COMP3702 / COMP7702 Artificial Intelligence

Review Lecture

COMP4702/COMP7703 - Machine Learning

- www.itee.uq.edu.au/~comp4702/

Competition

- 26 working classifiers
  - All entered classifiers worked!
- 50 letters
- 1 bonus mark to all entrants
- 1 bonus mark to top 4 classifiers

Review of Lecture Material

1. Introduction to AI
2. Uninformed Search
3. Informed Search
4. Adversarial Search
5. Uncertain Knowledge
6. Principles of Machine Learning
7. Symbolic Machine Learning
8. Statistical Machine Learning
9. Neural Networks
10. Natural Language Processing and Semantic Modelling
11. Language and Robots

Artificial Intelligence (1)

- How can we discuss artificial intelligence if we don't know what intelligence is?
- Russell and Norvig list definitions based on either a human/cognitive quality basis or an objective measure of rationality

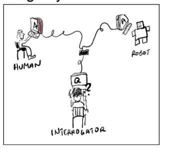
"The exciting new effort to make computers think ... *machines with minds*, in the full and literal sense" (Haugeland, 1985) "The automation of activities that we associate with human thinking, activities such as decision-making, problem solving, learning ..." (Bellman, 1978)	"The study of mental faculties through the use of computational models" (Charniak and McDermott, 1985) "The study of the computations that make it possible to perceive, reason, and act" (Winston, 1992)
"The art of creating machines that perform functions that require intelligence when performed by people" (Kurzweil, 1990) "The study of how to make computers do things at which, at the moment, people are better" (Rich and Knight, 1991)	"A field of study that seeks to explain and emulate intelligent behavior in terms of computational processes" (Schalkoff, 1990) "The branch of computer science that is concerned with the automation of intelligent behavior" (Luger and Stubblefield, 1993)

- Russell and Norvig identify four possible goals of definitions (implicitly exemplified in the table above)

Systems that think like humans.	Systems that think rationally.
Systems that act like humans	Systems that act rationally

Artificial Intelligence (2)

- Turing test
  - Operational test of intelligent behaviour with an interrogator, a computer, and a person
  - Agent requires knowledge, reasoning, language, and learning
  - Positive: standard test, bypasses the true nature of intelligence, removes bias
  - Negative: focus on symbolic tasks, compares machine with 'human' intelligence, far too restrictive, difficult to do in practice
- Extension: Total Turing Test
  - Agent also requires computer vision to analyse a video signal and robotics for manipulating objects



Artificial Intelligence (3)

- Physical Symbol System Hypothesis
- (Simon and Newell, 1976)
- *A physical symbol system has the necessary and sufficient means for intelligent action.*
- A system:
  – Consists of a set of entities, called symbols,
  – Contains a collection of these symbol structures
  – Contains a collection of processes that operate on expressions to produce other expressions: processes of creation, modification, reproduction and destruction

Artificial Intelligence (4)

- Searle's thought experiment
  – Monolingual English speaker in a room with a set of rules to correlate Chinese characters input to a set of Chinese characters to output

Artificial Intelligence (5)

- Strong AI
  – Duplication of intelligence
  – Aims to understand intelligence
- Weak AI
  – Simulation of intelligence
  – Aims to make computers more 'useful'

Artificial Intelligence (6)

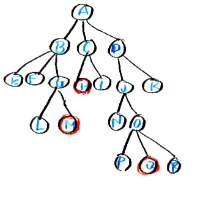
- Agents:
  – Simple reflex
  – Model-based reflex
  – Goal-based
  – Utility based
- Design of each type of agent, and progress from sensor inputs to actuator outputs

Uninformed Search (1)

- Problem formulation
  – Representation
  – Initial state
  – Goal test
  – Example rules

Uninformed Search (2)

- Algorithms:
  – Breadth first
  – Depth first
  – Depth limited
  – Iterative deepening
- General principles:
  – Search queue/fringe
  – Node expansion
- Performance measures:
  – Time and space complexity
  – Optimality and completeness



