Multimedia Information Retrieval and Technology

Lecture 11 Probabilistic information retrieval

By : Fangyu Wu

Room: SD555

Xi’an Jiaotong-Liverpool University
西交利物浦大学

- Review of basic probability theory
- Classical probabilistic retrieval model
  - Probability ranking principle, etc.
  - Binary independence model (≈ Naïve Bayes text cat)

Review of basic probability theory

The conditional probability *P*(*A*|*B*) expresses the probability of event *A* given that event *B* occurred.

Eg:
P(B)=50%: Lunch at Yushanfang;
P(A|B)=60%: When having lunch at yushanfang, choose noodle over others.

Review of basic probability theory

For two events *A* and *B*, the joint event of both events occurring is described by **the joint probability** *P*(*A*, *B*).

The fundamental relationship between joint and conditional probabilities is given by ***the chain rule***:

$$P(A, B) = P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

Eg:

P(B)=50%: Lunch at Yushanfang;

P(A|B)=60%: When having lunch at yushanfang, choose noodle over others.

What's P(A,B)? What event does it stand for?

Xi'an Jiaotong-Liverpool University
西交利物浦大学

Review of basic probability theory

the *chain rule*:

$$P(A,B) = P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

Again,Eg:

P(B)=50%: Lunch at Yushanfang;

P(A|B)=60%: When having lunch at yushanfang, choose noodle over others.

What event does P(B|A) stand for?

Xi'an Jiaotong-Liverpool University
西交利物浦大学

Review of basic probability theory

$$P(B) = P(A, B) + P(\bar{A}, B)$$

$\bar{A}$, the complement of an event.

The **Bayes' rule** for inverting conditional probabilities:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \left[\frac{P(B|A)}{\sum_{X \in \{A,\bar{A}\}} P(B|X)P(X)}\right] P(A)$$

A way of updating probabilities.

Xi'an Jiaotong-Liverpool University
西交利物浦大学

Review of basic probability theory

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \left[\frac{P(B|A)}{\sum_{X\in\{A,\bar{A}\}} P(B|X)P(X)}\right] P(A)$$

An initial estimate of how likely the event *A* is when we do not have any other information;
This is the ***prior probability P(A)***.

Bayes’ rule lets us derive a ***posterior probability P(A|B)*** after having seen the evidence *B*, based on the *probability* of *B* occurring in the two cases that *A* does or does not hold.

Review of basic probability theory

The odds of an event, provide a kind of multiplier for how probabilities change:

$$O(A) = \frac{P(A)}{P(\bar{A})} = \frac{P(A)}{1 - P(A)}$$

Probabilistic IR topics

- Review of basic probability theory
- Classical probabilistic retrieval model
  - ? Probability ranking principle, etc.
  - ? Binary independence model (≈ Naïve Bayes text cat)

The document ranking problem

We have a collection of documents
User issues a query
A list of documents needs to be returned

Ranking method is the core of an IR system:

We want the "best" document to be first, second best second, etc….

Idea: Rank by probability of relevance of the document w.r.t. information need

$$P(R=1|\text{document}_i, \text{query})$$

The basis of the **Probability Ranking Principle**

The Probability Ranking Principle (PRP)

“If a reference retrieval system’s response to each request is a ranking of the documents in the collection in order of decreasing probability of relevance to the user, where the probabilities are estimated as accurately as possible on the basis of whatever data have been made available to the system, the overall effectiveness of the system will be the best that is obtainable on the basis of those data.”

[1960s/1970s] S. Robertson, W.S. Cooper, M.E. Maron;
van Rijsbergen (1979:113); Manning & Schütze (1999:538)

Probability Ranking Principle

- Let *x* represent a document in the collection.
- Let *R* represent **relevance** of a document w.r.t. the given (fixed) query.
- Let **R=1** represent relevant and **R=0** not relevant**.**

- Need to find p(R=1|x) - probability that a document *x* is **relevant** w.r.t. the given (fixed) query**.**

The Binary Independence Model

PRP in action: Rank all documents by *p*(*R=1*|*x*)

If a set of retrieval results is to be returned, rather than an ordering, the ***Bayes Optimal Decision Rule***, the decision which minimizes the risk of loss, is to simply return documents that are more likely relevant than nonrelevant:

$$d\ is\ relevant\ iff\ P(R=1|d,q) > P(R=0|d,q)$$

Probability Ranking Principle (PRP)

Theorem: Using the PRP is optimal, in the sense that it minimizes the expected loss (Bayes risk) under 1/0 loss.

⍰Provable if all probabilities are correct, etc.

1/0 loss: either returning a nonrelevant document or failing to return a relevant document (such a binary situation where you are evaluated on your *accuracy* is called *1/0 loss*).

- Review of basic probability theory
- Classical probabilistic retrieval model
  - Probability ranking principle, etc.
  - Binary independence model (≈ Naïve Bayes text cat)

The Binary Independence Model

How do we compute all those probabilities?

- Do not know exact probabilities, have to use estimates
- Binary Independence Model (BIM) — the simplest model

The Binary Independence Model

⍰Binary: Boolean model of relevance
Documents and queries are both represented as binary term incidence vectors.

Binary Independence Model

Traditionally used in conjunction with PRP

- **"Binary" = Boolean**: documents are represented as binary incidence vectors of terms :$\vec{x} = (x_1, x_2, \cdots, x_M)$, where $x_t = 1$ if term *t* is present in document *x*, $x_t = 0$ if *t* is not present in document.
- **"Independence":** terms occur in documents independently

Different documents can be modeled as the same vector

The model recognizes no association between terms.

