Crowd-powered recommendation for continuous digital media access and exchange in social networks

FP7-610594

D4.1

Crowd Engagement Algorithms

Dissemination level:	Public
Contractual date of delivery:	September 2014
Actual date of delivery:	Month 12, September 30, 2014
Workpackage:	WP4 Crowd Engagement Algorithms
Task:	T4.1 Multimedia Content analysis for Crowd Engagement T4.2 Reciprocal Recommendation
Type:	Report
Approval Status:	PMB Final Draft
Version:	1.0
Number of pages:	21
Filename:	CrowdRec_WP4_D4.1_TUB_30092014_V1.0.pdf
Abstract	This deliverables summarizes CrowdRec Year 1 activities within WP4, which began in M7. WP4 is dedicated to Crowd Engagement. Algorithms for multimedia content analysis for crowd engagement were developed on the basis of SoundCloud use scenarios. Initial work on reciprocal recommendation is presented. Then, we showcase our initial work on incentivization schemes that motivate the crowd to contribute information to

	content streams. Finally, we provide an overview of our activities planned for the second year.

co-funded by the European Union

History

Version	Date	Reason	Revised by
0.1	July 2014	Creation of template	Frank Hopfgartner
0.2	August 2014	Addition of material on incentivization	TUB
0.5	September 2014	Completion of the material and finalization	Martha Larson
1.0	September 2014	Formatting, abstract and executive summary	Frank Hopfgartner

Author list

Organization	Name	Contact Information
TUB	Frank Hopfgartner	frank.hopfgartner@tu-berlin.de
TUD	Martha Larson	m.a.larson@tudelft.nl
TID	Alexandros Karatzoglou	alexk@tid.es
Gravity	Domonkos Tikk	Domonkos.tikk@gravityrd.com
TUD	Karthik Yadati	n.k.yadati@tudelft.nl
TUB	Michael Meder	meder@dai-lab.de
TUD	Babak Loni	b.loni@tudelft.nl
MOV	Roberto Turrin	roberto.turrin@moviri.com

Executive Summary

This deliverable summarizes our activities within WP4 that were performed during the first 12 months of CrowdRec. The work package, which began in M7 (April 2014), focuses on crowd engagement algorithms that "activate the crowd". We see algorithms that activate the crowd as algorithms that motivate users to perform certain tasks such as explicitly contributing metadata describing items or implicitly providing feedback on relevant items in a set of items.

The work package is split into four tasks. While two of them (T4.1 and T4.2) started at the beginning of the WP at M7 of the project, the latter two (T4.3 and T4.4) start in M13.

We argue that for some domains, a preceding content analysis can be of benefit to provide better recommendations. More precisely, we argue for content analysis in the multimedia domain where the so-called semantic gap hinders the interpretation of multimedia items. Therefore, in T4.1, we focus on multimedia content analysis for crowd engagement. We first developed an approach for discovering time-points in electronic dance music that has been released on the SoundCloud platform. The main aim was to identify a so-called "drop", i.e., the building up of a tension, followed up by the re-introduction of the full bass line. In order to evaluate our method, we first used timed comments on the SoundCloud platforms that mention the term drop as soft label for the appearance of such drop in the sounds. Moreover, we designed a mechanical turk task to annotate the data. Building on our observation that timed comments can be used as labels, we propose a benchmarking task "Crowdsourcing timed comments about music" which is currently run as "brave new task" at the benchmarking campaign MediaEval.

In Task 4.2, we focus on reciprocal recommendation. During the first period of the project, we aimed to bridge stream recommendation and reciprocal recommendation. We outline in this deliverable a general approach on crowd engagement in recommender systems research. Then, we summarize work on recommending users for items. We introduce a novel algorithm, RankSLDA, which extends the supervised latent Dirichlet allocation (SLDA) algorithm with a learning-to-rank component that allows for the exploitation of inherent collaborative effects that are present in question answering communities.

Next, we provide an outlook on Task 4.3 (Incentivize interactive user contributions). This task starts in M13, we have, however, performed some initial studies on the use of elements known from games to motivate users to contribute or annotate content. We performed initial studies on a gamified enterprise bookmarking system, aiming to see which role gamification can play in such enterprise setting. Moreover, we developed the SoundComment system, a serious game that aims to motivate users to annotate timed comments on the SoundCloud platform.

Finally, we provide an outlook onto the activities that we planned for the second year of the project.

Table of Contents

1 INTRODUCTION
1.1 Active Information and Incentivization
1.2 Multimedia Recommendation
1.3 The Reciprocity Revolution
2 MULTIMEDIA AND CROWD ENGAGEMENT
2.1 Crowd-Supported Audio Detection of Events in Music
2.2 Crowd-Validation of Timed Comments
2.3 Derivative Detection in Music Social Networks
3 RECIPROCAL RECOMMENDATION
3.1 Crowd Engagement in Recommender Systems Research
3.2 Recommending Users for Items
4 OUTLOOK ON INCENTIVIZATION
4.1 Designing Incentivization Systems
4.2 SoundComment
4.1 Outlook onto Crowd Engagement Algorithms
5 REFERENCES

List of Figures

Figure 1: A SoundCloud track with user comments. A listener can read the comments as the track plays over the time-point at which the comment was added.
Figure 2: Pipeline for music event detection ("the drop") using timed user comments from SoundCloud.
Figure 3: F1 scores for baseline condition that uses boosting (right bar, red) and for a condition that adds information from user comments to the baseline (blue bar, left).
Figure 4: Screenshot of the Gamified Bookmarking System DAIKnow
Figure 5: Screenshot of the SoundComment system.