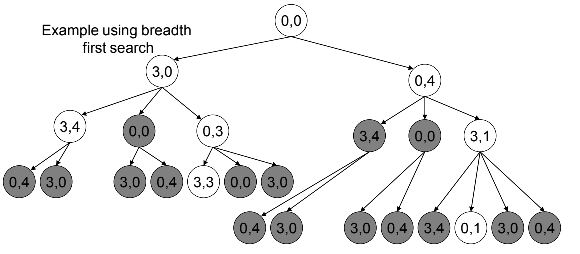
Uninformed Search (3)

- Water jug actions

Actions	*Constraints*	*Meaning*
a1 (x,y) → (3,y)		fill up jug3L
a2 (x,y) → (x,4)		fill up jug4L
a3 (x,y) → (0,y)		empty jug3L
a4 (x,y) → (x,0)		empty jug4L
a5 (x,y) → (0,x+y)	[0 ≤ x+y ≤ 4]	pour all of jug3L into jug4L
a6 (x,y) → (x+y,0)	[0 ≤ x+y ≤ 3]	pour all of jug4L into jug3L
a7 (x,y) → (x+y-4,4)	[x+y > 4]	fill up jug4L from jug3L
a8 (x,y) → (3,x+y-3)	[x+y > 3]	fill up jug3L from jug4L

Uninformed Search (4)

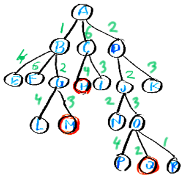
- Repeated States
  - Failure to detect repeated states can turn a linear problem into an exponential one



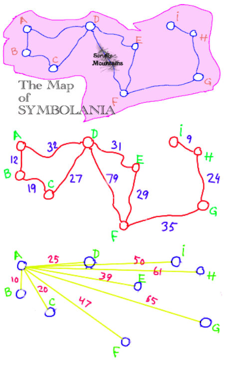
Informed Search (1)

- Algorithms:
  - Greedy search
  - A*

- Heuristic function, admissibility, dominance
- General principles
  - Search queue/fringe
  - Node expansion
- Performance measures:
  - Time and space complexity
  - Optimality and completeness



Informed Search (2)



- The country of Symbolania has 9 major cities, neatly named after the first 9 letters in the alphabet (red).
- There is a network of roads connecting the cities. The distances are provided in blue. Note that the Scruffy Mountains make the road from F to D much longer than expected.
- We need to find our way by car from our present location to the City of A (A has an airport that will take us home). From a birds-eye view the distances are as in the illustration (red numbers). This is the kind of information that we have available with respect to the goal at all states (assuming that we know the coordinates of A and the city that we're in).
- So, our heuristic function, h(), returns the Euclidean distance as shown above, e.g. h(F)=47, h(H)=61, etc.

Informed Search (3)

- Local Search
  - Hill climbing
    - Take actions that improve the current state (greedy local search)
  - Simulated annealing
    - Select actions stochastically (may make things worse temporarily) according to a decreasing "temperature"
  - Genetic algorithms
    - Search using a population of states, in which each state is subjected to "evolutionary selection" and "evolutionary processes" to create a new (more "fit") population of states

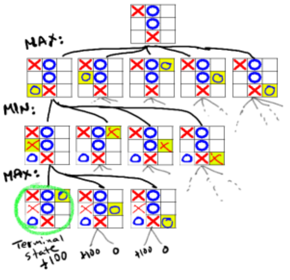
Adversarial Search (1)

Game Search Problem

- Initial state
  - Board position and player
- Successor function
  - Returns a list of (action, state) pairs, indicating a legal move and resulting state
- Terminal test
  - Determines when the game is over
- Utility function
  - Numeric value for a terminal state representing its utility for a given player

Adversarial Search (2)

- Mini-max algorithm
- Alpha-beta pruning
- Using heuristic functions to make imperfect decisions



- Alternating evaluations
- Depth first
- Pruning using alpha-beta
- Multiplayer?

