variables, interaction is a property between three variables. Interaction is when measuring one variable changes the importance of another variable.

Measuring interaction

Conditional information is the information x gives about c when a third variable y is already known:

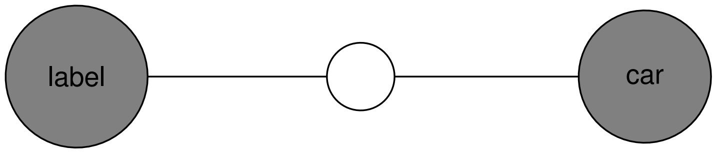
$$\begin{aligned} \mathcal{I}(c, x|y = \mathrm{y}) &= \mathcal{H}(c|y = \mathrm{y}) - \mathcal{H}(c|x, y = \mathrm{y}) && \text{(actual conditional information)} \\ \mathcal{I}(c, x|y) &= \sum_{y=\mathrm{y}} p(y = \mathrm{y})\mathcal{I}(c, x|y = \mathrm{y}) && \text{(expected conditional information)} \end{aligned}$$

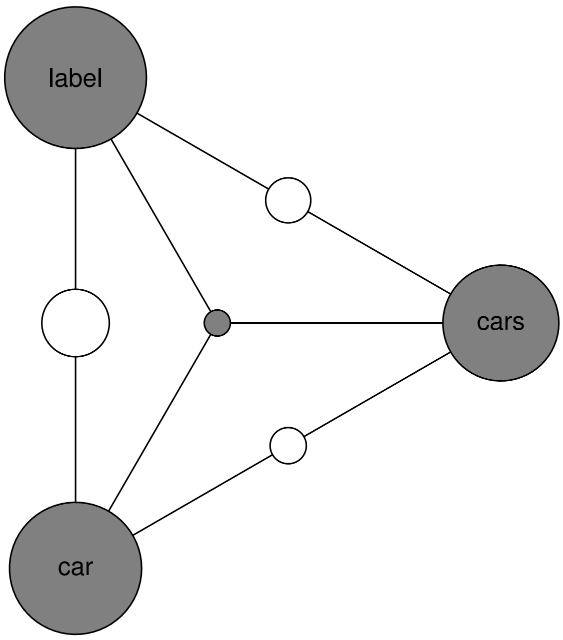
Interaction measures how y changes the information in x:

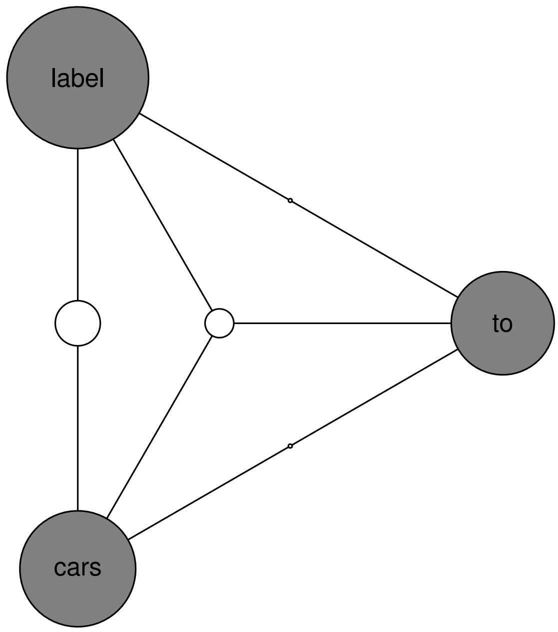
$$\begin{aligned} \mathcal{I}(c, x, y = \mathrm{y}) &= \mathcal{I}(c, x|y = \mathrm{y}) - \mathcal{I}(c, x) && \text{(actual interaction)} \\ \mathcal{I}(c, x, y) &= \mathcal{I}(c, x|y) - \mathcal{I}(c, x) && \text{(expected interaction)} \end{aligned}$$

Expected interaction is symmetric in all three variables: $\mathcal{I}(c, x, y) = \mathcal{I}(c, y, x) = \mathcal{I}(y, x, c)$, etc. A **positive interaction** means that y makes x more informative about c. A **negative interaction** means that y makes x less informative about c. For example, suppose c is tomorrow's weather and (x, y) are weather reports from two different stations. Once you get the report from one station, the other report is less useful to you.

These definitions of information and interaction are conveniently visualized with an **information graph**.







