Week 4: User Profiles and Content Based RSs – *Cont*.

CSX4207/ITX4207: Decision Support and Recommender Systems

Asst. Prof. Dr. Rachsuda Setthawong

Objectives

- To understand text classification methods
- To be familiar with additional algorithms in this approach
- To understand evaluation measures and to be able to apply them
- To understand the limitations of content-based approach

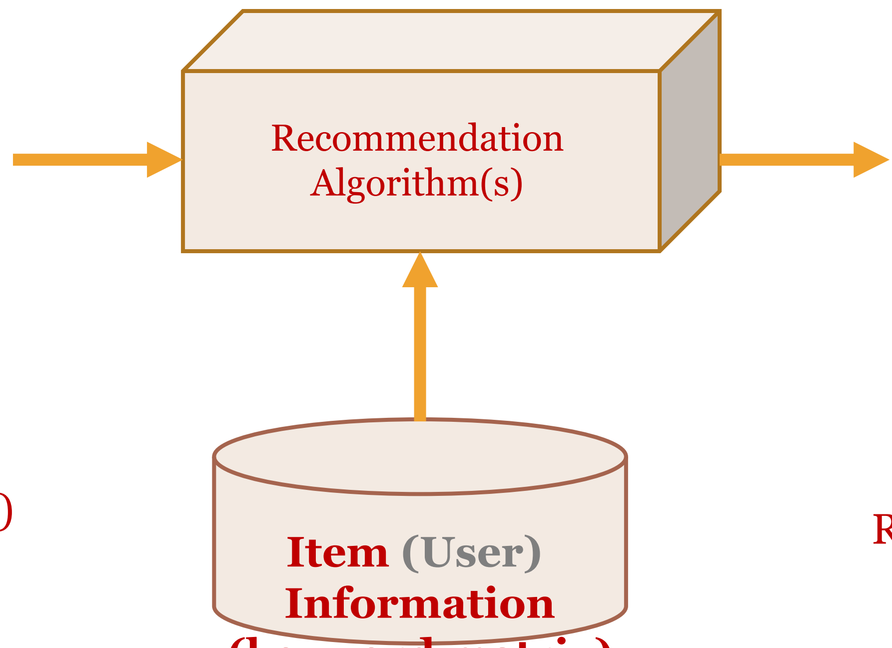
Outlines

- Recommendation as Text Classification Problems
  - Naïve Bayes
  - Additional Probabilistic Methods
- Additional Algorithm
- Limitations of Content-based Recommendation Methods
- Evaluation Measures

Asst. Prof. Dr. Rachsuda Setthawong

How to Generate Recommendation Using Content Based Approach

Target
User's Preferences (keyword vectors)
- Present/Absent (0/1)
- Frequency
- TF-IDF
- etc.



Item (User) **Information (keyword matrix)**
- Present/Absent (0/1)
- Frequency
- TF-IDF
- etc.

Prediction

Rank

Recommendation

Content-based:
"Show me more of the same what I've liked"

Asst. Prof. Dr. Rachsuda Setthawong

Recommendation as Text Classification Problem

- Probabilistic-based approach:
  - Given Boolean representation with 2 classes:
    - ***like*** (class label = 1)
    - ***dislike*** (class label = 0)
  - For each unseen document, calculate probability of ***like*** and ***dislike*** using **Bayes theorem**

$$P(Y|X) = \frac{\prod_{i=1}^{d} P(X_i|Y) \times P(Y)}{P(X)}$$

    - where, Y is is class label,
    - X is the set of attributes of the document, and
    - d is no. of attributes.

(Note: The calculation of P(X) can be omitted.)

  - ACTION: An unseen document should be recommended to the user if P(Label=1|X) > P(Label=0|X).

Asst. Prof. Dr. Rachsuda Setthawong

Bayes Classifier
(A Simplified Model of Bayes Theorem)

- The function that assigns a class label $\hat{y} = C_k$ using the following formula:

$$\hat{y} = \underset{k \in \{1,...,K\}}{\operatorname{argmax}} \; p(C_k) \prod_{i=1}^{n} p(x_i \mid C_k).$$

~ the variable ***Y*** in Bayes theorem

where,

$p(C_k)$ is prior probability (class probability)

$p(x=v \mid C_k)$ is posterior probability

(the probability distribution of v given a class C_k)

https://en.wikipedia.org/wiki/Naive_Bayes_classifier#Constructing_a_classifier_from_the_probability_model

Asst. Prof. Dr. Rachsuda Setthawong

Example 1: should DocID 6 be recommended to the user whose profiles are given below?