Uncertain Knowledge (1)

- Probability theory
- Prior probability
- Conditional (posterior) probability
- Independence
- Conditional independence
- Probabilistic inference
- Bayes' rule

Uncertain Knowledge (2)

- Probabilistic Inference:
- There is a disease X that affects 1%.
- Doctor A examines everything above your waist. In diagnosing X, Dr A misclassifies 25% who are sick, and 10% who are healthy.
- If Dr A diagnoses you as having X, what is the chance that you actually have it?

$P(X) = 0.01$

$P(ADiagX \mid X) = 0.75$
$P(ADiagX \mid \neg X) = 0.10$

$$P(X \mid ADiagX) = \frac{P(ADiagX \mid X)P(X)}{P(ADiagX)}$$

$$P(X \mid ADiagX) = \frac{0.75 \cdot 0.01}{0.1065} = 0.07$$

$P(ADiagX) = P(ADiagX \mid X)P(X) + P(ADiagX \mid \neg X)P(\neg X)$

$P(ADiagX) = 0.75 \cdot 0.01 + 0.10 \cdot 0.99 = 0.0075 + 0.099 = 0.1065$

Uncertain Knowledge (3)

- Probabilistic Inference:
- Doctor B examines everything below your waist. In diagnosing X, Dr B misclassifies 10% who are sick, and 20% who are healthy.
- If Dr B diagnoses you as having X, what is the chance that you have it?

$P(X) = 0.01$

$P(BDiagX \mid X) = 0.90$
$P(BDiagX \mid \neg X) = 0.20$

$$P(X \mid ADiagX) = \frac{P(ADiagX \mid X)P(X)}{P(ADiagX)}$$

$$P(X \mid BDiagX) = \frac{0.90 \cdot 0.01}{0.207} = 0.04$$

$P(ADiagX) = P(ADiagX \mid X)P(X) + P(ADiagX \mid \neg X)P(\neg X)$

$P(BDiagX) = 0.90 \cdot 0.01 + 0.20 \cdot 0.99 = 0.009 + 0.198 = 0.207$

Uncertain Knowledge (4)

- Probabilistic Inference:
- Due to their disparate methods, we regard Dr A and Dr B's diagnoses as independent (note: not in regard to X).
- What is the chance of having X if both Dr A and Dr B diagnoses you as having X?

$P(A \mid B)=P(A)$
$P(A \wedge B)=P(A)\,P(B)$

$$P(X \mid ADiagX \wedge BDiagX) = \frac{P(ADiagX \wedge BDiagX \mid X)P(X)}{P(ADiagX \wedge BDiagX)}$$

$$P(X \mid ADiagX \wedge BDiagX) = \frac{0.675 \cdot 0.01}{0.022} = 0.307$$

$P(ADiagX \mid X)=0.75$
$P(-ADiagX \mid X)=0.25$
$P(BDiagX \mid X)=0.90$
$P(-BDiagX \mid X)=0.10$

$P(ADiagX \wedge BDiagX \mid X) = P(ADiagX \mid X)P(BDiagX \mid X)$

$P(ADiagX \wedge BDiagX \mid X) = 0.75 \cdot 0.90 = 0.675$

$P(ADiagX \wedge BDiagX) = P(ADiagX)P(BDiagX)$

$P(ADiagX \wedge BDiagX) = 0.1065 \cdot 0.207 = 0.022$

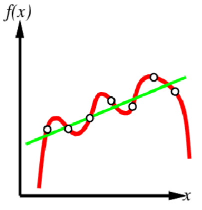
Principles of Machine Learning (1)

- Techniques – how do they work?
- Methodology – how should they be used?

- Occam's razor

Principles of Machine Learning (2)

- Learning Problem:
- Examples
- Supervised, reinforcement, or unsupervised?
- Representation
- Training and forming a hypothesis
- Testing and generalization



Principles of Machine Learning (3)

- Learning Models / Algorithms
  - – Current-best-hypothesis
  - – Decision trees (ID3)
  - – Bayesian learning and Naïve Bayes'
  - – Neural network
- Aspects:
  - – Representation
  - – Process

Principles of Machine Learning (4)

- Searching hypothesis space
  - – False positives
  - – False negatives
  - – True positives
  - – True negatives
- Information Theory
  - – Information content
  - – Gain
  - – Remainder