Example: the interaction between the presence of “car” and “cars”

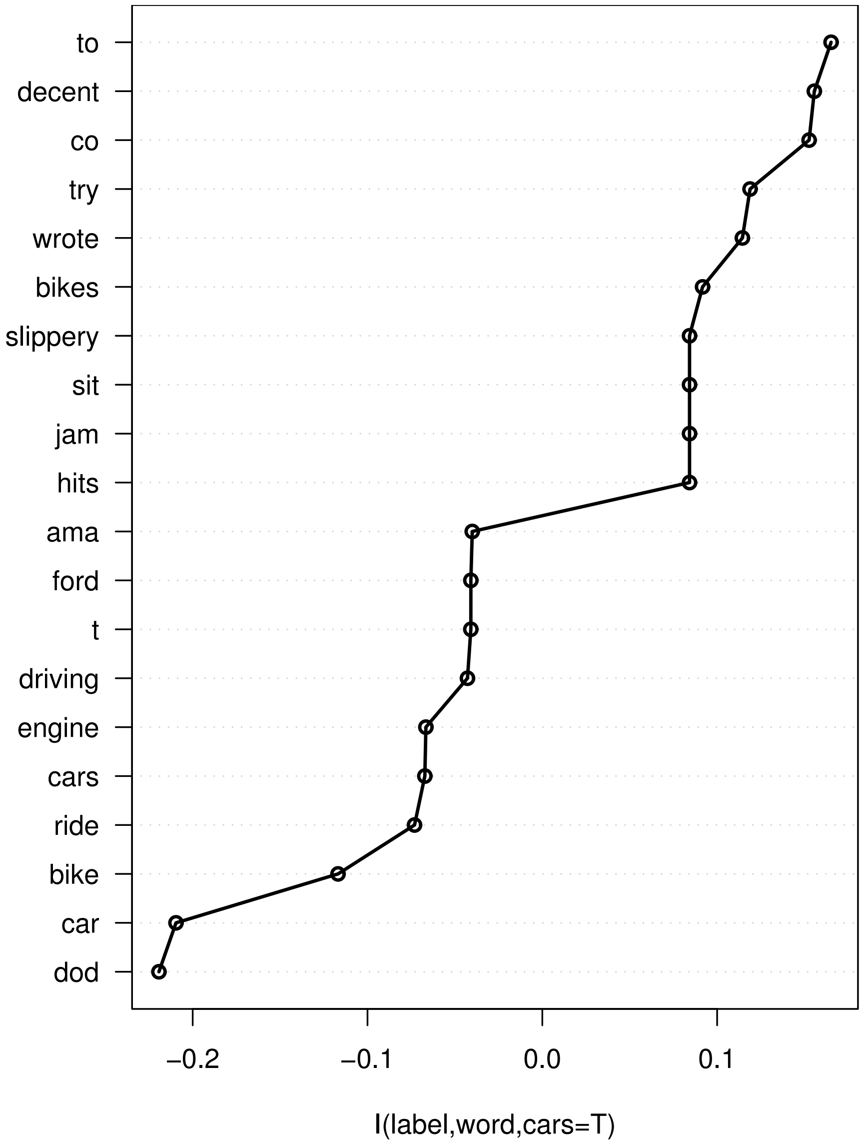
```
, , cars = FALSE

      car
label  FALSE TRUE
  auto 35    34
  moto 93     3

, , cars = TRUE

      car
label  FALSE TRUE
  auto 12    19
  moto  2     2
```

$$
\begin{aligned}
\mathcal{I}(label, car) &= 0.227 \\
\mathcal{I}(label, car|cars = F) &= 0.233 \\
\mathcal{I}(label, car, cars = F) &= 0.006 \qquad \text{(positive interaction)} \\
\mathcal{I}(label, car|cars = T) &= 0.004 \\
\mathcal{I}(label, car, cars = T) &= -0.223 \qquad \text{(negative interaction)} \\
p(cars = T) &= 0.175 \\
\mathcal{I}(label, car, cars) &= -0.034 \qquad \text{(negative on average)}
\end{aligned}
$$

Note that the expected interaction can be zero, even though the actual interactions are nonzero (but of opposite sign).

References

[1] Aleks Jakulin and Ivan Bratko. “Quantifying and Visualizing Attribute Interactions.” `http://ai.fri.uni-lj.si/~aleks/Int/index.html`



to
decent
co
try
wrote
bikes
slippery
sit
jam
hits
ama
ford
t
driving
engine
cars
ride
bike
car
dod
−0.2
−0.1
0.0
0.1
I(label,word,cars=T)