(class label = 1) → ***like***
(class label = 0) → ***dislike***

The target user's profile:

Doc-ID	python	amazon	forest	program	**Label**
1	1	1	1	0	1
2	0	0	1	1	0
3	1	1	0	0	1
4	1	0	1	1	1
5	0	0	0	1	0

Unseen Document

Doc-ID	python	amazon	forest	program	**Label**
6	**1**	**1**	**0**	**0**	**?**

Asst. Prof. Dr. Rachsuda Setthawong

Example 1: should DocID 6 be recommended to the user whose profiles are given below? – *Cont*.

The target user's profile:

Doc-ID	python	amazon	forest	program	**Label**
1	1	1	1	0	1
2	0	0	1	1	0
3	1	1	0	0	1
4	1	0	1	1	1
5	0	0	0	1	0

$$P(Y|X) = \frac{\prod_{i=1}^{d} P(X_i|Y) \times \boxed{P(Y)}}{\cancel{P(X)}}$$

ignore

Unseen Document

Doc-ID	python	amazon	forest	program	**Label**
6	**1**	**1**	**0**	**0**	**?**

Prior Probabilites:
P(Label=1) = 3/5
P(Label=0) = 2/5

Asst. Prof. Dr. Rachsuda Setthawong

Example 1: should DocID 6 be recommended to the user whose profiles are given below? – *Cont.*

The target user's profile:

Doc-ID	python	amazon	forest	program	**Label**
1	1	1	1	0	1
2	0	0	1	1	0
3	1	1	0	0	1
4	1	0	1	1	1
5	0	0	0	1	0

$$P(Y|X) = \frac{\prod_{i=1}^{d} \boxed{P(X_i|Y)} \times P(Y)}{\cancel{P(X)}}$$

ignore

Unseen Document

Doc-ID	python	amazon	forest	program	**Label**
6	**1**	**1**	**0**	**0**	**?**

posterior probability :

P(python=1|Label=1) = 3/3
P(amazon=1|Label=1) = 2/3
P(forest=0|Label=1) = 1/3
P(program=0|Label=1) = 2/3

P(python=1|Label=0) = 0/2
P(amazon=1|Label=0) = 0/2
P(forest=0|Label=0) = 1/2
P(program=0|Label=0) = 0/2

Example 1: should DocID 6 be recommended to the user whose profiles are given below? – *Cont.*

The model built from the target user's profile:

Prior Probabilites:
P(Label=1) = 3/5
P(Label=0) = 2/5

posterior probability :

P(python=1|Label=1) = 3/3
P(amazon=1|Label=1) = 2/3
P(forest=0|Label=1) = 1/3
P(program=0|Label=1) = 2/3

P(python=1|Label=0) = 0/2
P(amazon=1|Label=0) = 0/2
P(forest=0|Label=0) = 1/2
P(program=0|Label=0) = 0/2

$$\boxed{P(Y|X)} = \frac{\prod_{i=1}^{d} P(X_i|Y) \times P(Y)}{\cancel{P(X)}}$$

Unseen Document

Doc-ID	python	amazon	forest	program	**Label**
6	**1**	**1**	**0**	**0**	**?**

P(Label=1|X) = [P(python=1|Label=1) × P(amazon=1|Label=1) ×
P(forest=0|Label=1) × P(program=0|Label=1) × P(Label=1)] / (P(X)
= [(3/3 × 2/3 × 1/3 × 2/3) × (3/5)]/P(X)
= (0.149 × 3/5)/P(X) = **0.089/P(X)**

Asst. Prof. Dr. Rachsuda Setthawong

Example 1: should DocID 6 be recommended to the user whose profiles are given below? – *Cont*.