1. Introduction

The goal of WP4 “Crowd Engagement algorithms” is to develop algorithms that “activate the crowd” to improve the performance of recommender systems. The work carried out in WP4 makes an essential contribution to allowing the project to achieve its second objective “Crowd Engagement”.

This deliverable summarizes the achievements of CrowdRec related to crowd engagement algorithms during the first year of the project. It focuses on Task 4.1 “Multimedia Content Analysis for Crowd Engagement” and Task 4.2 “Reciprocal Recommendation” which started in M7 of the project (April 2013). It also provides a brief outlook on Task 4.3 “Incentivize interactive user contributions” and Task 4.4 “Recommendation Explanation”, which are due to begin in the second year of the project (October 2013).

This introduction section presents the perspective of CrowdRec on how crowd engagement can, and should, be used to improve the performance of recommender systems. This perspective distinguishes CrowdRec use of crowd engagement from techniques that are currently used in recommender systems. It provides the basis for CrowdRec’s innovative contribution in the area of crowd engagement algorithms for recommendation. We tie the overall CrowdRec perspective to the specific choices made within the project on where to focus and how to best activate the crowd to improve the recommender systems. After the introduction, the deliverable discusses the tasks in turn, and concludes with its outlook on the future.

1.1 Active Information and Incentivization

We understand algorithms that “activate the crowd” to be algorithms that motivate (trigger or inspire) users to contribute information needed by recommender systems. This information can take the form of contributions representing explicit information (e.g., metadata describing items) or of interactions representing implicit information (e.g., relevant preferences among a certain set of items). It can also take the form of validation: verification of information that exists in the system, but for which the confidence is low. In the CrowdRec DoW we refer to crowd engagement algorithms that “activate the crowd to improve recommender systems”. During the first year of the project, the consortium has found the term “active information” to be helpful in expressing this concept, and often refers to recommender systems integrating active information, when discussing or developing crowd engagement algorithms.

A variety of factors trigger users to contribute information. These are referred to as “incentives”. The process of adding incentives to a system in order to elicit user contributions is called “incentivization”. Incentives are categorized into two groups, intrinsic incentives, which include satisfaction in exercising skills and enjoyment, and extrinsic incentives, which includes monetary compensation and awards. CrowdRec investigates both types of motivation.

CrowdRec's use of intrinsic motivation builds on user interest, the fundamental attractive force at work in recommender systems. For example, conventional recommender systems use ratings or clicks, which are produced by user as a result of interest. Using reciprocal techniques, discussed in greater detail below, we reinforce the basic attraction of a user to an item in order to generate additional information, which in turn improves recommendations.

CrowdRec's use of extrinsic motivation builds on gamification techniques. Gamification is an interaction layer, which adds a system of incentives to an underlying system. The incentives are typically classical game mechanics, such as points and badge systems. However, gamification can also include the introduction of supplementary goals, such as profile completion. CrowdRec also makes use of commercial crowdsourcing platforms, where other incentives are combined with an underlying monetary incentive in the form of a micropayment paid in return for performance of a microtask.

1.2 Multimedia Recommendation

CrowdRec's study of user engagement starts from the observation that some domains in which recommender systems are applied a priori appear to have a higher potential to benefit from crowd-contributed information than others. WP4 specifically identifies the area of multimedia content recommendation. Task 4.1 "Multimedia Content Analysis for Crowd Engagement" is dedicated to this domain.

In general, recommender systems suffer from the cold start problem, which is conventionally formulated as follows. Before an item has been rated by users, or before a user rates items, it is impossible to generate recommendations for that user or item. We observe that a priori there are two reasons why the cold start problem is particularly important for multimedia recommender systems.

First, multimedia content analysis can be used to generate metadata descriptions for multimedia items (such as music or video). More recently, it has also been used to make direct signal-to-signal comparisons capable of generating an item-to-item similarity score. The crowd is a very promising source of labels that are needed to train classifiers and comparison algorithms. The crowd can also be activated in order to verify the output of automatic algorithms in borderline cases.

Second, although multimedia may receive enough user interactions at a global level (downloads, listens) to make recommendation, user interaction that is supplied at a time-point specific level remains impossibly sparse. Such information is necessary in order to create recommendation systems capable of recommending certain moments within large multimedia streams (i.e., segments of music or scenes in a video). It is also necessary in order to refine signal-based item-to-item similarities.

Incentivizing users to contribute information on multimedia starts with the insight, set out in the description of Task 4.1 in the CrowdRec DoW, that contributions to specific multimedia items can be triggered, if these items can be matched with users who find them interesting. Building on this insight, we create algorithms that attempt to bring users in contact with multimedia that interests them, and then exploit the labels contributed by these users to identify more multimedia segments that are in need of human-contributed information.

1.3 The Reciprocity Revolution

The fundamental concept that the CrowdRec project uses to achieve Crowd Engagement is *reciprocity*. Simply expressed, reciprocity can be considered a radical broadening of the perspective from which recommender system algorithms are developed. Conventional systems focus on providing a list of items that are useful to a user. CrowdRec systems are concerned with connecting users with useful items, and also connecting items with useful users. We understand a “useful user” to be someone who has the ability, availability and motivation to contribute information that is helpful for the recommender system. That user may belong to the community of users that already use the recommender system, or may also be drawn from another community, e.g., workers on a crowdsourcing platform who are requested to contribute key information that improves recommendations.

