Explaining Recommendations by Means of User Reviews

Tim Donkers
University of Duisburg-Essen
Duisburg, Germany
tim.donkers@uni-due.de

Benedikt Loepp
University of Duisburg-Essen
Duisburg, Germany
benedikt.loepp@uni-due.de

Jürgen Ziegler
University of Duisburg-Essen
Duisburg, Germany
juergen.ziegler@uni-due.de

ABSTRACT
The field of recommender systems has seen substantial progress in recent years in terms of algorithmic sophistication and quality of recommendations as measured by standard accuracy metrics. Yet, the systems mainly act as black boxes for the user and are limited in their capability to explain why certain items are recommended. This is particularly true when using abstract models which do not easily lend themselves for providing explanations. In many cases, however, recommendation methods are employed in scenarios where users not only rate items, but also provide feedback in the form of tags or written product reviews. Such user-generated content can serve as a useful source for deriving explanatory information that may increase the user's understanding of the underlying criteria and mechanisms that led to the results. In this paper, we describe a set of developments we undertook to couple such textual content with common recommender techniques. These developments have moved from integrating tags into collaborative filtering to employing topics and sentiments expressed in reviews to increase transparency and to give users more control over the recommendation process. Furthermore, we describe our current research goals and a first concept concerning extraction of more complex argumentative explanations from textual reviews and presenting them to users.

ACM Classification Keywords
H.3.0. Information Storage and Retrieval: General

Author Keywords
Recommender Systems; Deep Learning; Explanations

INTRODUCTION
Today's *Recommender Systems* (RS) have been shown to generate recommendations for items that match the user's interest profile quite accurately. The underlying methods are usually based either on purchase data or ratings of other users (collaborative filtering), or on structured data explicitly describing the items (content-based filtering). Yet, algorithmic maturity does not necessarily lead to a commensurate level of user satisfaction [11]. Aspects related to user experience such as the amount of control users have over the recommendation process or the transparency of the systems may also contribute substantially to the user's acceptance and appraisal of the recommendations [11]. Still, many state-of-the-art methods appear to the user as black boxes and do not provide a rationale for the recommendations, which may negatively influence intelligibility, and thus user's comprehension and trust [18].

When recommendation methods are applied in the real world, users can often provide textual feedback on items in the form of short tags or written reviews. Textual feedback from other customers is known to strongly influence the current user's decision-making [2]. However, perusing all reviews associated with the items in a recommendation set is time-consuming and mostly infeasible. The information available in review data is currently also hardly exploited for making the otherwise opaque recommendation process more transparent. While research has more recently begun to investigate the role of, for instance, product features mentioned, topics addressed, or general sentiments expressed, for improving algorithmic precision [25, 5, 1], their potential for increasing intelligibility of recommendations has not been fully exploited yet. The same applies for aspects such as presence, polarity and quality of arguments found in the reviews for or against a product. Thus, extracting and summarizing relevant arguments and presenting them as textual explanations seems to offer a promising avenue to supporting users better in their decision process.

RELATED AND PRIOR WORK
One popular way of increasing transparency of RS is to use textual explanations [20]. When sufficient content information is available, item attributes may be aligned with user preferences to explain a recommendation [22]. Such data can also be used together with context information to point out arguments for recommended items, and may simultaneously serve as a means to critique recommendations [13]. For item-based collaborative filtering, the static variant used e.g. by Amazon ("Customers who bought this item also bought...") is quite popular. For model-based methods, it is still difficult to improve transparency through explanations. The approach proposed in [25] actually exploits review data, but is limited to identifying sentiments on a phrase-level to highlight product features the user is particularly interested in. In other cases where reviews have been analyzed semantically, this has served primarily to improve model quality, e.g. by inferring hidden topics, extracting content aspects, or mining user opinions [5, 1]. More advanced approaches that, for instance, extract argumentative explanations from review texts to support the user's decision-making do not exist. One exception where argumentation techniques have been integrated into RS is presented in [4]. Yet, this approach depends on the availability of explicit item features and, since implemented in defeasible logic, requires defining postulates manually.