The model built from the target user's profile:

Prior Probabilites:
P(Label=1) = 3/5
P(Label=0) = 2/5

posterior probability :

P(python=1|Label=1) = 3/3
P(amazon=1|Label=1) = 2/3
P(forest=0|Label=1) = 1/3
P(program=0|Label=1) = 2/3

P(python=1|Label=0) = 0/2
P(amazon=1|Label=0) = 0/2
P(forest=0|Label=0) = 1/2
P(program=0|Label=0) = 0/2

$$\boxed{P(Y|X)} = \frac{\prod_{i=1}^{d} P(X_i|Y) \times P(Y)}{\cancel{P(X)}}$$

Unseen Document

Doc-ID	python	amazon	forest	program	**Label**
6	**1**	**1**	**0**	**0**	**?**

P(Label=0|X) = [P(python=1|Label=0) × P(amazon=1|Label=0) × P(forest=0|Label=0) × P(program=0|Label=0) × P(Label=0)] / (P(X)
= [(0/2 × 0/2 × 1/2 × 0/2) × (2/5)]/P(X)
= 0/P(X) = **0**

Asst. Prof. Dr. Rachsuda Setthawong

Example 1: should DocID 6 be recommended to the user whose profiles are given below? – *Cont*.

$$\hat{y} = \boxed{\operatorname{argmax}}_{k \in \{1,...,K\}} p(C_k) \prod_{i=1}^{n} p(x_i \mid C_k).$$

- P(Label=1|X) > P(Label=0|X)

(class label = 1) → ***like***
(class label = 0) → ***dislike***

0.089/P(X) > 0

∴ DocID 6 should be recommended to the user.

Asst. Prof. Dr. Rachsuda Setthawong

Pros of Naïve Bayes

- **Boolean representation** is simple with the assumption of positional independence.
- **The component of the classifier** can be easily updated.
- **The learning time complexity** remains linear to the number of examples.

Cons of Naïve Bayes

- It does **not enforce conditional independence of events**,
  - e.g., “San Fancisco”, “United State of America”
- The **cold start problem** (**new user**’s problem) still exists.
  - **Solution 1:** let the user manually label a set of documents.
  - **Solution 2:** ask the user to provide a list of interest words for each topic category.

Asst. Prof. Dr. Rachsuda Setthawong

Additional Probabilistic Methods

- Alter the posterior probability of the the original Bayes theorem;

$$P(Y|X) = \frac{\prod_{i=1}^{d} \boxed{P(X_i|Y)} \times P(Y)}{P(X)}$$

  - ***Estimate probability of term*** v_i occurring in a document of class C ***by relative frequency*** of v_i in all documents of the class.

$$P(v_i|C = c) = \frac{CountTerms(v_i, docs(c))}{AllTerms(docs(c))}$$

  - Laplace smoothing

$$\hat{P}(v_i|C = c) = \frac{CountTerms(v_i, docs(c)) + 1}{AllTerms(docs(c)) + |V|}$$

CountTerms(v_i, *docs*(*c*)) = **the number of appearances of term** v_i in documents labeled with *c*

AllTerms(*docs*(*c*)) = **the number of all terms** in the documents labeled with *c*

|V| = **the number of distinct terms** appearing in all documents (vocabulary)

Asst. Prof. Dr. Rachsuda Setthawong

Example 2: should DocID 6 be recommended to the user whose profiles are given below?

(class label = 1) → ***like***
(class label = 0) → ***dislike***

The target user's profile:

DocID	Words	Label
1	recommender intelligent recommender	1
2	recommender recommender learning	1
3	recommender school	1
4	teacher homework recommender	0

Unseen Document

DocID	Words	Label
6	recommender recommender recommender teacher homework	?

Asst. Prof. Dr. Rachsuda Setthawong

Example 2: should DocID 6 be recommended to the user whose profiles are given below? – *Cont.*

The target user's profile:

DocID	Words	Label
1	recommender intelligent recommender	1
2	recommender recommender learning	1
3	recommender school	1
4	teacher homework recommender	0

6 distinct terms:

1. recommender
2. intelligent
3. learning
4. school
5. teacher
6. homework

Unseen Document

DocID	Words	Label
6	recommender recommender recommender teacher homework	?