Simply stated, introducing reciprocity into a recommender algorithm opens the door for usefulness of recommendation to flow in both ways. We refer to the reconceptualization of recommendation as both inferring useful items and inferring useful users as the “Reciprocity Revolution”. Algorithms developed from the perspective of reciprocity lead to recommender systems that are characterized by a synergy between items to be recommended and the users that consume those items. The ultimate goal of this synergy is to strengthen the overall system, i.e., users that are highly engaged should obtain more satisfaction from the recommendations provided by the system, and that this satisfaction should also improve business aspects of recommender systems, such as clicks, transactions, and retention.

2. Multimedia and Crowd Engagement

In this section, we present the work carried out in Task 4.1 “Multimedia Content Analysis for Crowd Engagement”, which began in M7 of the project. These algorithms were developed on the basis of the business requirements of SoundCloud as reported in D2.1 “First Iteration Requirements plus Evaluation Specifications” (cf. Section 5 p. 28). Specifically, SoundCloud requires a recommender that is capable of continuous play recommendation: i.e., users log in to SoundCloud and are immediately listening to a continuous stream of songs that they find appealing.

These requirements were further refined into a set of technical specifications during the CrowdRec plenary meeting in Barcelona 7 February 2014. Specifically, SoundCloud is interested in technology that uses signal processing as a basis for the calculation of item-to-item similarities (i.e., the level of similarity between two tracks). Such similarities are the underlying technology for the continuous play recommender. SoundCloud also saw further users for them in the future. Further, SoundCloud was interested in the ability to leverage the content of timed comments that users add to SoundCloud tracks. In order to use these comments for recommendation, it is necessary to understand which ones are useful, i.e., have a relationship to the music at that time point, and which ones should be discarded. On the basis of these requirements, we defined two research efforts that are described in this section.

2.1 Crowd-supported Audio Detection of Events in Music

Timed comments are comments that users make at certain points within a continuous media stream. On SoundCloud, a user's comment is represented below the stream with a small version of the user icon added at the time-point where the user made the comment, as illustrated in Figure 1.

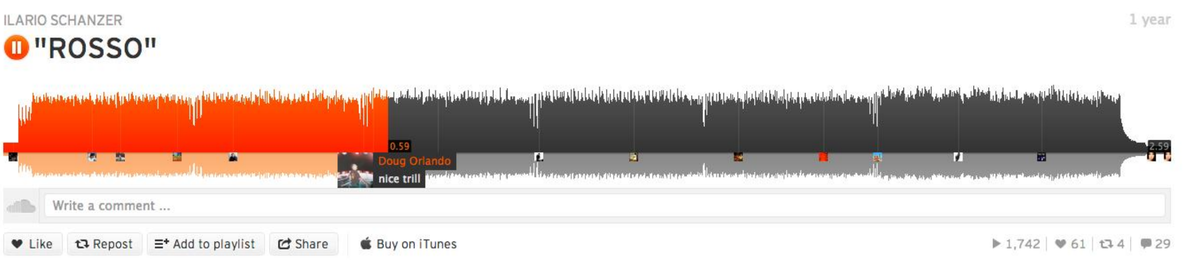

Figure 1: A SoundCloud track with user comments. A listener can read the comments as the track plays over the time-point at which the comment was added.

We point out that use of timed comments is not restricted to music on SoundCloud, but is rapidly becoming more widely spread in video as well, e.g., http://www.metacafe.com and http://www.youku.com, with YouTube offering the same functionality in a slightly different form as so-called deep-link comments. Here, however, we restrict our investigations to music. Specifically we chose the genre of Electronic Dance Music (EDM), which is a popular genre of music, with a large representation on SoundCloud. We chose to focus our work on EDM because of its importance for SoundCloud, but also because it has not been extensively studied in the music information retrieval literature.

The goal of our work is to improve the timed-comments that users contribute on SoundCloud in such a way that they are suitable to support recommender systems. We are interested in two aspects. We are interested in filtering comments to identify those, which are potentially most helpful. The discussion of the SoundComment application in the outlook below provides further information on this direction of inquiry. Second, we are interested in building detectors that are able to discover new time-points in music that possibly should be commented by users, but have not yet been commented. In both cases, our approaches are driven by the fact that they bring music content in contact with contributors who are incentivized to provide input.

In the first five months of T4.1, we developed an approach for discovering time-points in EDM music in which a specific event, called "the drop" had occurred. We are interested in the particular event since it is important to SoundCloud. As an April Fool's joke in 2013, SoundCloud had announced a drop detection functionality[1], which received large resonance from the community. Our initial

[1] http://blog.soundcloud.com/2013/04/01/dropometer/

motivation to study the drop, derived from the obvious significance of the drop as a music event to the user community.

Further motivation was the technical challenge presented by the "drop". A drop is defined as a building up of tension, and is followed by the re-introduction of the full bass line. As such it is an event with a temporal structure, and differs from other events studied in music information retrieval, which have a beginning and an end, but no internal structure. We hope that starting with the drop will allow us to extend our work to other music events that are socially significant, and have similar challenging characteristics.

