Review Form for Multimedia Communications Seminar SS 2018

=============================================================================

(Please replace all the []’s with your ratings/text. To be filled in English)

Paper Title

-----------

Evaluation of Content Recommender Systems in Learning Scenarios

Previous work

-------------------------

Please judge the extent to which this paper recognizes and summarizes previous work in this area. This should include a number of references.

1 = no previous work mentioned

2 = very little amount of pointers to previous work, or very unclear relevance

3 = some pointers to previous work, but important references are missing

4 = mostly complete overview of previous work

5 = very good summary of previous work, given the space constraints

Your Score: 2

Style and Organization

----------------------

Please judge the writing style of this paper. Is the writing style scientific? Is the language used correctly? Is the outline and overall organisation suitable?

1 = this is not a scientific paper

2 = there are a lot of style problems, organization is unclear

3 = there are some style problems, but it is ok overall

4 = only minor style or organizational problems

5 = great style, language, and overall organization

Your Score: 5

Formatting

-------------

Please judge how well the formatting guidelines were followed. Does the paper contain an unnumbered abstract? Are authors and their affiliation given properly? Was the ACL style file used? Doe the paper contain a reference section?

1 = this does not follow the formatting guidelines at all

2 = some part of guidelines are followed, but overall the formatting is very remote from the guidelines

3 = formatting is ok, but needs improvements

4 = only minor details do not follow the formatting guidelines

5 = ready to print! formatting guidelines were followed entirely.

Your Score: 4

State of Completion

-------------------

Please judge how far this work is from completion. This includes marking missing experiments, missing sections in the paper, and missing information.

1 = extremely preliminary - no experiments were conducted

2 = very preliminary: some things were done, but still a lot is missing towards a final version

3 = paper is still incomplete, but it is clear where the author is heading

4 = only a few experiments / sections are missing, paper is close to complete

5 = this paper comprises a complete work

Your Score: 4

Overall Rating

--------------

Please give an overall rating for this paper. If a lot of papers compete for a limited amount of slots, this would be the main score for deciding on paper acceptance.

1 = this paper is not worth presenting

2 = I would rather not see this paper in the workshop

3 = the paper is ok for the workshop, if enough slots are available

4 = I would rather like to see the paper in the workshop

5 = the paper will be great for the workshop

Your Score: 4

Reviewer Confidence

-------------------

Please state how confident are in you decisions

1 = I have no idea about the topic of this paper, so please don't trust my judgments

2 = I have very little idea about the topic, so my judgments should be taken with a grain of salt

3 = I have some idea about the topic, but I am not an expert and I could have missed some things

4 = I am pretty confident that my judgments are correct, since I know the topic well

5 = I am an expert and thus very confident that my judgments are correct

Your Score: 4

I feel pretty confident about this review, although I don’t have any prior experience regarding Recommender Systems. My own topic is closely related (user modeling in e-learning), I have experience in user-centered methods and my own …. Fachbereich has a more sophisticated paper-culture regarding references, structure and a critical view, compared to FB20.

Comments to the authors (please refer to https://cs.brown.edu/~sk/Memos/Paper-Reviews/)

-----------------------

The paper describes different evaluation methods for content recommender systems in (e-)learning scenarios. After a quick introduction in issues specific to recommender systems in learning, the author dives deeper into different approaches that are commonly used to evaluate proposed systems. In the final sections, it is uncovered that the current trend to tend towards automated offline experiments is not suitable to actually evaluate recommender systems in learning, since the user component is underevaluated.

The paper follows a very clear structure, establishes basics necessary to understand the following sections and comes to a satisfying conclusion. The scope narrows down from recommender systems [RS] in general to RS in learning to commonly used evaluation methods in this area and finally to a specific look at and evaluation of evaluation methods, and thus manages to answer the question of how RS in learning should be evaluated, contrary to common practice. However, the final sections feel a little bit rushed in comparison to the rest of the paper.

The language is professional, scientific and easy to understand, with only minor typos and errors that do not interfere with understandability. Elements like itemized lists and subsections were used to break large pieces of text, making the paper pleasant to read.

The figures provided by the author are very relevant to the text and highlight relevant findings of the final study. Unfortunately, some figures appear to be a little bit underexplained, which could be improved.

The formatting is good and does follow the guidelines. There are two minor issues here: the author’s affiliations are missing and on pages 3 and 4 there’s a little hiccup with section headlines.

The major shortcoming of the paper however is the lack of a critical perspective, expressed in the very few references provided by the author and the rather abrupt ending. While the three references used are in fact very relevant to the topic and provide great guidance through the subjects, lots of claims remain unreferenced. These would have been important to properly estimate the feasibility of the paper’s findings, making the author’s final thoughts and concerns regarding the results even more important. Unfortunately, these are a little short as well: while the important findings are concisely summarized, they are not questioned, but just taken as a fact. A *critical* discussion would greatly improve the paper, especially considering the already established potential unreliability of self-reported user evaluation.