$\hat{P}(\text{recommender}|\text{Label} = 1) = (5+1)/(8+6) = 6/14$

$\hat{P}(\text{homework}|\text{Label} = 1) = (0+1)/(8+6) = 1/14$

$\hat{P}(\text{teacher}|\text{Label} = 1) = (0+1)/(8+6) = 1/14$

$\hat{P}(\text{recommender}|\text{Label} = 0) = (1+1)/(3+6) = 2/9$

$\hat{P}(\text{homework}|\text{Label} = 0) = (1+1)/(3+6) = 2/9$

$\hat{P}(\text{teacher}|\text{Label} = 0) = (1+1)/(3+6) = 2/9$

$P(\text{Label}=1) = 3/4$

$P(\text{Label}=0) = 1/4$

$P(\text{Label}=1|DocID_6) = ?$

$P(\text{Label}=0|DocID_6) = ?$

$$\hat{P}(v_i|C = c) = \frac{CountTerms(v_i, docs(c)) + 1}{AllTerms(docs(c)) + |V|}$$

$$P(Y|X) = \frac{\prod_{i=1}^{d} P(X_i|Y) \times P(Y)}{P(X)}$$

CountTerms(v_i, *docs*(*c*)) = **the number of appearances of term** v_i in documents labeled with *c*

AllTerms(*docs*(*c*)) = **the number of all terms** in the documents labeled with *c*

|V| = **the number of distinct terms** appearing in all documents (vocabulary)

Asst. Prof. Dr. Rachsuda Setthawong

Example 2: should DocID 6 be recommended to the user whose profiles are given below? – *Cont.*

Unseen Document

DocID	Words	Label
6	recommender recommender recommender teacher homework	?

(The word '**recommender**' presents **3 times** in DocID 6.)

$$\hat{P}(v_i|C=c) = \frac{CountTerms(v_i, docs(c)) + 1}{AllTerms(docs(c)) + |V|}$$

$$P(Y|X) = \frac{\prod_{i=1}^{d} \boxed{P(X_i|Y)} \times P(Y)}{\cancel{P(X)}}$$ ignore

$\hat{P}(\text{recommender}|\text{Label} = 1) = 3/7$

$\hat{P}(\text{homework}|\text{Label} = 1) = 1/14$

$\hat{P}(\text{teacher}|\text{Label} = 1) = 1/14$

$\hat{P}(\text{recommender}|\text{Label} = 0) = 2/9$

$\hat{P}(\text{homework}|\text{Label} = 0) = 2/9$

$\hat{P}(\text{teacher}|\text{Label} = 0) = 2/9$

P(Label=1) = 3/4   P(Label=0) = 1/4

$P(\text{Label}=1|v_1 \ldots v_n) = [(3/7)^{\boxed{3}} * 1/14 * 1/14 * \boxed{3/4}] / P(v_1 \ldots v_n) \approx \mathbf{0.0003} / P(v_1 \ldots v_n)$

$P(\text{Label}=0|v_1 \ldots v_n) = [(2/9)^{3} * 2/9 * 2/9 * 1/4] / P(v_1 \ldots v_n) \approx 0.0001 / P(v_1 \ldots v_n)$

$P(\text{Label}=1|v_1 \ldots v_n) > P(\text{Label}=0|v_1 \ldots v_n)$

∴ DocID 6 should be recommended to the user.

Useful References: https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/

Outlines

- Recommendation as Text Classification Problems
  - Naïve Bayes
  - Additional Probabilistic Methods
- Additional Algorithm
- Limitations of Content-based Recommendation Methods
- Evaluation Measures