We designed and implemented a drop detector in a process that made use of crowd input in two ways. First, we used timed comments from SoundCloud that mentioned a "drop" as weak labels in order to identify training data. Second, we created a task on Mechanical Turk that was designed to engage specifically those workers active on Mechanical Turk that have an interest in EDM, and were able to identify the drop.

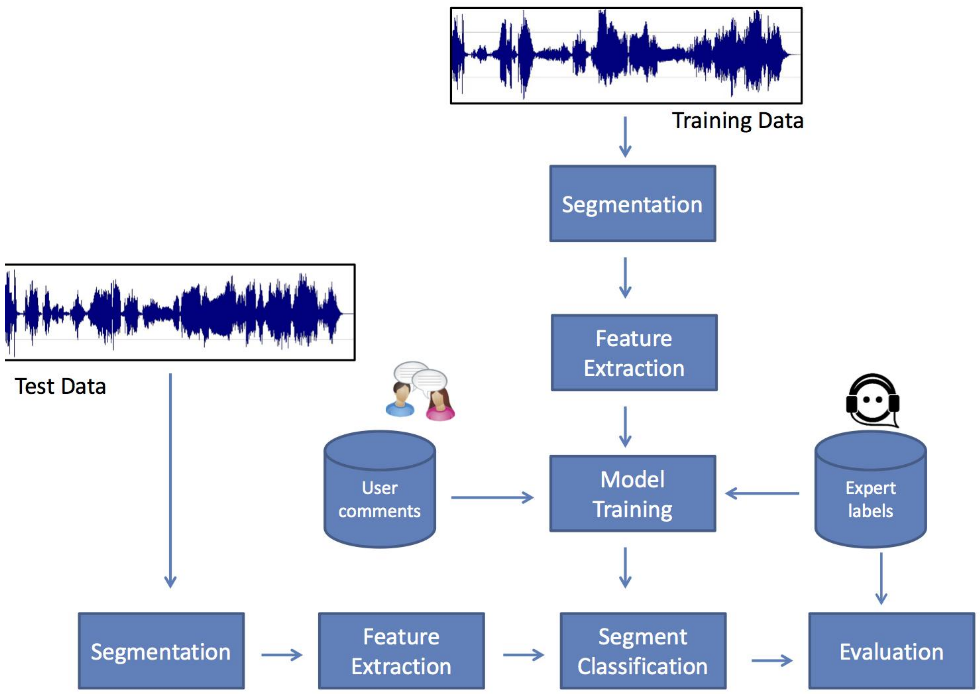

Figure 2: Pipeline for music event detection ("the drop") using timed user comments from SoundCloud.

Our work on using timed user comments followed the pipeline illustrated in Fig. 2. Our drop detection algorithm consists of two steps, first a segmentation that identifies possible drop candidates, and then a model that zeros in on temporal structure. In order to test the contribution of user timed comments to improvement of drop detection, the system is seeded with a small number of expert labels identifying segments of music tracks that contain a drop. These are mixed with a proportion of low fidelity labels derived from user comments also identifying segments of music tracks that contain a drop. The segments are low fidelity because when SoundCloud users mention a “drop”, they might be referring to the ability of the artist, to drops in the track in general, or to a drop event that took place at a moment located quite distant from the time-point at which they placed the comment.

We carried out an experiment that integrated information from user timed comments in a boosting setting. The results are reported in Fig. 3. There are 60 tracks in the training set, with 2-3 drops each, and 20 tracks in the test set. The classifier starts with a given number n of tracks of the training set for which expert labels (high fidelity labels) are provided. It uses the rest of the tracks for boosting. Then it further refines the classifier by adding labels extracted from drop comments (low fidelity labels). In Fig. 3, results are reported for tolerance window size of 11 seconds and number of seed tracks for which high fidelity ground truth was exploited by the classifier: 5 tracks, 10 tracks, 20 tracks, 30 tracks, 40 tracks, 50 tracks.

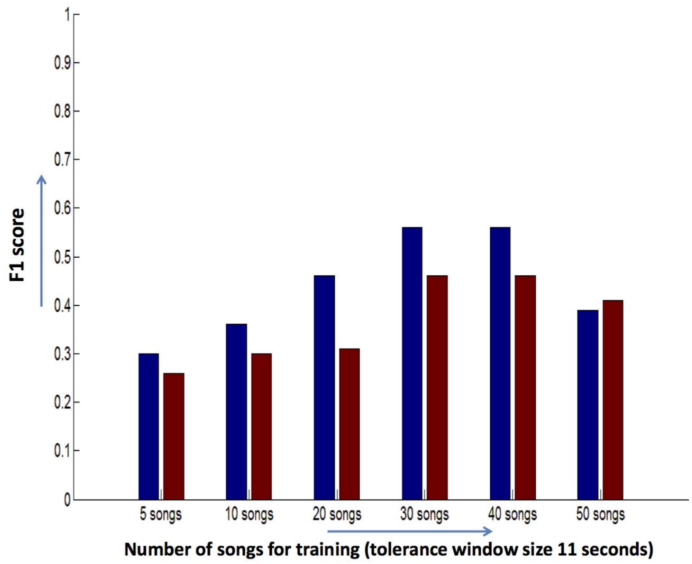

Figure 3: F1 scores for baseline condition that uses boosting (right bar, red) and for a condition that adds information from user comments to the baseline (blue bar, left).