In prior work of ours [6, 7], we proposed a model-based recommendation method that exploits textual data. *TagMF* enhances a standard *matrix factorization* [12] algorithm with tags users

ExSS '18, March 11, Tokyo, Japan

provided for the items. By learning an integrated model of ratings and tags, the meaning of the learned latent factors becomes more transparent [7] and the user may interactively change their effect on the generated recommendations [6]. In contrast to other attempts that integrate additional data, this improves comprehensibility of recommendations as well as user control over the system which is typically limited to (re-)rating single items. Moreover, our user study showed that not only objective accuracy as measured in offline experiments, but also perceived recommendation quality benefits from complementing ratings with additional tag data. Apparently, tags can introduce semantics into the underlying abstract model that are natural to understand so that users notice some kind of inner consistency in the recommendation set. Besides, the relations between users and tags introduced by our method allow to explicitly describe user preferences in textual form: We can automatically derive which tags are considered important to an individual user, even in cases when the user has never tagged items him- or herself. These tags can then be presented as an explanation of his or her formerly latent preference profile [7]. Despite its advantages, our method requires the relationship between items and additional data to be quantified in numerical form. If this requirement is not met, it seems a natural extension to exploit other, more general forms of user-generated content such as product reviews.

In [9], we presented an interactive recommending approach relying on review data. We extended the concept of *blended recommending* [15] by automatically extracting keywords from product reviews and identifying their sentiment. Then, we used the extracted (positive or negative) item descriptions to offer filtering options usually not available in contemporary RS. In our system, we present them as facet values that can be selected and weighted by the user to influence the recommendations. We conducted a user study that provided evidence that users were able to find items matching interests that are difficult to take into account when only structured content data is exploited. Without requiring users to actually read the reviews, the method seems promising for improving the recommendation process in terms of user experience, especially when users have to choose from sets of "experience products". Although in principle any algorithm can be integrated in the system's hybrid configuration, we have not yet utilized the advantages of model-based recommender methods in combination with the ones of exploiting user-generated reviews.

A CONCEPT FOR EXPLAINING RECOMMENDATIONS

Building on our prior work, we in the following present a novel concept that relies on extracting and summarizing arguments about products from textual reviews in order to provide users with adapted model-based recommendations, and in particular, explanations that are personalized according to the current user's preferences and styles of decision-making.

While tagging helps to classify items by attributing specific properties, the descriptive nature of tags limits their applicability as an explanatory element. User reviews, on the other hand, constitute semantically rich information sources that incorporate sentiment and often an intended sequence of arguments, i.e. an argumentation flow. Reviews go beyond mere objective descriptions by (sometimes implicitly) invoking a stance towards a target. Users reading a review can get an idea of motives and reasoning behind its author's words. This verbal provision of subjective information constitutes a comprehensible context on which arguments and their interrelations can be grounded. However, automatically identifying presence, polarity and quality of argumentation structures has not yet been considered in RS research, although it is well known that textual feedback of other users may strongly influence decision-making [2]. In particular, the potential of arguments has not been exploited for explaining recommendations. Due to their persuasive nature, arguments can be thought of providing intelligible reasons that support recommendations, and may thus increase system transparency and trust in the results.

Since users differ with respect to dispositions and preferences, they may attach different levels of importance to arguments. For example, one user may insist on closeness to the beach when booking a hotel, while another user lays more focus on cleanliness or friendliness of the staff. Although reading the same review, these two users would, most likely, attend to completely different aspects. A sophisticated argument extractor should therefore mimic human scanning behavior by adaptively assigning individual attention weights to arguments.

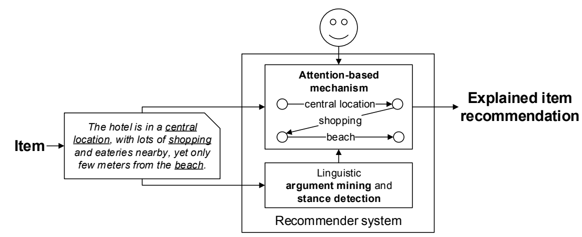

Figure 1. In our framework, a review is analyzed linguistically and via an attention-based mechanism. This allows to implement an argumentation flow based on information provided in the review while deeply integrating the user. Eventually, a personalized recommendation is presented together with individual arguments for or against this product.

