Use of calculator is permitted.
This is an open book examination.
Any written or printed material is permitted.

School of Computing

May 2018

COMP3771

User Adaptive Intelligent Systems

Answer all 3 questions

Time allowed: 2 hours

Question 1

Learning4Life is a newly set up education provider aiming for learners aged over 21. The vision is to use MOOCs (Massive Open Online Courses) to recruit learners and to deliver learning materials, support and assessment. While the resources are open-access, the business model relies on the value-added personalised support for learning and certification. This will be attractive to learners who aim to acquire specific knowledge and skills to improve their employability.

To provide illustrations in your answers use any one of the following examples t: 1) a foreign language graduate wanting to work as a translator in EU or UN organisations; or 2) a self-employed music teacher wanting to further his/her portfolio of music-related services (such as public performance, private tuition, composition and arrangement of music scores); or 3) an IT professional wanting to pick up some artistic / creative skills to improve their user interface design.

(a) Use Jameson's schema, which was introduced in this module, to explain how a learner's subscription information can be used to offer personalised recommendation of courses.

[10 marks]

Sample solution for Question 1:

Full marks are awarded only if understanding is demonstrated by giving examples from the scenario.

*Ans 1(a) Jameson's schema contains the main stages of user model acquisition, user model and user model application [**1 mark**]. Expect a brief description of the types of subscription information available for capture (e.g. learner's login id and/or email address, demographic details, qualifications acquired, interests and topic preferences for learning) [**2 marks**]. The method of model acquisition would be mainly explicit via a subscription form at the outset [**1 mark**] and subsequent edit/update [**1 mark**]. The profile information will be stored [**1 mark**] based on some form of computational representation (e.g. relational DB, or XML, or semantics) [**1 mark**] which can be retrieved by a recommendation system [**1 mark**]. Depending on the algorithm(s) used in the recommender system, other sources for data on content (i.e. catalogue of courses from external suppliers, pool of assessment instruments) [**1 mark**] and on other learners [**1 mark**] will be needed to detect relevance & apply strategy for adaptation.*

(b) Suggest one explicit and one implicit method to collect user information for adaptation, in addition to the subscription information from learners. For each method, state how the collected information can be used to influence the output of recommendation.

[4 marks]

*Ans 1(b) Possible explicit methods – like/dislike or star ratings (to pick the higher rated ones for recommendation); tagging (to select items which have most relevant tags to learner's interest) [**2 marks**].*

*Possible implicit methods – clicks or urls visited (to check real interests versus declared interests in subscription), timestamps for browsing (to detect if there is any short-term vs long-term interests) [**2 marks**].*

(c) Imagine that a learner has picked and started two courses.Explain what is a 'scrutable learner model' in this context, how it can be applied for further recommendations and what benefits it may bring. You may draw inspiration from the paper by Wasinger et al (2013) that was discussed in lectures.

[6 marks]

*Ans 1(c) A 'scrutable learner model' is a user model that can be inspected via a user friendly interface for the user to check the correctness of the information held [**1 mark**] and the rationale for its recommendation [**1 mark**]. In this case, the model should have captured the outcome of the learning done to date and the subscription/additional information captured from the learner as discussed in (a) and (b) [**1 mark**]. There may be an option for the user/learner to provide feedback on the accuracy / appropriateness of the recommendations [**1 mark**].*
*By allowing a learner to understand better on how their user model and feedback contributes to the resulting course recommendation, the learner will feel in control of the recommendations. For example, the learner can 'rate' the components of their model or reasoning, and this will be factored into the algorithms for future recommendation [**2 marks**].*

Question 2

A charity for carers has an idea of developing an online resource that promotes an all-round healthy lifestyle for its members. The aim is to encourage the carers to make wise choices for a balanced lifestyle and not to overlook their own needs while caring for others. The following aspects have been identified as important - emotional state, social relationships and physical fitness (including a right level of activity, relaxation and an appropriate diet).

(a) In preparation for a 'personalised reading recommender', the charity plans to start constructing user models for its online members. Illustrate a suitable vector representation for a member Alice who is interested in gardening and yoga, but not others such as cooking, walking and so on. Extend it to a matrix representation that will be suitable for User-User Collaborative Filtering algorithm.

[3 marks]

Marking scheme for Qu2

*Ans 2(a) Member's interests vector = (gardening, cooking, walking, yoga, …, etc.); hence, Alice = (1,0,0,1,0,…,0). **[1 mark].** A table similar to below:*

	User1	*User2*	…	*UserN*
Int1	*1 or 0*	*1 or 0*		*1 or 0*
Int2	…			
Int3				
…				
IntM				

Correct column and row labels

1's and 0's – 1 for having an interests, 0 for no interest or unknown <other rating scale is acceptable provided it's explained> ***[1 mark for partially correct answer; 2 marks for a complete answer]***

(b) As there are a number of health oriented mobile apps available on the market, the charity also aims to provide on its website a recommender service for health apps based on the use of stereotypes. Appropriate apps will be recommended to each of the charity's members depending on their profiles.