These experiments revealed the high potential of adding information from user-contributed timed comments to improve music event detection. Results were reported in a submission to ISMIR 2014 [1].

During this investigation we established that there was a clear difference between high fidelity labels provided by experts and low fidelity labels derived from user timed comments. We carried out an analysis that revealed that only 20% of timed-comments that mention a drop are located near the actual drop of a SoundCloud EDM track. This means that although timed comments are already useful for identifying a drop, they would be even more useful if user comments could first be validated before being used as labels.

2.2 Crowd-validation of Timed Comments

Our observation that crowd validation of timed comments has high potential to make timed comments useful for recommender systems motivated us to propose a benchmarking task related to the validation of timed comments on SoundCloud. A benchmarking task is a challenge that includes a problem formulation and a data set that is publicly released for members of the research community to solve. The basic goal of our benchmarking task is for participants to develop an algorithm that generates a single accurate label given multiple noisy labels collected using a crowdsourcing platform. The data set for the task consisted of a set of 591 15-second segments in 355 SoundCloud tracks. The tracks were selected from the larger SoundCloud data set that was described in D2.2 “First Reference Framework Release and Evaluation Report” and also mentioned in D5.1 “Basic deployment and report”.

Each segment is associated with three different validations contributed by three different crowdworkers who listened to the segment. The crowdworkers contributions were collected on Amazon Mechanical Turk using a human intelligence task (HIT) that was designed to attract the segment of the MTurk population with expertise in EDM by appealing to their interest in and experience with EDM.

Our benchmarking task is currently running as a task entitled “Crowdsorting timed comments about music” at the 2014 MediaEval Multimedia Benchmark.[2] It is running as a so-called “Brave New Task”, which means that it has high-risk high-reward status within the benchmark. If the initial year is successful, the task could potentially be expanded later. More details are reported in D6.2 “First Year Dissemination Report”.

The dataset for the task was released on 1 September 2014 on the Open Science Framework.[3] Analysis of the results of the benchmark will allow us to determine the extent to which active information acquired on a commercial crowdsourcing platform can be used to complement information contributed by users in the form of comments. The next step is to acquire active information directly from the user-community of SoundCloud. For this purpose, we have developed the concept of SoundComment, described in further detail below.

2.3 Derivative Detection in Music Social Networks

Within the SoundCloud community, musicians share their music, and learn from each other. By nature, the music on SoundCloud is comprised of many derivative works. We define a derivative work as one piece of music being used as the basis to create another piece of music. General examples include one composer borrowing a melodic theme or a style from another. Because

[2] http://multimediaeval.org/mediaeval2014/crowd2014/
[3] https://osf.io/h92g8

SoundCloud music is often produced electronically or using digital editing equipment, examples of derivative works on SoundCloud are more tightly linked at the level of the signal than is true of music in general. This makes the application of multimedia analysis algorithms a particularly promising area of investigation. Derivative works on SoundCloud include DJ-mixes, remixes, mash-ups and playlists.

As mentioned above, SoundCloud sees great potential in algorithms that can make item-to-item similarity predictions on the basis of the signals. Between M7-M12 of the project, we worked on laying the groundwork for a line of research within Task 4.1 that would result in large-scale deployment of track-to-track similarities.

The first step was the creation of a baseline Python implementation of MASK, an algorithm previously developed by TID (see [2]) MASK stands for Masked Audio Spectral Keypoints and is a fingerprinting algorithm capable of encoding acoustic information and discriminating between transformed versions of those documents and other documents. SoundCloud anticipated that the transformations handled by MASK were similar enough to the transformations characterizing many of their derivative works, to make this a productive line of inquiry.

Next, experiments were carried out to determine the extent to which MASK could be applied to the derivative detection challenges most important to SoundCloud. To this end, we gathered a database of 150 documents of length 30s each, approx. 45% general TV/movie audio, 40% music and 15% speech. We transformed each of the audio files using typical acoustic transformations. We used the NIST-TRECVid Video-copy detection transformations. Then, we generated several versions of each transformation, at different SNR levels, mixing the signal with Gaussian noise

Then, a matching system was implemented. The purpose of the system was to identify the original source signal, given one of the target signals (i.e., signals that had been transformed and to which noise had been added). We tested the system with whole segments (30s of signal) and also with 10s of signal, to evaluate the robustness of the matching when less signal is available.

On the basis of the outcomes of experiments with the matching system, a data-driven algorithm was implemented to select region-to-region pairs compares that are then used in the MASK fingerprint to derive the final binary fingerprint. The new algorithm is based on mutual information. This algorithm represents a significant extension of the original system, in which the comparison regions had been manually selected. Evaluation was carried out using our database. Results demonstrated that the accuracy was substantially improved even in very low SNR values.

The initial results are promising, but the work also revealed the critical necessity of crowd contributions in developing signal-based item-to-item matching algorithms. In order to refine the algorithms specifically on SoundCloud data, a large amount of data is needed in which tracks that contain mixes of two songs are hand labeled with the identity of the original songs.

Without such labeled data, signal based matching algorithms cannot be further developed or refined. An initial indication that users could be incentivized to provide such data is the fact that for a few

SoundCloud tracks, users have included this information in the metadata. As such, we conclude that it may be possible to create a system to incentivize users to contribute this information. The basic challenge of this system would be scaling it to so that it can incentivize the collections of the large amounts of high-quality labeled data necessary for further development.
At the moment of writing, the future of derivative works on SoundCloud is not clear. On August 21, 2013, SoundCloud began a new licensing program. The New York Times cites derivative works as being problematic for SoundCloud's new licensing deals.[4] We expect that in a general music retrieval scenario derivative works to remain important. This importance is anchored in the creative process by which artists create new music. It is independent of the status of derivative works on SoundCloud.