Going into more detail, there are fortunately very few things to further delve into:

The abstract is concise and to the point, although the first sentence confused me at first. Commonly, the fact that a topic receives lots of research is not the reason why this specific piece of work should be relevant. However it makes perfect sense in this case, since commonly used practices of the referenced publications are the focus. There’s a sentence in the middle saying the same thing twice. (Look for “adapt”.)

The aforementioned problem of missing references appears immediately in the second sentence of the introduction: “Recently it was noticed, that accuracy metrics might not be enough to evaluate recommender systems [...]” Who noticed this? When? How? Another example is just a few lines down: “[…], the goals of [RS] are not restricted to accurate recommendations, rather additional metrics such as usefulness, novelty, diversity or serendipity should be used.” This pattern holds true through the entirety of the paper: statements about other people’s research and conclusions should always be backed by citations.

The final lines of the introductional section provide a great overview over what to expect from this paper (and is something I should consider for my own paper as well ;) ). Still, I would propose to try to slightly shorten it, since it makes up a large part of the introduction.

Section II starts strong, and provides another good summary about this section that is guided by the first reference. However, inside the subsections themselves, it remains unclear where specific information originated. The only citation provided adresses the headlines of these detailed explanations. Did this paper uncover these findings or is this a summary of other work already? Sometimes the initial source is referenced again, but more often, the reader would not know where a specific bit of information originated.

The subsections receive a brief critical discussion (issues regarding implicit feedback interpretation, confirmation bias), which greatly improves them. While further criticism could be found, the rather narrow scope of our assignment leads to this being a good length for these paragraphs.

The summary at the end of the section however could provide some more information: are there other things that need to be considered but are not discussed by the mentioned source? What are the things that can be transmitted from e-commerce?

I will use this section to highlight the only language errors I encountered when reading through the paper:

- as far as I understand the English language, some commas could be deleted. (I’m not too sure about this, though.)

- several occurrences of missing ‘ in phrases like “[…] the user’s tase [...]”, “A user’s knowledge [...]”

In section III, the evaluation methods for RS are discussed. At first, a summary of evaluation goals or potential metrics are discussed, before evaluation methods are explored. In subsection “System information”, I wondered how to assess whether items are “useful to the user” in order to calculate the precision.

The methodology sections are great and provide information on the respective approach itself, as well as advantages and disadvantages of these methods. In the middle of “Offline Experiments” there’s a reference to whether “a user actually bought the recommended items”, which confused me since up to this point e-commerce was strictly out of scope.

I also think that there’s a misplaced headline: “2) User Studies” on page 3 on top of the right column is missing and it appears on page 4, on top of the right column instead. This has also led to a problem with section IV’s headline appearing right above the figure.

Regarding the figures on page 4, at first I wished for more context for figure 2 (Number of studies taken into consideration, how these were selected, …), but thinking about it, the citation of reference 2 should suffice. Figure 1 however is actually more complicated than the description provided in the text. Maybe its prominence should be reduced.

The last sentence of the first paragraph of section IV is missing a verb. Additionally, the transition between these paragraphs did not make it completely clear what the evaluation wanted to accomplish. After reading the whole paper it is easy to understand the goal in hindsight, but at first it appeared a little bit chaotic. My other notes include a question regarding the “test data” (How was it compiled?), and the suggestion to include a reference to the explanation of accuracy and recall provided earlier.

The following summary of the study and it’s findings are to the point. In figure 3, I wondered about “neighborhood sizes” and the “@” used in (a), but these might be out of scope.

The only remaining issue to note is the aforementioned lack of a proper discussion of these findings.

With more references and a more critical view in the final section, this paper would be a 5 overall. The structure is very clear and easy to follow, the main findings of the current references are summed up and reduced to the most relevant parts, and the presentation in terms of style and text structure builds hope for a good presentation during the workshop.

Confidential comments to the committee (this part will only be shown to supervisors, not the author)

--------------------------------------

I have explicitly stated most of my critique in the public part of this review. However, I would like to emphasize that the main part of this paper seems to consist of just summarizing three other papers without any additional thoughts. From what I have learned in FB03‘s culture towards papers, this seems to be rather lazy: the proper discussion of findings and the presentation of relevant and comparable literature is the main workload of such a review paper.

I don‘t want to assume any bad intentions from the author‘s part, and maybe his references and discussions are simply not in place yet. However, I tried to make a point describing the missing features, and sincerely hope that the author understands what I feel is missing. The general direction, style and structure of the paper are great and could provide a great report overall.