Binary Independence Model

Similarly, query q is represented by the binary term incidence vector $\vec{q}$.

To make a **probabilistic retrieval strategy** precise, we need to estimate how terms in documents contribute to relevance:

- How tf, df, document length, and other statistics influence judgments about document relevance
- How they can be reasonably combined to estimate the probability of document relevance.

Binary Independence Model

- “Relevance” of each document is independent of relevance of other documents.

The Binary Independence Model

Under the BIM, we model the probability $P(R|d,q)$ that a document is relevant via the probability in terms of term incidence vectors $P(R|\vec{x},\vec{q})$.

Using Bayes rule

$$p(R=1|\vec{x},\vec{q}) = \frac{p(\vec{x}|R=1,\vec{q})\,p(R=1|\vec{q})}{p(\vec{x}|\vec{q})}$$

$$p(R=0|\vec{x},\vec{q}) = \frac{p(\vec{x}|R=0,\vec{q})\,p(R=0|\vec{q})}{p(\vec{x}|\vec{q})}$$

The Binary Independence Model

For a given query:

$$p(R=1 \mid x) = \frac{p(x \mid R=1)\,p(R=1)}{p(x)}$$

$$p(R=0 \mid x) = \frac{p(x \mid R=0)\,p(R=0)}{p(x)}$$

- p(R=*1*), p(R=*0*) – Probability of what?
- p(x|R=1), p(x|R=0) - ?

The Binary Independence Model

$$p(R=1\,|\,x)=\frac{p(x\,|\,R=1)\,p(R=1)}{p(x)}$$

$$p(R=0\,|\,x)=\frac{p(x\,|\,R=0)\,p(R=0)}{p(x)}$$

- p(*R=1)*, p(*R=0*) - prior probability of retrieving a relevant or non-relevant document
- p(*x*|*R*=1), p(*x*|*R*=0) - probability that if a relevant (not relevant) document is retrieved, it is *x*.

$$p(R=0\,|\,x)+p(R=1\,|\,x)=1$$

Xi'an Jiaotong-Liverpool University 西交利物浦大学

Probabilistic Retrieval Strategy

How do we compute all these probabilities?

We never know the exact probabilities, and so we have to use **estimates:**

-- Statistics about the actual document collection are used to estimate these probabilities.

Probabilistic Ranking

Basic concept:

“For a given query, if we know some documents that are relevant, terms that occur in those documents should be given greater weighting in searching for other relevant documents.

By making assumptions about the distribution of terms and applying **Bayes Theorem**, it is possible to derive weights theoretically.”

Van Rijsbergen

Binary Independence Model

Under the BIM, given query ***q***, for each document ***d*** need to compute ***p(R|q,d).***

This is replaced with computing ***p(R|*** $\vec{q},\vec{x}$ **)** where $\vec{x}$ is binary term incidence vector presenting ***d.***

Interested only in ranking, we use odds and Bayes' Rule:

$$O(R \mid \vec{q}, \vec{x}) = \frac{p(R=1 \mid \vec{q}, \vec{x})}{p(R=0 \mid \vec{q}, \vec{x})} = \frac{\dfrac{p(R=1 \mid \vec{q})\, p(\vec{x} \mid R=1, \vec{q})}{p(\vec{x} \mid \vec{q})}}{\dfrac{p(R=0 \mid \vec{q})\, p(\vec{x} \mid R=0, \vec{q})}{p(\vec{x} \mid \vec{q})}}$$

Xi'an Jiaotong-Liverpool University
西交利物浦大学

Binary Independence Model

$$O(R|q,\vec{x}) = \frac{p(R=1|q,\vec{x})}{p(R=0|q,\vec{x})} = \frac{p(R=1|q)}{p(R=0|q)} \cdot \frac{p(\vec{x}|R=1,q)}{p(\vec{x}|R=0,q)}$$

Constant for a given query

Needs estimation

How can we estimate the probability of a term incidence vector occurring?

$$\frac{p(\vec{x}|R=1,q)}{p(\vec{x}|R=0,q)}$$

probability that if a relevant (not relevant) document is retrieved, it is *x*

Binary Independence Model

Naïve Bayes Conditional Independence Assumption: the presence or absence of a word in a document is independent of the presence or absence of any other word.

$$\frac{p(\vec{x} \mid R=1,q)}{p(\vec{x} \mid R=0,q)} = \prod_{i=1}^{n} \frac{p(x_i \mid R=1,q)}{p(x_i \mid R=0,q)}$$

So:

$$O(R \mid q,\vec{x}) = O(R \mid q) \cdot \prod_{i=1}^{n} \frac{p(x_i \mid R=1,q)}{p(x_i \mid R=0,q)}$$

Query: Obama health plan
Doc1: Obama rejects allegations about his own bad health
Doc2: The plan is to visit Obama
Doc3: Obama raises concerns with US health plan Reforms

Estimate the probability that the above documents are relevant to the query. Use a contingency table. These are the only three documents in the collection

Maximum Likelihood Estimate (MLE),

For trials with categorical outcomes (such as noting the presence or absence of a term), one way to estimate the probability of an event from data is simply to count the number of times an event occurred divided by the total number of trials.

This is referred to as the ***relative frequency*** of the event.

Estimating the probability as the relative frequency is the ***maximum likelihood estimate*** **(or *MLE*)**, because this value makes the observed data maximally likely.