Additional Algorithm

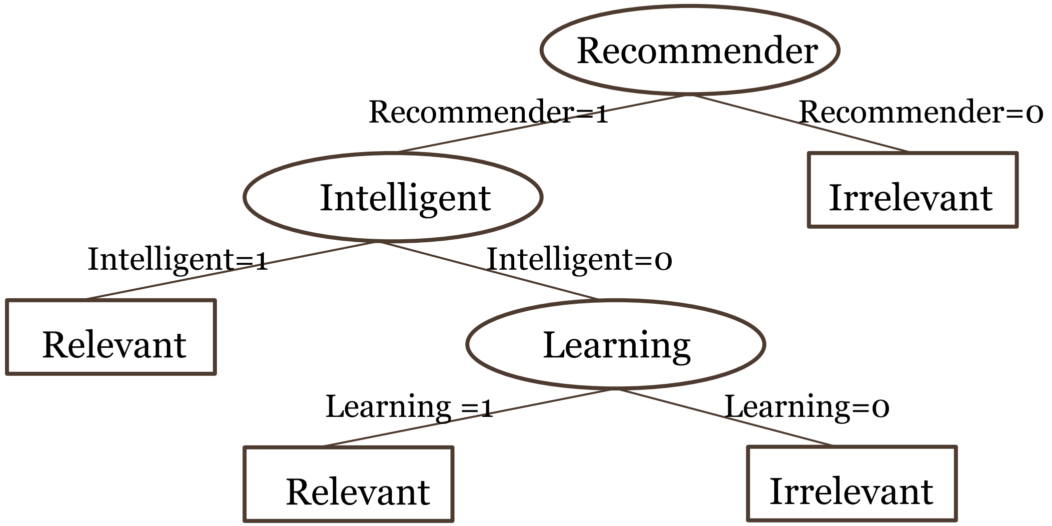
- Decision Tree

Applying Decision Trees to the Recommendation Problem

Asst. Prof. Dr. Rachsuda Setthawong

- Basic idea:
  - Construct a tree whose inner nodes are keywords in documents (item features).
  - Tree's branches (edges) relates to ***present*** (= 0) or ***absent*** (= 1) of related keywords (inner nodes).
  - Leave nodes relates to recommendations (relevant/irrelevant).

An Example of User Profile's Tree

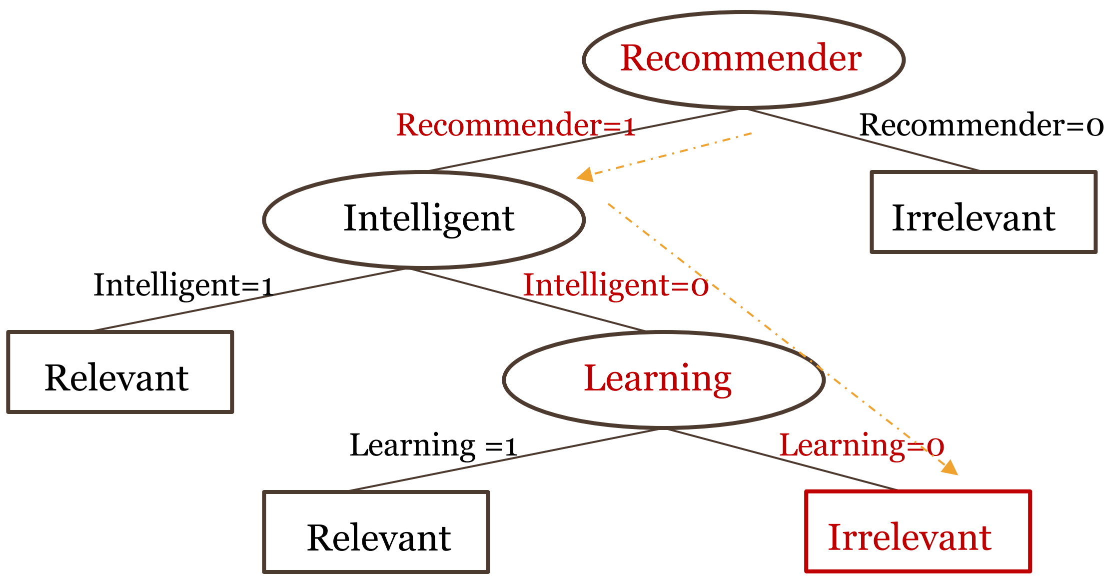


Asst. Prof. Dr. Rachsuda Setthawong

Generating a Recommendation:
Traverse the User Profile's Tree Using an UnSeen Item

Unseen Document

DocID	Words	Label
6	recommender recommender recommender teacher homework	?



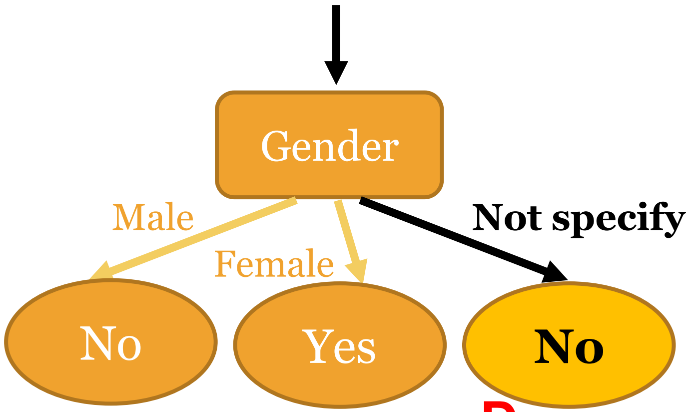
Asst. Prof. Dr. Rachsuda Setthawong

Constructing A Decision Tree Using Hunt's Algorithm

- Let D_t be the set of training records that reach a node t
- General Procedure:
  - If D_t is an empty set, then t is a leaf node labeled by the default class, y_d
    Recursively apply the procedure to each subset.
  - If D_t contains records that belong the same class y_t, then t is a leaf node labeled as y_t
  - If D_t contains records that belong to more than one class, use an attribute test to split the data into smaller subsets.