Principles of Machine Learning (5)

- Assessing the performance of a supervised learning algorithm
  - – Training and test sets

Symbolic Machine Learning (1)

Current Best Learning

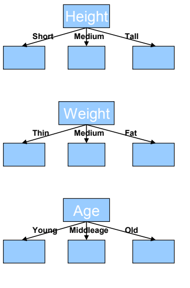
- Iteratively refining hypothesis by traversing a concept space (various types of representations)
- False negative -> Generalization
- False positive -> Specialization

  - colour(ball, red) Generalization Specialization
    colour(X, red)
  - shape(X, round) ∧ size(X, small) ∧ colour(X, red)
    shape(X, round) ∧ colour(X, red)
  - shape(X, round) ∧ size(X, small)
    shape(X, round) ∧ size(X, small) ∨ size(X, large)

Symbolic Machine Learning (2)

Decision Trees: ID3

	Height	Weight	Age	Gender	Sick?
1	Tall	Fat	Young	Female	Yes
2	Tall	Thin	Middleage	Male	Yes
3	Short	Medium	Old	Male	No
4	Medium	Medium	Old	Female	No
5	Medium	Fat	Young	Male	No
6	Tall	Thin	Young	Female	Yes
7	Short	Medium	Middleage	Male	No
8	Medium	Fat	Young	Female	No
9	Tall	Thin	Old	Female	Yes
10	Tall	Thin	Young	Female	Yes
11	Short	Medium	Middleage	Male	No
12	Tall	Medium	Young	Male	Yes
13	Tall	Fat	Young	Female	Yes
14	Short	Thin	Old	Male	No
15	Medium	Thin	Old	Female	Yes
16	Tall	Fat	Young	Female	Yes
17	Tall	Thin	Middleage	Male	Yes
18	Short	Thin	Young	Female	Yes
19	Medium	Fat	Old	Female	No
20	Tall	Thin	Young	Male	Yes



Symbolic Machine Learning (3)

Decision Trees: ID3

- choosing top attribute

$$Gain(A) = I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) - Remainder(A)$$

1st term is the same for all attributes, so rank attributes based on remainder only.

$$Remainder(A) = \sum_{i=1}^{m} \frac{p_i + n_i}{p+n} I\left(\frac{p_i}{p_i+n_i}, \frac{n_i}{p_i+n_i}\right)$$

m = number of possible values of attribute

Symbolic Machine Learning (4)

Decision Trees: ID3

	Height	Weight	Age	Gender	Sick?
1	Tall	Fat	Young	Female	Yes
2	Tall	Thin	Middleage	Male	Yes
3	Short	Medium	Old	Male	No
4	Medium	Medium	Old	Female	No
5	Medium	Fat	Young	Male	No
6	Tall	Thin	Young	Female	Yes
7	Short	Medium	Middleage	Male	No
8	Medium	Fat	Young	Female	No
9	Tall	Thin	Old	Female	Yes
10	Tall	Thin	Young	Female	Yes
11	Short	Medium	Middleage	Male	No
12	Tall	Medium	Young	Male	Yes
13	Tall	Fat	Young	Female	Yes
14	Short	Thin	Old	Male	No
15	Medium	Thin	Old	Female	Yes
16	Tall	Fat	Young	Female	Yes
17	Tall	Thin	Middleage	Male	Yes
18	Short	Thin	Young	Female	Yes
19	Medium	Fat	Old	Female	No
20	Tall	Thin	Young	Male	Yes

- E.g. Height attribute – has 3 possible values: tall, medium, short.

$$Remainder(Height) =$$

$$\frac{p_{tall}+n_{tall}}{p+n} I\left(\frac{p_{tall}}{p_{tall}+n_{tall}}, \frac{n_{tall}}{p_{tall}+n_{tall}}\right)$$

$$+ \frac{p_{medium}+n_{medium}}{p+n} I\left(\frac{p_{medium}}{p_{medium}+n_{medium}}, \frac{n_{medium}}{p_{medium}+n_{medium}}\right)$$

$$+ \frac{p_{short}+n_{short}}{p+n} I\left(\frac{p_{short}}{p_{short}+n_{short}}, \frac{n_{short}}{p_{short}+n_{short}}\right),$$