We propose a conceptual framework that automates the process of extracting arguments from reviews (see Figure 1) to come up with personalized argumentative item-level explanations for recommendations: We suggest to apply feature-based methods from *computational linguistics* in order to derive argumentative structures, including argument polarity in the form of stances. Beyond that, detecting personally important arguments requires a deep integration into the process of calculating recommendations. We aim at utilizing *deep learning* methods that—matching the attention analogy—enable the system to focus on important concepts subject to a user variable. Put together, this leads to the following challenges:

- Linguistically analyzing review texts via argument mining and stance detection.
- Identifying important concepts for a target user via an attention-based mechanism.
- Deriving an argumentation flow via multiple applications of the attention-based mechanism.
- Unifying the linguistic analyses and the attention-based mechanism.

Linguistic Analyses: Computational linguistic approaches can be distinguished based on granularity of the analysis.

Document-level analyses, for instance, aim at determining a sentiment or stance towards the subject of decision, e.g. a hotel. However, for our purpose such an approach is too shallow as a review generally not only consists of utterances regarding the target item as a whole, but usually also includes remarks on sub-aspects. Sentiments or stances towards these sub-aspects may deviate from the review's overall polarity. For example, a guest may generally like a hotel albeit he or she found the bed uncomfortable. In addition, a review might address sub-aspects not important to the current user. Therefore, we propose a more fine-grained approach relying on aspect-level argument mining that is capable of assigning polarity to a variety of mentioned entities.

Automated argument mining refers to identifying linguistic structures consisting of at least one explicit claim and optional supporting structures such as premises [17]. Depending on which theory of argumentation one follows [21, 10], these structures are more or less specialized. However, analyzing user-generated reviews depicts a difficult task for an argument mining tool as they deviate considerably from texts on which linguistic argument mining is typically performed, e.g. legal text or scientific writing: Reviews are usually shorter, noisy, less densely packed with arguments, and contain arguments in a way that is often not as sophisticated.

Another difficulty arises when one wants to decide whether an extracted argument is in favor or against a particular product. However, a notion of polarity is necessary to be able to select adequate arguments supporting or opposing a recommendation. Therefore, we propose to relate each argument to a stance expressed by the review author. In this regard, a stance target is not limited to the subject of decision itself but may essentially address anything towards which one can have a stance. It follows that the linguistic analyses need to identify the target and establish a relation to the subject of decision. Please note that stance detection involves identifying a subjective disposition that might often be implicit [14]. Moreover, stances, as opposed to e.g. sentiments, do not necessarily integrate a polarity aspect (e.g. "the hotel is modern", "the food is local").

Attention-based Mechanism: Since we are interested in providing users with personalized explanations, solely performing linguistic analyses is not sufficient: Users differ not only in their interests, but have unique styles of decision-making. As a consequence, assessment of the importance of an argument found in a review as well as its accordance with the stances expressed by the review author have to be adapted towards the current user. In order to achieve this, we propose an attention-based approach that considers a vectorial user representation to identify personally important concepts.

Attention-based deep learning approaches [3] have proved to be very successful at identifying significant local features in tasks from several domains such as image processing [24] or machine translation [16]. An attention function can be described as a (soft-)search for a set of positions in a linguistic source where the most relevant information is concentrated. In the context of review-based RS, we propose to use attention as a means to identify important and distinguishing concepts for the subject of decision, i.e. a target item. Technically speaking, we suggest to compute a weighted sum over the sequence of vectors derived from the words contained in a review. The assigned weights would indicate the relative importance of each vector and thus resemble the amount of attention a particular word is receiving.

