Diploma Programme
Programme du diplôme
Programa del Diploma

Markscheme

May 2023

Computer science

Higher level

Paper 3

9 pages

Subject details: **Computer science HL paper 3 markscheme**

Mark allocation

Candidates are required to answer **all** questions. Total 30 marks.

General

A markscheme often has more specific points worthy of a mark than the total allows. This is intentional. Do not award more than the maximum marks allowed for that part of a question.

When deciding upon alternative answers by candidates to those given in the markscheme, consider the following points:

- Each statement worth one point has a separate line and the end is signified by means of a semi-colon (;).
- An alternative answer or wording is indicated in the markscheme by a "/"; either wording can be accepted.
- Words in (…) in the markscheme are not necessary to gain the mark.
- If the candidate's answer has the same meaning or can be clearly interpreted as being the same as that in the markscheme then award the mark.
- Mark positively. Give candidates credit for what they have achieved and for what they have got correct, rather than penalizing them for what they have not achieved or what they have got wrong.
- Remember that many candidates are writing in a second language; be forgiving of minor linguistic slips. In this subject effective communication is more important than grammatical accuracy.
- Occasionally, a part of a question may require a calculation whose answer is required for subsequent parts. If an error is made in the first part then it should be penalized. However, if the incorrect answer is used correctly in subsequent parts then **follow through** marks should be awarded. Indicate this with "**FT**".
- Question 4 is marked against markbands. The markbands represent a single holistic criterion applied to the piece of work. Each markband level descriptor corresponds to a number of marks. When assessing with markbands, a "best fit" approach is used, with markers making a judgment about which particular mark to award from the possible range for each level descriptor, according to how well the candidate's work fits that descriptor.

General guidance

Issue	Guidance
Answering more than the quantity of responses prescribed in the questions	• In the case of an "identify" question read all answers and mark positively up to the maximum marks. Disregard incorrect answers. • In the case of a "describe" question, which asks for a certain number of facts *eg* "describe two kinds", mark the first two correct answers. This could include two descriptions, one description and one identification, or two identifications. • In the case of an "explain" question, which asks for a specified number of explanations *eg* "explain two reasons …", mark the first two correct answers. This could include two full explanations, one explanation, one partial explanation *etc*.

1. (a) *Award **[2]** max*
Provides the necessary computing resources for a recommender system;
Resources available quickly/short lead time;
More customisable/greater control than PaaS;
Cost efficiency/pay for what they use;
Storage resiliency;
Reduced maintenance of on-premises data centre;
IaaS provides security/compliance certifications;
Flexibility to choose computing resources;
Offers scalability/easier to scale IT resources;
Disaster recovery provided;

(b) *Award **[2]** max*
An optimization algorithm to minimize the cost/loss function (predicted-actual);
Starts with an initial guess (random point)/uses randomly selected sample/uses a single datapoint at each iteration/incrementally updates the parameters iteratively;
Which makes the path to optimum is less direct/faster/moves in the opposite direction to the gradient/moves in the direction of steepest descent;
Makes small changes to item-feature and user feature matrices/used in the matrix factorisation prediction-error-update process;

2. (a) *Award **[4]** max*
Benefits *[2 max]*
Abundant data readily available/Explicit data only may not be enough/Users who don't provide explicit data can still get good recommendations;
Leads to a much greater range of data that can be analysed;
Authentic data/captures the subconscious thoughts/feelings/opinions of users that may be more insightful than the explicit behavioural data;
Helps fill gaps where explicit data is not available;

Concerns *[2 max]*
May compromise the users right to privacy or their right to anonymity;
The user is not aware that the data is being collected so it could be seen as unethical;
Revealing a user's identify may cause them harm (potential lawsuit/bad publicity);
May include unsuitable data based on a user's impulses rather than considered opinions;
May lead to considerably more data being collected that is of limited value;
Wrong conclusions may be drawn due to ambiguity of data (e.g. long time on a page may be due to confusion or a coffee break);
Dataset has too much noise, which may lead to overfitting;

Example Conclusions *[1 max]*
It depends on the nature of the data and how sensitive it is;
It depends on how the data is stored, with whom the data is shared, and for what purposes;
Even though implicit data is beneficial, users should be given the right to decline its use;

Notes for examiners:
- *A generic conclusion/opinion (e.g. benefits outweigh the concerns, or vice versa) is no marks, but if included, award **[4 max]** for two benefits and two concerns.*
- *If no opinion is given **[3 max]** for benefits and concerns i.e. **[2]** + **[1]** or **[1]** + **[2]***
- *If an insightful conclusion is provided, award **[4 max]** i.e. **[1]** benefit + **[2]** concern + **[1]** conclusion or **[2]** benefit + **[1]** concern + **[1]** conclusion.*

(b) *Award **[4]** max*
Each item in the database is represented as a point in a multi-dimensional feature space/ represent each item as a vector of features;
K hyperparameter is set;
Choose the method to calculate distance (e.g. Euclidean, Manhattan, Cosine similarity) (between target item and other items);
K-number of closest items are calculated/finds distance to other data points/a similarity score is determined;
Those that fall within this proximity are recommended;
The predicted rating could be calculated as a weighted average of the k items' ratings.