  Recursively apply the procedure to each subset.

Age	Gender	PlaySport	PassCourse?
16	Male	No	**No**
17	Female	No	Yes
20	Male	Yes	**No**
20	Male	No	**No**



D_t of Gender = 'Not Specify' is empty.

Age	Gender	Play-Sport	Pass-Course?
16	Male	No	No
20	Male	Yes	No
20	Male	No	No

Age	Gender	Play-Sport	Pass-Course?
17	Female	No	Yes

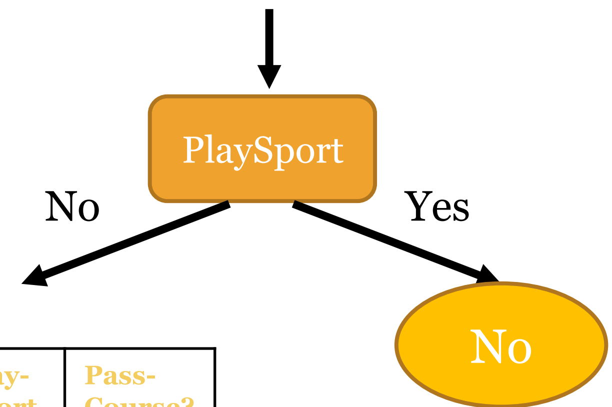
Asst. Prof. Dr. Rachsuda Setthawong

Constructing A Decision Tree Using Hunt's Algorithm

- Let D_t be the set of training records that reach a node t

- General Procedure:
  - If D_t is an empty set, then t is a leaf node labeled by the default class, y_d
  - If D_t contains records that belong the same class y_t, then t is a leaf node labeled as y_t
  - If D_t contains records that belong to more than one class, use an attribute test to split the data into smaller subsets.

Age	Gender	PlaySport	PassCourse?
16	Male	**No**	No
17	Female	**No**	Yes
20	Male	**Yes**	No
20	Male	**No**	No



Recursively apply the procedure to each subset.

Age	Gender	Play-Sport	Pass-Course?
16	Male	No	No
17	Female	No	Yes
20	Male	No	No

D_t of 'PlaySport = 'Yes'

Age	Gender	Play-Sport	Pass-Course?
20	Male	Yes	**No**

Constructing A Decision Tree Using Hunt's Algorithm

- Let D_t be the set of training records that reach a node t

- General Procedure:
  - If D_t is an empty set, then t is a leaf node labeled by the default class, y_d
  - If D_t contains records that belong the same class y_t, then t is a leaf node labeled as y_t
  - If D_t contains records that belong to more than one class, use an attribute test (e.g., Gender) to split the data into smaller subsets.

  Recursively apply the procedure to each subset.

Age	Gender	PlaySport	PassCourse?
16	Male	No	**No**
17	Female	No	Yes
20	Male	Yes	**No**
20	Male	No	**No**

PlaySport

No

$D_{t\ of}$ 'PlaySport = 'No'

Age	Gender	Play-Sport	Pass-Course?
16	Male	No	**No**
17	Female	No	**Yes**
20	Male	No	**No**

??