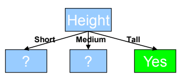
where e.g. $\frac{p_{short}}{p_{short}+n_{short}}$ = estimate of prob(sick | short), here = 1/5,

so $\frac{n_{short}}{p_{short}+n_{short}} = 4/5.$

Symbolic Machine Learning (5)

Decision Trees: ID3

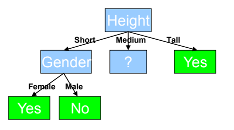
	Height	Weight	Age	Gender	Sick?
1	Tall	Fat	Young	Female	Yes
2	Tall	Thin	Middleage	Male	Yes
3	Short	Medium	Old	Male	No
4	Medium	Medium	Old	Female	No
5	Medium	Fat	Young	Male	No
6	Tall	Thin	Young	Female	Yes
7	Short	Medium	Middleage	Male	No
8	Medium	Fat	Young	Female	No
9	Tall	Thin	Old	Female	Yes
10	Tall	Thin	Young	Female	Yes
11	Short	Medium	Middleage	Male	No
12	Tall	Medium	Young	Male	Yes
13	Tall	Fat	Young	Female	Yes
14	Short	Thin	Old	Male	No
15	Medium	Thin	Old	Female	Yes
16	Tall	Fat	Young	Female	Yes
17	Tall	Thin	Middleage	Male	Yes
18	Short	Thin	Young	Female	Yes
19	Medium	Fat	Old	Female	No
20	Tall	Thin	Young	Male	Yes



Symbolic Machine Learning (6)

Decision Trees: ID3

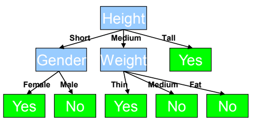
	Height	Weight	Age	Gender	Sick?
1	Tall	Fat	Young	Female	Yes
2	Tall	Thin	Middleage	Male	Yes
3	Short	Medium	Old	Male	No
4	Medium	Medium	Old	Female	No
5	Medium	Fat	Young	Male	No
6	Tall	Thin	Young	Female	Yes
7	Short	Medium	Middleage	Male	No
8	Medium	Fat	Young	Female	No
9	Tall	Thin	Old	Female	Yes
10	Tall	Thin	Young	Female	Yes
11	Short	Medium	Middleage	Male	No
12	Tall	Medium	Young	Male	Yes
13	Tall	Fat	Young	Female	Yes
14	Short	Thin	Old	Male	No
15	Medium	Thin	Old	Female	Yes
16	Tall	Fat	Young	Female	Yes
17	Tall	Thin	Middleage	Male	Yes
18	Short	Thin	Young	Female	Yes
19	Medium	Fat	Old	Female	No
20	Tall	Thin	Young	Male	Yes



Symbolic Machine Learning (7)

Decision Trees: ID3

	Height	Weight	Age	Gender	Sick?
1	Tall	Fat	Young	Female	Yes
2	Tall	Thin	Middleage	Male	Yes
3	Short	Medium	Old	Male	No
4	Medium	Medium	Old	Female	No
5	Medium	Fat	Young	Male	No
6	Tall	Thin	Young	Female	Yes
7	Short	Medium	Middleage	Male	No
8	Medium	Fat	Young	Female	No
9	Tall	Thin	Old	Female	Yes
10	Tall	Thin	Young	Female	Yes
11	Short	Medium	Middleage	Male	No
12	Tall	Medium	Young	Male	Yes
13	Tall	Fat	Young	Female	Yes
14	Short	Thin	Old	Male	No
15	Medium	Thin	Old	Female	Yes
16	Tall	Fat	Young	Female	Yes
17	Tall	Thin	Middleage	Male	Yes
18	Short	Thin	Young	Female	Yes
19	Medium	Fat	Old	Female	No
20	Tall	Thin	Young	Male	Yes

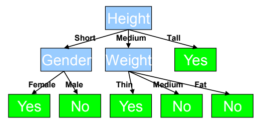


Symbolic Machine Learning (8)

Decision Trees: ID3

- Testing

	Height	Weight	Age	Gender	Sick?
T1	Medium	Fat	Young	Male	?
T2	Tall	Fat	Old	Male	?
T3	Short	Medium	Old	Male	?
T4	Medium	Medium	Old	Female	?