3. *Award **[6]** max*
CF has a **cold start problem;**
Cannot make recommendations for new users due to a lack of data/no preferences;
Cannot recommended new items due to a lack user interactions with them;
Necessarily to use an alternative (e.g. Content-based/hybrid) in the beginning/New users may be asked for preferences when they sign up/pilot group used to gather data.

CF **suffers from data sparsity** problem;
All users do not give feedback or rate items that they have used which makes similarity data unavailable;
This results in little training and testing data;
Data has to be predicted using different techniques;

CF is based on the premise that **users and items have been profiled** correctly;
Current behaviour of users may change in the future;
Users may give different rating at different times/Users may change their views;

CF suffers from **poor scalability**;
Computations are expensive and grow non-linearly when number of users and items added;
The database can become extremely large;

CF suffers with **popularity bias**/crowding out;
Popular items have more interactions, so get recommended/new artists fail to get recommended;
Leads to a feedback loop where popular items become even more popular/lesser-known artists are ignored/the goal of NextStar is to introduce the new artists;

CF may struggle with **heterogeneous** (diverse) **populations**/works better with homogenous populations;
Because recommendations assume people have similar interests;
Atypical people will be recommended content that typical people enjoyed/Atypical people would have to filter out content that doesn't interest them.

*Award **[3]** and **[3]***

4. *Award **[12]** max*

Effective RS should recommend content/items of user's choice or preference.

Development challenges:

- Challenge of accuracy and coverage.
- Challenge of new users and new items (cold start).
- How to deal with data sparsity.
- Data cleaning challenge, single items represented with multiple names, duplicate user identity.
- Scalability issues.
- Adaptability to changing tastes.
- Dealing with malicious users who give false ratings.
- User studies to collect data are expensive and time consuming.
- Challenges of test and training data generation, bias introduced in data can skew the results.
- Choice of machine learning approach (e.g. supervised, semi-supervised, unsupervised, reinforcement learning).
- Challenge of overfitting with supervised and ANN.
- Choice of algorithm may result in better recommendations, but each has its own strengths and weakness.
- Content based filtering focuses on item attributes, characteristics. Collaborative filtering has the potential to adjust its recommendation within a short period of time. However, attributes may not be correctly defined. May lead to over-fitting and lack of diversity.
- Collaborative filtering based on past interactions, on user ratings and feedback. User data may not be always available. Need to select the right method k-NN or matrix factorization to predict unavailable data. Suffers with the cold start, data sparsity, poor scalability, and popularity bias problems.

Choice of training/testing data

- If models are trained on high-quality data, they can make good predictions. Training data is used to learn, and test data is used to validate the performance.
- Reinforcement learning can be used to build labelled datasets for supervised learning models.
- Choice is made between using exemplary training data and actual complete data.
- Biased training data can result in skewed outcomes.
- Performance of prediction affected when data set is imbalanced.
- Overfitting of training data can negatively affect the predictive accuracy.

Methods of evaluation

- Selecting suitable metrics for evaluation.
- Statistical measures MSE, RMSE, F-measure for evaluating precision and accuracy.
- For correct evaluation, need quantitative as well as qualitative data.
- Offline evaluation.
- Other metrics like presentation of results, intuitiveness of results, novelty of recommendation are difficult to measure.
- Challenge of combining resurfacing and discovery items in the same list.
- Difficult to decide on a baseline model to compare with.

Conclusions

- A pre-launch gathering of data from a pilot group may help overcome a cold start.
- A content-based model using KNN may be a good starting point because it works well with little data.
- A collaborative filtering model may be brought online after a period of time.
- A hybrid design model might improve recommender system.
- RS should interact with users by polling their opinion and taking new data into account.
- Prediction error due to noise generated by the user could be reduced by asking for re-rating.
- Some concepts are difficult to measure like novelty of a recommendation hence optimized evaluation of RS is not easy.
- No generic best RS exists, appropriate solution is different. For example, a RS for movies would be different from an RS for academic research paper publications.
- Trade-off between accuracy and diversity of RS, between large set of data to improve result and cost and time involved lead to some compromise in the development of an effective RS.

Please see markband on page 9.

Marks	Level descriptor
No marks	• No knowledge or understanding of the relevant issues and concepts. • No use of appropriate terminology.
Basic 1–3 marks	• Minimal knowledge and understanding of the relevant issues or concepts. • Minimal use of appropriate terminology. • The answer may be little more than a list. • No reference is made to the information in the case study or independent research.
Adequate 4–6 marks	• A descriptive response with limited knowledge and/or understanding of the relevant issues or concepts. • A limited use of appropriate terminology. • There is limited evidence of analysis. • There is evidence that limited research has been undertaken.
Competent 7–9 marks	• A response with knowledge and understanding of the related issues and/or concepts. • A response that uses terminology appropriately in places. • There is some evidence of analysis. • There is evidence that research has been undertaken.
Proficient 10–12 marks	• A response with a detailed knowledge and clear understanding of the computer science. • A response that uses terminology appropriately throughout. • There is competent and balanced analysis. • Conclusions are drawn that are linked to the analysis. • There is clear evidence that extensive research has been undertaken.

Total: [30]