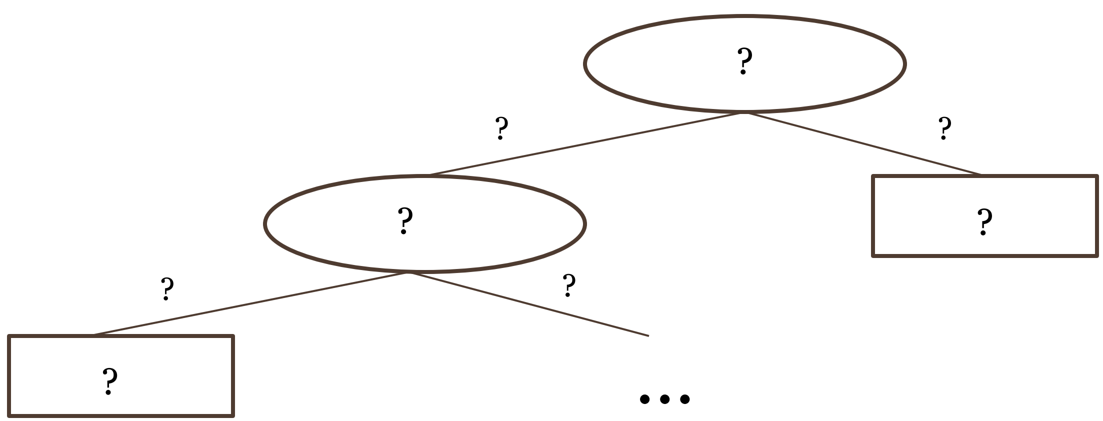
Yes

Age	Gender	Play-Sport	Pass-Course?
20	Male	Yes	No

Asst. Prof. Dr. Rachsuda Setthawong

Let’s Construct Decision Tree

DocID	Words	Label
1	computer java programming computing	1
2	computer programming python programming	1
3	python amazon rainforest python	0
4	python anaconda rainforest rainforest	0
5	program computer programming	1



Remark about A Decision Tree Constructing

- We didn't discuss about how to select the best split in this course.
- For further reading, see also
  - https://en.wikipedia.org/wiki/Decision_tree_learning

Suggestions on Decision Tree Learners

- A decision tree model performs well when the number of features is relatively small.
- The model is more appropriate to apply on "meta" features
  - E.g., author name, and genre than on TF-IDF calculations.

Asst. Prof. Dr. Rachsuda Setthawong

Outlines

- Recommendation as Text Classification Problems
  - Naïve Bayes
  - Additional Probabilistic Methods
- Additional Algorithm
- Limitations of Content-based Recommendation Methods
- Evaluation Measures

Asst. Prof. Dr. Rachsuda Setthawong

Limitations of Content-based Recommendation Methods

- Limit judgement on quality/relevance of a document (item)
  - It cannot judge up-to-date-ness, utility, and beauty
- Content's size and extraction issue
  - Too short; feature vector's sparsity
  - Not be automatically extractable (e.g., multimedia).
- A need of ramp-up phase
  - Need some training data for its initialization.
- Overspecialization
  - Suggest "more of the same“ items.

Asst. Prof. Dr. Rachsuda Setthawong

Outlines

- Recommendation as Text Classification Problems
  - Naïve Bayes
  - Additional Probabilistic Methods
- Additional Algorithm
- Limitations of Content-based Recommendation Methods
- Evaluation Measures

Asst. Prof. Dr. Rachsuda Setthawong

Basic Evaluation Measures

- Ground truth is given.
  - Accuracy
  - Precision
  - Recall
  - F-measure
- Otherwise,
  - Real users's evaluation (questionnaires, survey, etc.)

Asst. Prof. Dr. Rachsuda Setthawong

Accuracy

Predicted Answer

Actual Answer

	isRelevant=Yes	isRelevant=No
isRelevant=Yes	a	b
isRelevant=No	c	d

$$\text{Accuracy} = \frac{a+d}{a+b+c+d} = \frac{TP+TN}{TP+TN+FP+FN}$$

Asst. Prof. Dr. Rachsuda Setthawong

Precision, Recall and F-measure

p

Predicted Answer

Actual Answer		isRelevant=Yes	isRelevant=No
r	isRelevant=Yes	a	b
	isRelevant=No	c	d

$$\text{Precision (p)} = \frac{a}{a+c}$$

$$\text{Recall (r)} = \frac{a}{a+b}$$

$$\text{F-measure (F)} = \frac{2rp}{r+p} = \frac{2a}{2a+b+c}$$

Asst. Prof. Dr. Rachsuda Setthawong

Example 3:

User ID	Items	Prediction (Recommendation)	True Answer
1	1	Relevant	Relevant
1	2	Relevant	Relevant
1	3	Not Relevant	Not Relevant
1	4	Relevant	Not Relevant
1	5	Not Relevant	Not Relevant
1	6	Relevant	Relevant
1	7	Relevant	Relevant
1	8	Not Relevant	Relevant
1	9	Not Relevant	Not Relevant
1	10	Not Relevant	Relevant