Personalizing explanations requires attention to be distributed with respect to user preferences, i.e. users act as the context that moderates the (soft-)attention's output. Thus, it is necessary to calculate attention weights subject to a vectorial user representation. We propose to model users analogous to our previous work [8] by embedding one-hot user vectors, along with word vectors, into a densified joint information space. This would allow to numerically estimate the degree to which the current user's preferences are in line with concepts expressed in a review. Assume, for example, the following portion of a review: "The food is good, but beds are uncomfortable." If the user has shown interest in good food in the past, the attention mechanism should assign a large weight to *food*. Although the lack of *bed* comfort might be relevant to others, this argument should be neglected and receive a weight close to zero if the system has not detected a relationship between the current user and this particular aspect before.

Argumentative Flow: Argumentation can be considered more convincing if it consists not only of an aggregation of potentially important words or phrases. We assume arguments to become more understandable and effective in case they follow a coherent argumentative flow, i.e. dependencies between successively extracted arguments. The analogy in our proposed concept is the repeated application of the (soft-)attention mechanism while considering previous output as context. Informally, this mimics a conversational exchange between user and attention mechanism where the latter continuously tries to identify concepts more and more tailored specifically towards the user's preferences. The result would be a succession of relevant word sets that exhibit sequential properties and are thus tied to each other. As a consequence, the output sequence would reflect the whole path of how the system came up with a particular recommendation and thus explain which concepts played an important role during the process (see Figure 1). It must be noted that such a procedure would be closely related to so-called *memory networks* [23, 19], which also work with multiple hop operations on an attention-based memory. Since our concept imposes an artificial argumentative structure on the raw review text, it will be interesting to investigate the argumentative flow intended by the review's author compared to the one automatically derived by the system.

Unification of Linguistic Analyses and Attention: Up to this point, we have covered two independent approaches to process review texts: (1) linguistic analyses aiming at mining arguments and detecting stances, and (2) attention-based extraction of words relevant to the current user. However, both on their own are limited in expressiveness: the linguistic analyses lack the personalization component while the attention mechanism operates on word-level, thus incapable of extracting complete arguments. Consequently, following our superordinate research goal of presenting users more informative, personalized explanations, we plan to exploit the benefits of both approaches by coupling them closely together. A first attempt, for instance, would be to check whether the surrounding context of a word that received a large amount of attention was identified as an argument by the linguistic processor. If this condition is met, the system can interpret the argument as a possible candidate for a personalized explanation. Assume the system detects e.g. the arguments "the food is good" and "beds are uncomfortable", then identifying *food* as an important individual concept should lead to the first argument being chosen for explaining the recommendation.

CONCLUSIONS

In this paper, we have discussed the state of research regarding usage of product reviews in RS—with a focus on explanations. We set our prior work into context, where we already used tags to increase a recommender's transparency and analyzed product reviews in order to provide extended interaction possibilities. Based on this, we pointed out a possible way to exploit this rich information source for presenting users with intelligible, personalized explanations by extracting more complex arguments. We outlined several challenges and proposed a concept to address them. For future work, our research goals are to implement this novel concept, use it to additionally personalize recommendations, and to evaluate it in several domains with a focus on the influence on user experience in comparison to other explanation methods.