Statistical Machine Learning (1)

- Bayesian
  - All hypotheses are used to predict, using the prior probability of each hypothesis
- Maximum Likelihood (ML)
  - Assume a uniform prior on the hypotheses and ignore the prior
- Maximum a posteriori (MAP)
  - Make predictions based on the most probable hypothesis only

Statistical Machine Learning (2)

- Naïve Bayes Classifier – assumes that all attributes are conditionally independent

Statistical Machine Learning (3)

	Height	Weight	Age	Gender	Sick?
1	Tall	Fat	Young	Female	Yes
2	Tall	Thin	Middleage	Male	Yes
3	Short	Medium	Old	Male	No
4	Medium	Medium	Old	Female	No
5	Medium	Fat	Young	Male	No
6	Tall	Thin	Young	Female	Yes
7	Short	Medium	Middleage	Male	No
8	Medium	Fat	Young	Female	No
9	Tall	Thin	Old	Female	Yes
10	Tall	Thin	Young	Female	Yes
11	Short	Medium	Middleage	Male	No
12	Tall	Medium	Young	Male	Yes
13	Tall	Fat	Young	Female	Yes
14	Short	Thin	Old	Male	No
15	Medium	Thin	Old	Female	Yes
16	Tall	Fat	Young	Female	Yes
17	Tall	Thin	Middleage	Male	Yes
18	Short	Thin	Young	Female	Yes
19	Medium	Fat	Old	Female	No
20	Tall	Thin	Young	Male	Yes

	Height	Weight	Age	Gender	Sick?
T1	Medium	Fat	Young	Male	?
T2	Tall	Fat	Old	Male	?
T3	Short	Medium	Old	Male	?
T4	Medium	Medium	Old	Female	?

$$P(Sick \mid T1) = \frac{P(T1 \mid Sick)P(Sick)}{P(T1)}$$

Independence between feature values $x_1, x_2 \ldots x_n$ assumed

$$P(Sick = yes \mid T1) = \alpha \prod_i P(x_i \mid Sick = yes)P(Sick = yes)$$

$$= \alpha \frac{1}{12} \cdot \frac{3}{12} \cdot \frac{8}{12} \cdot \frac{4}{12} \cdot \frac{12}{20} = \alpha 0.083 \cdot 0.25 \cdot 0.67 \cdot 0.33 \cdot 0.60 = 0.0028\alpha$$

$$P(Sick = no \mid T1) = \alpha \frac{4}{8} \cdot \frac{3}{8} \cdot \frac{2}{8} \cdot \frac{5}{8} \cdot \frac{8}{20} =$$

$$\alpha 0.50 \cdot 0.375 \cdot 0.25 \cdot 0.625 \cdot 0.40 = 0.0117\alpha$$

Renormalising:

$$P(Sick = yes \mid T1)$$

$$= 0.0028\alpha / (0.0028 + 0.0117)\alpha = 0.19$$

Neural Networks (1)

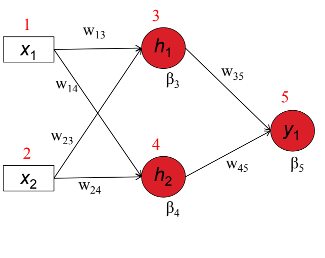
- Single and Multi layer neural networks
  - Weighted input function
  - Transfer functions (Threshold and Sigmoid)
  - Constructing a neural network from a truth table
- Gradient descent and Backpropagation
- Batch and Online Learning

Neural Networks (2)