Predicted Answer

Actual Answer	isRelevant= Yes	isRelevant= No
isRelevant=Yes	4 (a) TP	2 (b) FN
isRelevant=No	1 (c) FP	3 (d) TN

Asst. Prof. Dr. Rachsuda Setthawong

Example 3:

User ID	Items	Prediction (Recommendation)	True Answer
1	1	Relevant	Relevant
1	2	Relevant	Relevant
1	3	Not Relevant	Not Relevant
1	4	Relevant	Not Relevant
1	5	Not Relevant	Not Relevant
1	6	Relevant	Relevant
1	7	Relevant	Relevant
1	8	Not Relevant	Relevant
1	9	Not Relevant	Not Relevant
1	10	Not Relevant	Relevant

Predicted Answer

Actual Answer

	isRelevant= Yes	isRelevant= No
isRelevant=Yes	4 (a) TP	2 (b) FN
isRelevant=No	1 (c) FP	3 (d) TN

Asst. Prof. Dr. Rachsuda Setthawong

Example 3:

User ID	Items	Prediction (Recommendation)	True Answer
1	1	Relevant	Relevant
1	2	Relevant	Relevant
1	3	Not Relevant	Not Relevant
1	4	Relevant	Not Relevant
1	5	Not Relevant	Not Relevant
1	6	Relevant	Relevant
1	7	Relevant	Relevant
1	8	Not Relevant	Relevant
1	9	Not Relevant	Not Relevant
1	10	Not Relevant	Relevant

Predicted Answer

Actual Answer

	isRelevant= Yes	isRelevant= No
isRelevant=Yes	4 (a) TP	2 (b) FN
isRelevant=No	1 (c) FP	3 (d) TN

Asst. Prof. Dr. Rachsuda Setthawong

Example 3:

User ID	Items	Prediction (Recommendation)	True Answer
1	1	Relevant	Relevant
1	2	Relevant	Relevant
1	3	Not Relevant	Not Relevant
1	4	Relevant	Not Relevant
1	5	Not Relevant	Not Relevant
1	6	Relevant	Relevant
1	7	Relevant	Relevant
1	8	Not Relevant	Relevant
1	9	Not Relevant	Not Relevant
1	10	Not Relevant	Relevant

Predicted Answer

Actual Answer

	isRelevant= Yes	isRelevant= No
isRelevant=Yes	4 (a) TP	2 (b) FN
isRelevant=No	1 (c) FP	3 (d) TN

Example 3: Accuracy

Predicted Answer

Actual Answer		isRelevant= Yes	isRelevant= No
	isRelevant=Yes	4 (a) TP	2 (b) FN
	isRelevant=No	1 (c) FP	3 (d) TN

$$\text{Accuracy} = \frac{a+d}{a+b+c+d} = \frac{TP+TN}{TP+TN+FP+FN}$$

= (4+3)/(4+2+1+3)

= 0.7

Asst. Prof. Dr. Rachsuda Setthawong

Example 3: Precision

	Predicted Answer: isRelevant= Yes	Predicted Answer: isRelevant= No
Actual Answer: isRelevant=Yes	4 (a) TP	2 (b) FN
Actual Answer: isRelevant=No	1 (c) FP	3 (d) TN

$$\text{Precision (p)} = \frac{a}{a+c}$$

$= 4 / (4+1)$

$= 0.8$

Asst. Prof. Dr. Rachsuda Setthawong

Example 3: Recall

Predicted Answer

Actual Answer	isRelevant=Yes	isRelevant=No
isRelevant=Yes	4 (a) TP	2 (b) FN
isRelevant=No	1 (c) FP	3 (d) TN

$$\text{Recall (r)} = \frac{a}{a+b}$$

$= 4 / (4+2)$

$= 0.67$

Asst. Prof. Dr. Rachsuda Setthawong

Example 3: F-measure (F)

Predicted Answer

Actual Answer

	isRelevant= Yes	isRelevant= No
isRelevant=Yes	4 (a)	2 (b)
isRelevant=No	1 (c)	3 (d)

$$\text{F - measure (F)} = \frac{2rp}{r+p} = \frac{2a}{2a+b+c}$$

= (2*4) / [(2*4)+2+1)

= 0.73