CS 485: Lecture 37

Brendan Rehon

26 April 2006

1 Introduction

Two references:

Kumar, Raghavan, Rajapolan, Tomkins. *Recommendation Systems: A Probabilistic Analysis.*

Jon Kleinberg and Mark Sandler. *Using Mixture Models for Collaborative Filtering.*

We wish to find the probability matrix **A** where each entry gives the probability that a given buyer purchases a given item.

$$\mathit{buyers}\ \underset{\mathbf{m}\times\mathbf{n}}{\overset{\mathit{items}}{\begin{pmatrix} & & \\ & \mathbf{A} & \\ & & \end{pmatrix}}}$$

However, this matrix tends to be too large to update efficiently. So we subdivide the items into categories, and we rewrite **A**:

$$\mathbf{A} = \mathit{buyers}\ \underset{\mathbf{m}\times\mathbf{k}}{\overset{\mathit{categories}}{\begin{pmatrix} & & \\ & \mathbf{P} & \\ & & \end{pmatrix}}}\ \mathit{categories}\ \underset{\mathbf{k}\times\mathbf{n}}{\overset{\mathit{items}}{\begin{pmatrix} & & \\ & \mathbf{W} & \\ & & \end{pmatrix}}}$$

where we use purchases of items as an approximation for **W**.

If I have a rank k matrix $\mathbf{A}$ whose rows sum to 1, can I factor it into $\mathbf{P}$ and $\mathbf{W}$ whose rows sum to 1? Yes. (We cannot guarantee the entries of $\mathbf{P}$ and $\mathbf{W}$ will be in [0,1].)

First we factor $\mathbf{A}$ using spectral value decomposition.

$$\mathbf{A} = \mathbf{U}\sum\mathbf{V} = \underbrace{(\mathbf{U}\sum)}_{\mathbf{P}}\underbrace{\mathbf{V}}_{\mathbf{W}}$$

We normalize the rows of $\mathbf{V}$ to get $\mathbf{W}$. Then we find $\mathbf{P}$ as follows: correct for row normalization of $\mathbf{W}$ by normalizing columns of $\mathbf{U}\sum$.

Replace $\sum_{\ell=1}^{n}\mathbf{A}_{i\ell} = 1$ with $\sum_{\ell=1}^{n}\sum_{j=1}^{k}\mathbf{P}_{ij}\mathbf{W}_{j\ell} = 1$

Then:

$$\sum_{\ell=1}^{n}\sum_{j=1}^{k}\mathbf{P}_{ij}\mathbf{W}_{j\ell} = 1$$
$$\iff \sum_{j}^{k}\mathbf{P}_{ij}\underbrace{\sum_{\ell=1}^{n}\mathbf{W}_{j\ell}}_{1} = 1$$
$$\iff \sum_{j}^{k}\mathbf{P}_{ij} = 1$$

which confirms that if we normalize the columns of $\mathbf{P}$ we normalize the rows of $\mathbf{A}$.

2 Finding Categories

Two cases:

- **(1)** clusters disjoint (where clusters $\equiv$ categories)
- **(2)** clusters can overlap

If we knew the categories and the categories were disjoint, we could calculate $\mathbf{W}$:

$$\mathbf{W} = \textit{category}\overset{\textit{items}}{\begin{pmatrix} \textit{freq. of sales} & \mathbf{0} & \cdots \\ \mathbf{0} & \textit{freq. of sales} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}}$$

if we know the number of items sold (which is just the sum of columns).

Let us define *sample size* = the number of items in a transaction.

We assume the sample size is always two (which is the minimum size we can do something with). To enforce this, in any sales transaction, we just look at the first two items sold.

We have two products i, j. We explore if the following equality is true:

$$Prob(i,j) = Prob(i)Prob(j)$$

If it is not true, then i and j are correlated. That means we might put them in the same category.

3 Correlation Test for Model with Disjoint Clusters (or Categories)

Let:
$\#(i)$ be number of occurrences of i
$\#(i{,}j)$ be the number of occurrences of a pair of items i and j.

$$\textit{frequency of } i = \sum_{\mu=1}^{m} \mathbf{P}_{\mu C(i)} * (\underbrace{\mathbf{W}_{C(i)i}}_{\textit{doesn't depend on user}}) = \mathbf{W}_{C(i)i} \sum_{\mu=1}^{m} \mathbf{P}_{\mu C(i)}$$

$$\textit{frequency of pair } i,j = \sum_{\mu=1}^{m} \mathbf{P}_{\mu C(i)} \mathbf{W}_{C(i)i} \mathbf{P}_{\mu C(j)} \mathbf{W}_{C(j)j} = \mathbf{W}_{C(i)i} \mathbf{W}_{C(j)j} \sum_{\mu=1}^{m} \mathbf{P}_{\mu C(i)} \mathbf{P}_{\mu C(j)}$$

To determine if elements are in the same category consider two vectors

$$(P_{1C(i)}, P_{2C(i)}, \ldots, P_{mC(i)})W_{C(i)i}$$

$$(P_{1C(j)}, P_{2C(j)}, \ldots, P_{mC(j)})W_{C(j)j}$$

What if $C(i)$, $C(j)$ vectors are parallel? If $C(i)$, $C(j)$ vectors are distinguishable, vectors are not close to parallel.

Now we need to form a test (since we don't know $\mathbf{P}$ directly).

3.1 Test To Determine If Vectors are Parallel

If $\#(i)$ and $\#(j)$ are large, then $\#(i,j)$ is close to expected value of $\#(i,j)$.

$$E[\#(i,j)] = \sum_{\mu} \mathbf{P}_{\mu\mathbf{C(i)}} \mathbf{W}_{\mathbf{C(i)i}} \mathbf{P}_{\mu\mathbf{C(j)}} \mathbf{W}_{\mathbf{C(j)j}}$$

Consider

$$\frac{E[\#(i,j)]E[\#(i,j)]}{E[\#(i,i)]E[\#(j,j)]}$$

$$= \frac{\mathbf{W}_{\mathbf{C(i)i}}\mathbf{W}_{\mathbf{C(j)j}}\mathbf{W}_{\mathbf{C(i)i}}\mathbf{W}_{\mathbf{C(j)j}} \sum_{\mu} \mathbf{P}_{\mu\mathbf{C(i)}}\mathbf{P}_{\mu\mathbf{C(j)}} \sum_{\mu} \mathbf{P}_{\mu\mathbf{C(i)}}\mathbf{P}_{\mu\mathbf{C(j)}}}{\mathbf{W}_{\mathbf{C(i)i}}\mathbf{W}_{\mathbf{C(i)i}}\mathbf{W}_{\mathbf{C(j)j}}\mathbf{W}_{\mathbf{C(j)j}} \sum_{\mu} \mathbf{P}_{\mu\mathbf{C(i)}}\mathbf{P}_{\mu\mathbf{C(i)}} \sum_{\mu} \mathbf{P}_{\mu\mathbf{C(j)}}\mathbf{P}_{\mu\mathbf{C(j)}}}$$

$$= \frac{(\mathbf{x}\cdot\mathbf{y})^2}{(\mathbf{x})^2(\mathbf{y})^2} = \cos^2\theta$$

With $\frac{E[\#(i,j)]E[\#(i,j)]}{E[\#(i,i)]E[\#(j,j)]} = \cos^2\theta$ we have an effective test for determining if the two vectors are parallel.