REFERENCES

1. A. Almahairi, K. Kastner, K. Cho, and A. Courville. 2015. Learning Distributed Representations from Reviews for Collaborative Filtering. In *RecSys '15*. ACM, 147–154.
2. G. Askalidis and E. C. Malthouse. 2016. The Value of Online Customer Reviews. In *RecSys '16*. ACM, 155–158.
3. D. Bahdanau, K. Cho, and Y. Bengio. 2015. Neural Machine Translation by Jointly Learning to Align and Translate. In *ICLR '15*.
4. C. E. Briguez, M. C. D. Budán, C. A. D. Deagustini, A. G. Maguitman, M. Capobianco, and G. R. Simari. 2014. Argument-based Mixed Recommenders and Their Application to Movie Suggestion. *Expert Syst. Appl.* 41, 14 (2014), 6467–6482.
5. Q. Diao, M. Qiu, C.-Y. Wu, A. J. Smola, J. Jiang, and C. Wang. 2014. Jointly Modeling Aspects, Ratings and Sentiments for Movie Recommendation (JMARS). In *KDD '14*. ACM, 193–202.
6. T. Donkers, B. Loepp, and J. Ziegler. 2016a. Tag-Enhanced Collaborative Filtering for Increasing Transparency and Interactive Control. In *UMAP '16*. ACM, 169–173.
7. T. Donkers, B. Loepp, and J. Ziegler. 2016b. Towards Understanding Latent Factors and User Profiles by Enhancing Matrix Factorization with Tags. In *RecSys '16*.
8. T. Donkers, B. Loepp, and J. Ziegler. 2017. Sequential User-Based Recurrent Neural Network Recommendations. In *RecSys '17*. ACM, 152–160.
9. J. Feuerbach, B. Loepp, C.-M. Barbu, and J. Ziegler. 2017. Enhancing an Interactive Recommendation System with Review-based Information Filtering. In *IntRS '17*. 2–9.
10. J. B. Freeman. 1991. *Dialectics and the Macrostructure of Arguments: A Theory of Argument Structure*. De Gruyter.
11. J. A. Konstan and J. Riedl. 2012. Recommender Systems: From Algorithms to User Experience. *User Mod. User-Adap.* 22, 1-2 (2012), 101–123.
12. Y. Koren, R. M. Bell, and C. Volinsky. 2009. Matrix Factorization Techniques for Recommender Systems. *IEEE Computer* 42, 8 (2009), 30–37.
13. B. Lamche, U. Adıgüzel, and W. Wörndl. 2014. Interactive Explanations in Mobile Shopping Recommender Systems. In *IntRS '14*. 14–21.
14. W.-H. Lin, T. Wilson, J. Wiebe, and A. Hauptmann. 2006. Which Side are you on?: Identifying Perspectives at the Document and Sentence Levels. In *CoNLL-X '06*. Association for Computational Linguistics, 109–116.
15. B. Loepp, K. Herrmanny, and J. Ziegler. 2015. Blended Recommending: Integrating Interactive Information Filtering and Algorithmic Recommender Techniques. In *CHI '15*. ACM, 975–984.
16. T. Luong, H. Pham, and C. D. Manning. 2015. Effective Approaches to Attention-based Neural Machine Translation. In *EMNLP '15*. 1412–1421.
17. R. M. Palau and M.-F. Moens. 2009. Argumentation Mining: The Detection, Classification and Structure of Arguments in Text. In *ICAIL '09*. ACM, 98–107.
18. P. Pu, L. Chen, and R. Hu. 2012. Evaluating Recommender Systems from the User's Perspective: Survey of the State of the Art. *User Mod. User-Adap.* 22, 4-5 (2012), 317–355.
19. S. Sukhbaatar, A. Szlam, J. Weston, and R. Fergus. 2015. End-To-End Memory Networks. In *NIPS '15*. 2440–2448.
20. N. Tintarev and J. Masthoff. 2015. *Recommender Systems Handbook*. Springer US, Chapter Explaining Recommendations: Design and Evaluation, 353–382.
21. S. E. Toulmin. 2003. *The Uses of Argument*. Cambridge University Press.
22. J. Vig, S. Sen, and J. Riedl. 2009. Tagsplanations: Explaining Recommendations Using Tags. In *IUI '09*. ACM, 47–56.
23. J. Weston, S. Chopra, and A. Bordes. 2014. Memory Networks. (2014).
24. K. Xu, J. Ba, R. Kiros, K. Cho, A. C. Courville, R. Salakhutdinov, R. S. Zemel, and Y. Bengio. 2015. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. In *ICML '15*. 2048–2057.
25. Y. Zhang, G. Lai, M. Zhang, Y. Zhang, Y. Liu, and S. Ma. 2014. Explicit Factor Models for Explainable Recommendation Based on Phrase-Level Sentiment Analysis. In *SIGIR '14*. ACM, 83–92.