3 Reciprocal Recommendation

In this section, we present the work carried out in Task 4.2 "Reciprocal Recommendation", which began in M7 of the project. Early on in the project, we discovered that a major challenge of converting the potential of the "Reciprocity Revolution" into concrete recommender system algorithms was developing concepts within a new modality of thinking about recommender system problems (i.e., recommendation as a two-way street). During the first project period our work in reciprocal recommendation was dedicated to developing concepts, and also to initial work bridging stream recommendation and reciprocal recommendation. We discuss each here in turn.

3.1 Crowd Engagement in Recommender Systems Research

As already discussed in the introduction, CrowdRec's innovation in the area of crowd engagement is based on the idea that recommender systems should not only seek to match items to users, but also they should match users with items.

On one level, the benefits of matching items with users are very clear. For example, Netflix hires part-time workers in order to annotate specific movies, the so-called Netflix taggers[5]. The taggers improve Netflix recommendations, which have been key to the company's enormous success, and are essential to maintain their competitive edge.

On another level, the idea of incentivizing contributions as part of the recommender system, or as the underlying recommender algorithms remains exotic and underexplored in the community of scientists working in recommender system research and development. CrowdRec maintains that there is a subtle difference between algorithms that aim to increase the probability that a user clicks an item, and algorithms that aim to increase the probability that a user click on an item will benefit the system as a whole.

[4] http://www.nytimes.com/2014/08/21/business/media/popular-and-free-soundcloud-is-now-ready-for-ads.html
[5] http://www.forbes.com/sites/jaymcgregor/2014/07/07/netflix-wants-to-pay-you-to-sit-at-home-and-watch-movies/

Since the beginning of CrowdRec (i.e., even pre-dating T4.2) we have worked on laying the conceptual groundwork for the "Reciprocity Revolution" and for formulations of recommender system problems in terms of engagement rather than transactions alone. The result of this work has been two white papers published on the topic, and presented at workshops where the recommender system community is well represented [3],[4]. It has also fed into the conceptualization of the ACM RecSys Challenge 2014, which is described in more detail in D6.2 "First Year Dissemination Report". Here, we include a brief description of how recommender system evaluation interlocks with the development of algorithms that exploit reciprocity and target engagement.

A prerequisite that the "Reciprocity Revolution" takes foothold in the recommender system research community is that the research community also adopts the 3D benchmarking and evaluation model. As discussed in D2.1 "First Iteration Requirements plus Evaluation Specifications" the 3D model is the model used by CrowdRec to inform the process of requirements elicitation and also to design evaluations of recommender system technologies. The 3D model formulates the goals of recommender systems along three dimensions: User requirements, Business requirements, and Technical constraints. All three of these dimensions play a role in determining whether one recommender system algorithm outperforms another.

In conventional formulations of recommender system solution, user satisfaction with recommendation lists is the only factor that plays a role. Solutions to these problems can be evaluated along the user dimension only using standard evaluation metrics such as MAP (mean average precision) or NDCG (normalized discounted cumulative gain). However, in a formulation of a recommender system solution that aspires to take advantage of reciprocity, the user dimension is not adequate to assess the full benefit that reciprocity can bring to the system. If reciprocity works well, we expect user satisfaction with individual recommendations to rise, but we also expect users as a whole to be more engaged in the system. This means that it is also necessary to evaluate the system with respect to business metrics that measure user engagement and retention.

Future work on reciprocal recommendation in CrowdRec will build on these insights and on the outcomes of the RecSys Challenge 2014.

3.2 Recommending Users for Items

In order to tackle the problem of recommendation within social communities, recommending not only users to items, but also items to users, we introduced the RankSLDA algorithm. Its recommendations are reciprocal in that they reflect matches in both directions. The algorithm was developed to address the problem of recommending within question answering fora. However, it is formulation is general enough to be applied to in other social communities. This algorithm is also related to Task 3.2 (Social and Community Stream Analysis), which is discussed in deliverable D3.1 on page 12. An initial publication on the RankSLDA algorithms has been accepted to appear at ACM RecSys 2014 [5].

RankSLDA extends the Supervised Latent Dirichlet Allocation (SLDA) algorithm with a learning to rank paradigm that allows it to exploit the inherent collaborative effects that are present in question answering communities. It uses semantic analysis of the text to create a semantic model incorporating questions, answers and community feedback. The uniqueness of this method is that it models relationships between users and answers, but also between users and other users who answer the same questions. In contrast to other LDA methods, in this method, the topic model must explain. In this algorithm, the users are represented based on their expertise on different topics. The model learns a pairwise rank of users for a given question and generates a ranked list of users for that question.

The algorithm generates a list of users who are ranked in order of their relevance for new questions entering the community. Experimental results confirmed the ability of the algorithm to systematically outperform competitive baseline methods. RankSLDA proved particularly useful for identifying the top ranked user who should answer the question.