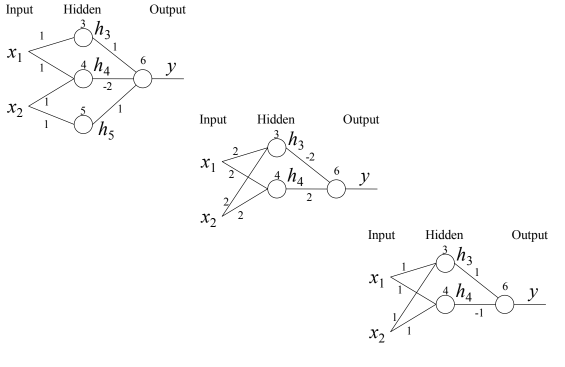
- Work out the neural network structure, including weight and bias values

Y = X1 XOR X2

X1	X2	Y
0	0	0
0	1	1
1	0	1
1	1	0



Neural Networks (3)



Natural Language Processing and Semantic Modelling (1)

N-gram word models

- Given a word sequence, what is the next word?
- Assume that each word only depends on a short linear history
- An n-gram model is a Markov chain of order n-1

$$P(w_1, w_2,..., w_k) = P(w_1)P(w_2 \mid w_1)P(w_3 \mid w_1, w_2)...P(w_k \mid w_1,..., w_k)$$

Natural Language Processing and Semantic Modelling (2)

N-gram word models

- Uni-gram model $P('w_1w_2w_3...w_n') = \prod_i P(w_i)$
  - Also known as Bag of Words
- Bi-gram model $P('w_1w_2w_3...w_n') = \prod_i P(w_i \mid w_{i-1})$
- Tri-gram model $P('w_1w_2w_3...w_n') = \prod_i P(w_i \mid w_{i-2:i-1})$
- Limitations include long distance effects in language and sparsity of words
- Improved by smoothing and using large data sets to train the models

Natural Language Processing and Semantic Modelling (3)

N-gram character model

- Using characters instead of words
- Can be used for language identification by determining the most probable language given the text

$$l^* = \arg\max_l P(l)\prod_{i=1}^{N} P(c_i \mid c_{i-2:i-1}, l)$$

Trigram character model

Natural Language Processing and Semantic Modelling (4)

Information Retrieval

- Using a query to find documents in a document collection that are relevant to a user and presenting the results
- We want to compute P(R=true|D, Q)
  - R – relevance
  - D – document
  - Q – query
- Choose documents that maximise:

$$\frac{P(r \mid D,Q)}{P(\neg r \mid D,Q)} = \frac{P(Q \mid D,r)P(r \mid D)}{\alpha(1-P(r \mid D))}$$

Natural Language Processing and Semantic Modelling (5)

Semantic Modelling

- Sentiment Analysis, Named Entity Recognition, Pattern Identification, Coreference Identification, Business Intelligence
- Latent Semantic Analysis, Co-occurrence, Naive Bayesian Term Similarity
- Information Visualisation
  - Concept Mapping – Leximancer
  - Temporal Data – Discursis

Language and Robots

- Social
- Vehicles
- Walking
- Commercial
- Language

Final Exam: Questions

- 6 Questions
- 60 marks total
- Question 1 is mainly content
  - (COMP7702 includes a discussion question)
- Questions 2 to 6 are mainly application

Final Exam: Weighting

- COMP3702
  - 50% if including mid-semester exam gives a better result
  - 60% if not including mid-semester exam gives a better result
- COMP7702
  - 50%

Final exam

- Wednesday 9 November 8:00am
- COMP3702
  - A-R 69-110
  - S-Z 67-141
- COMP7702
  - All 67-141
- Non-programmable calculator that is approved or labelled

Are you prepared for your examinations?

Do you:

- Have your current student ID card?
- Know where your exam is?
- Know what materials you are permitted to bring to the exam? (check with your course coordinator)
- Have an approved / labelled calculator (where calculators are permitted)

For each exam, ensure you:

- Have rechecked the timetable for exam date, time and venue and checked your emails
- Have your current student ID card on hand, and ready to present on entry to the exam room – should you forget it, you should report to the Student Centre before your exam
- Have spare pencils and pens, as well as any permitted materials
- Arrive at your exam venue 15 minutes before the scheduled start of the exam

Questions?