Imagine that the charity maintains a profile of each member that includes: `gender, age, marital status, list-of-interests,` where `list-of-interests` is a subset of `{outdoor-sports, indoor-activities, food, special-hobbies}`.

Three stereotypes are identified below.

```
I. FITNESS-FANATIC:
(age>21) and (age<35) and (list-of-interests includes outdoor-
sports, indoor-activities)
   P=0.7
```

Facet	Degree of interest	Rating (out of 100)
Have a smartphone	high	70
Use health app A	high	10
Use health app A	no	60
Use health app B	high	80
Gardening	moderate	60

```
II. YOUNG-CARER:
(age>14) and (age<24) and single and (list-of-interests includes
outdoor-sports, indoor-activities, food, special-hobbies)
   P=0.8
```

Facet	Degree of interest	Rating
Use health app A	high	20
Use health app A	no	40
Use health app B	no	40
Gardening	low	80

```
III. FOODIE:
(age>18) and (age<60) and (list-of-interests includes indoor-
activities, food)
   P=0.9
```

Facet	Degree of interest	Rating
Have a smartphone	moderate	30
Use health app A	high	70
Use health app A	no	10
Use health app B	high	50

(i) In the stereotype `FITNESS-FANATIC`, what does the value of the rating for the facet "Have a smartphone" mean?

[2 marks]

Ans 2(b) (i) Users who belong to FITNESS-FANATIC *have high interest in having a* smartphone*, and this is valid for 70% of the cases, i.e. for such users the probability of this facet to take the value "high" is 0.7.*

***(2 marks)** for a correct and complete description (up to 1 mark if the description is partially complete).*

(ii) Provide an example from the above stereotypes to illustrate a case of contradiction in stereotypes.

[3 marks]

*Ans 2(b) (ii) A correct answer should point at a conflict between stereotypes by suggesting a user profile for which the facets values and ratings are in conflict. For example: Male/female, 30, and interested in indoor-activities} – conflict with regard to '*Use health app A*' between* FITNESS-FANATIC *(not interested in using it) and* FOODIE *(very interested in using it).*

***(3 marks)** for a full and correct answer (up to 2 for mostly correct and up to 1 mark if the explanation is only partially correct).*

(iii) Consider a member who is a 19-year-old female, interested in food and indoor-activities. Should she be recommended health app A or health app B? Present your working step-by-step.

[12 marks]

Ans 2(b) (iii)

This member belongs to two stereotypes: YOUNG-CARER and FOODIE. ***(1 mark)***

For health app A = high (use both stereotypes to calculate combined probability)

*P(*YOUNG-CARER*) = 0.8*0.2 = 0.16*

*P(*FOODIE*) = 0.9*0.7 = 0.63*

*Combined P = 0.16 + (1-0.16)*0.63 = 0.69* ***(3 marks)***

For health app A = no (use both stereotypes to calculate combined probability)

*P(*YOUNG-CARER*)= 0.8*0.4 = 0.32*

*P(*FOODIE*) = 0.9*0.1 = 0.09*

*Combined P = 0.32 + (1-0.32)*0.09 = 0.38* ***(3 marks)***

For health app B = high (use FOODIE *to calculate probability)*

*P = 0.9*0.5 =0.45* ***(1 mark)***

For health app B = no (use YOUNG-CARER *to calculate probability)*

*P = 0.8*0.4 = 0.32* ***(1 mark)***

Hence for this member,

P(app A = high) = 0.69 > P(app B = high) = 0.45 : it is more likely to have high interest in app A - a difference of 24%

P(app A = no) = 0.38 > P(app B = no) = 0.32 : it is only slightly more likely to have no interest in app A than app B - a difference of 6%

So on balance, should recommend app A. ***(3 marks for recommendation and reasoning)***

Question 3

This question relates to a hypothetical online museum system which enables users to browse through a collection of digital images of paintings. A recent customer survey has shown that users are often lost in the digital space and unable to find interesting items in the collection. The IT team that maintains the online museum system has decided to integrate appropriate personalisation techniques.

(a) There are log files from the last two years recording every user click on a painting. There is no additional information about the users. The database of paintings stores additional details, such as author, art style (e.g. impressionism, cubism, modernism), year, country of origin. Following these conditions, identify a recommender algorithm that will be suitable for the online museum and describe how it will be applied.

[4 marks]

Ans 3 (a) Possible answers include:
Using the log files and the data about the paintings, implicit user profiles can be derived. For each user we can identify favourite authors, preferred art styles, art periods or art schools of interest. The profiles can be used in a content-based filtering algorithm.
Another possibility is to extract interest matrix which records for each painting the users who have visited that painting. The interest matrix can then be used in a collaborative filtering algorithm (either user-user or item-item). When a user visits a painting, he/she will be recommended similar paintings.

2 marks for a correctly identified method that is suitable for the scenario.
2 marks for a description how the method can be applied.