The algorithm has a potential impact on recommender systems going beyond the item-oriented evaluation results, however. It has the potential to decrease the rate of unanswered questions in the community, and also reduce the time that people who submit questions need to wait in order to receive answers. The overall affect of the algorithm is also to encourage participation in the community by promoting the involvement of the less active users. Such users fall into the "long tail of participation". They will be encouraged to become more tightly involved in the community, since they receive questions that are specifically targeted to their interests and abilities.

4 Outlook on Incentivization

In this section, we present the foundation that has been laid for Task 4.3 "Incentivize interactive user contributions", which will begin in M12 (October 2014).

4.1 Designing Incentivization Systems

In the next phase of the project, research will be carried out to develop incentivization schemes on the basis of principles of gamification or games with a purpose. Taking game design patterns and principles out of video games to apply them in non-game environments has established as a powerful concept in recent years. Deterding et al. [6] define gamification as *"the use of game design elements in non-game contexts"*. A benefit-oriented definition is provided by Huotari and Hamari [7] who see it as *"a process of enhancing a service with affordances for gameful experiences in order to support users' overall value creation"*. In today's online world, gamification elements are productively employed in many contexts. Stackoverflow uses a reputation leader board and users get points for helpful answers, which they can redeem later to vote on others' answers. LinkedIn motivates users to complete their profiles by presenting progress bars. CrowdRec sees the potential to motivate users to contribute information that can be exploited for recommendation tasks by relying on either game elements.

A common objective of gamification is to enhance motivation. Finding the right means to increase motivation is a non-trivial task. The motivations must be designed so as not to disrupt the users trust relationship with the social networking system or to be overly cognitively taxing or endangering the intuitiveness of interaction with the system. In order to prepare for Task 4.3, we have conducted an exploratory experiment on the impact of the integration of game design elements on users engagement as measured by their contributions to a bookmarking system.

For this experiment we used the DAIKnow Gamified Enterprise Bookmarking system[6] [8]. The system is depicted in Fig. 4. The experiment addressed two research questions: we investigated whether the users of this system perceived gamification as a positive or negative factor in their work environment, and also whether there is a correspondence between a user's perception of gamification, and the users' engagement with the system. The study results revealed that although some employees were already familiar with the idea of gamification and were convinced that it can have a positive effect on their work, a majority of participants stated that they are not convinced that gamification would motivate them.

Figure 4: Screenshot of the Gamified Bookmarking System DAIKnow

Gamification methods have been applied in various environments and for different purposes such as enterprise workplaces, education, pervasive health care, e-commerce, human resource management and many more. Although these studies indicate that gamification can lead to increased user activity, a detailed analysis of users' perception of gamification principles has hardly been studied. We are all individuals and are driven by different input factors such as our personality, as well as social or cultural differences. Especially in an enterprise scenario, it is of uttermost importance to measure

[6] http://daiknow.dai-labor.de

challenges and risks that occur due to these differences before introducing gamification methods though. On the one hand, we expect gamification to increase user participation within an enterprise. On the other hand, the visibility of user interaction (or lack thereof), e.g., the position of the employee on a leader board can increase the stress level of employees or even cause fear that their activities on a gamified system will be used as an indicator of their engagement with the company. We discovered a correlation between a positive attitude towards gamification, and user engagement measured in terms of interaction with the system. This experiment revealed that user perceptions of the incentivization mechanisms used by a system can play a defining role in the power of these mechanisms to incentivize them. It should be noted that the exploratory study was carried out in an enterprise environment, which means that care must be taken when generalizing the conclusions to other environments.

4.2 SoundComment

Building on the observations of Task 4.1 described above, a system was created to collect validation of SoundCloud comments from users. The system is called SoundComment[7] because its purpose is to collect feedback on useful (i.e., "sound") comments the users have contributed to SoundCloud tracks.