(b) In a further extension of the museum system, users are asked to indicate whether they would like to see the real painting when they visit the museum exhibiting this painting. The matrix R below indicates for each user $u_1, \ldots, u_5$ which paintings $p_1, \ldots, p_8$ he/she would like to see (value 1 indicates that the user would like to see the real painting).

$$R = \begin{bmatrix} & u_1 & u_2 & u_3 & u_4 & u_5 \\ p_1 & 1 & 1 & 0 & 1 & 1 \\ p_2 & 0 & 0 & 0 & 0 & 1 \\ p_3 & 0 & 1 & 0 & 1 & 0 \\ p_4 & 1 & 1 & 1 & 1 & 0 \\ p_5 & 1 & 1 & 0 & 0 & 1 \\ p_6 & 1 & 0 & 1 & 0 & 1 \\ p_7 & 1 & 0 & 0 & 1 & 1 \\ p_8 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

User u_3 visits the Tate gallery in London where painting p_1 is exhibited. Apply item-item filtering with 3-nearest neighbours to find out whether user u_3 may like to see painting p_1. Show your working. Illustrate one problem with this algorithm with an example.

[8 marks]

Ans 3(b)

Step 1: Calculate the similarity between p_1 and all other paintings (we will use cosine similarity):

$|p_1| = \sqrt{4} = 2 \quad |p_2| = \sqrt{1} = 1$

$|p_3| = \sqrt{2} \quad |p_4| = \sqrt{3} \quad |p_5| = \sqrt{3} \quad |p_6| = \sqrt{2} \quad |p_7| = \sqrt{3} \quad |p_8| = 0$

$sim(p_1, p_2) = \frac{p_1 * p_2}{|p_1| * |p_2|} = \frac{1}{2} \quad sim(p_1, p_3) = \frac{1}{\sqrt{2}} \quad sim(p_1, p_4) = \frac{3}{2\sqrt{3}}$

$sim(p_1, p_5) = \frac{3}{2\sqrt{3}} \quad sim(p_1, p_6) = \frac{1}{\sqrt{2}} \quad sim(p_1, p_7) = \frac{3}{2\sqrt{3}} \quad sim(p_1, p_8) = 0$

Step 2: Select k-nearest neighbours, where k=3. The three most similar neighbours are p_4, p_5, and p_7.

Step 3: Calculate the weighted sum:

$$weighted\_sum = \frac{\frac{3}{2\sqrt{3}}}{\frac{9}{2\sqrt{3}}} = \frac{3 * 2\sqrt{3}}{9 * 2\sqrt{3}} = \frac{1}{3}$$

The user is unlikely to be interested in the painting.

An example problem may point out that:

- *Painting p_8 was not visited by anybody, so will never be recommended to a user.*
- *The formation of the neighbourhood affects the precision of the algorithm (e.g. if the size of the neighbourhood was 2, it would have to choose 2 out of the 3 most similar paintings, the choice would have affected the result; a neighbourhood of size 4 would have produced a different result too).*

3 marks for a correct calculation of Step 1
1 mark for a correct calculation of Step 2
2 marks for a correct calculation and conclusion of Step 3
2 marks for an appropriate example of a problem (marks will not be awarded if the example is general and unrelated to the particular case).
[8 marks]

(c) The recommendation algorithm has to be assessed. The first criterion considered is accuracy. Suggest one method that can be used to evaluate the accuracy of the recommendation and describe how the method can be applied in this specific application (i.e. online museum system). Indicate another criterion that can be used to evaluate the recommendation algorithm, and justify why you have chosen this criterion.

[4 marks]

Ans 3 (c)
Two methods have been introduced to the students:
-Mean absolute error
-Precision and recall
Both can be applied in this case, the description should show how the information about the method is collected (e.g. for the mean absolute error, we need to consider the difference between the user predictions and the algorithm predictions, it is crucial to show how the user predictions will be obtained; for the precision and recall, the number of relevant and irrelevant recommendations is calculated, should describe how the relevance of an item will be assessed; both cases may involve some sampling of the recommendations, if the data size is big – this should be mentioned too).

The suggestions for another criterion should refer to usability, usefulness, applicability. The chosen criterion should clearly apply to the specific case. A research article with evaluation methods for recommender systems has been distributed at lectures. This question checks that the students can select methods for a specific application and apply in the specific context.

2 marks for suggesting an appropriate accuracy method with description how to apply.
2 marks for another criterion with appropriate justification.
[4 marks]

(d) The recommendations derived with the algorithm from (c) will be used to offer a user a PDA (personal digital assistant) with a user-adaptive guide which gives descriptions of paintings when the user visits the physical gallery (e.g. the Tate gallery in London). For this, it uses a representation of the current context. Suggest two parameters that should be considered for modelling context in the personal museum guide and justify your suggestion.

[4 marks]

Ans 3 (d) Possible context parameters include:
-Device context (description of the device and its affordances) – important for content rendering (e.g. will pictures/text/video/speech be used)
-User model (what the user likes or dislikes) – important for deciding the level of detail (if the user is interested in the painting more detail about it can be given)
-Location context (user position, paintings in a near proximity, position of related paintings) will be important to decide what other paintings the user can be directed to
-Social context (users have different behaviour when visiting museums in groups) – this can be related to the level of detail shown and suggestions given (the students were directed to an article describing different patterns of behaviour in museums and how users' expectations different because of this)

2x(1 mark for correct parameter + 1 mark for appropriate justification) ***[4 marks]***