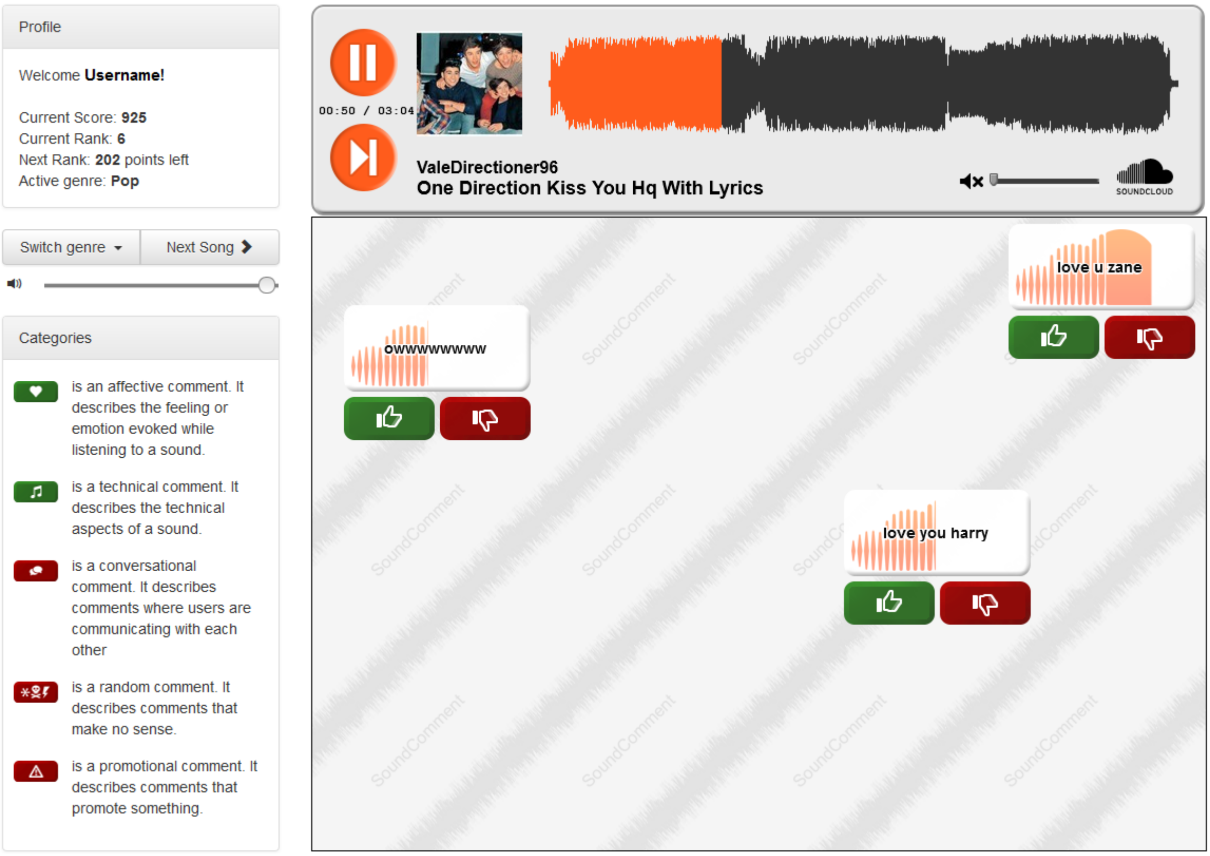

Figure 5: Screenshot of the SoundComment system.

[7] http://soundcomment.nl

The system, shown in Fig.5, allows a user to select a genre and then it plays SoundCloud tracks from that genre. As the track plays, the comments that users have contribute to the track materialize in random positions on the grey striped playing field. Each comment is accompanied by a set of buttons representing different categories into which the user can sort a comment. These categories reflect whether or not a comment contains a description of the track, and if so, what kind of description it contains. The user must click a button before the comment disappears. Each comment is superimposed on top of a SoundCloud cloud, which acts as a status bar. The number of lines left in the cloud indicates how many seconds the user has left to click the correct button for the comment. When a button is clicked, the comment either disappears or a second level of buttons is displayed with further category information. If the system accepts the user answers, it displays the number of points earned at the moment that the cloud disappears.

The system is collecting validations concerning SoundCloud comments. It assigns points based on the level of consensus of the users category with other users who have categorized the same comments. Ultimately, an automatic multimedia analysis algorithm such as the drop detector described above could contribute a confidence that could also help to generate an appropriate score.

This system integrates two game design elements. First, it includes a point system and a leader board that shows where the user stands among other users playing the system. Second, it includes gameplay that requires users to exercise ear/hand/eye coordination. Users must listen read and also click the buttons to earn points. As the user continues playing, the comments appear more rapidly, and more comments are shown at once.

In general, SoundComment makes use of gamification, but also leverages intrinsic motivations. It is designed to appeal to people who are interested in specific music and who derive enjoyment from exercising their expertise by categorizing comments.

4.3 Outlook onto Crowd Engagement Algorithms

In Year 2 of CrowdRec, WP4 will achieve the goals of further extending crowd engagement algorithms. These extensions will build on the foundations established in Year 1. However, the work carried out in WP4 will also be guided by new requirements. Specifically, in M14, a second round of requirements elicitation will be carried out (D2.3 “Second Iteration Requirements”) and in M20, a third round (D2.5 “Third Iteration Requirements”). Requirements are elicited from the industrial partners of the CrowdRec project, including the recommender system vendors, and the large-scale social networks.

Here, we provide a task-by-task summary of the work that is planned next. For Task 4.1, “Multimedia analysis for crowd engagement” the next steps will be to investigate whether the quality of validations that can be collected via SoundComment is comparable with those of expert annotators, and those collected on Mechanical Turk. Initially, we will continue to investigate “the drop”, and gradually expand to include other music events that are socially relevant. We gauge social relevance by analysis of SoundCloud comments to determine which even attract the attention of user. Our guiding hypothesis is that these events will be useful in generating detailed item-to-item similarities that can improve music recommender systems.

For Task 4.2 “Reciprocal Recommendation” we will continue to investigate scenarios in which a tighter match between users and items increases the engagement of users the system. We focus on algorithms that are inspired by the “Reciprocity Revolution”, in other words, they view recommendation as a two way street.

For Task 4.3 “Incentivizing Interactive user contributions” we will focus on systems that engage users to improve news recommendation. We will investigate the news recommendation domain because of the specific challenges that it involves. First, news recommendation stands to benefit from human-annotations of news items. It is difficult for content classifiers to determine the abstract topic of a news item, or the attitude that a news item takes to the specific entities that it discusses. It is also difficult to establish abstract links between news items and related information sources. These are tasks that humans can perform easily, given the appropriate motivation. The shelf-life of any given news item is limited. For this reason, an incentivization system must also be able to motivate users to timely contributions, if is to contribute to increasing the performance of a news recommendation system.

We see two methods to design a gamified news annotation system. One approach, referred to as simple gamification, applies popular game design elements such as leader boards, badges, and points to motivate users to perform a specific task. Another approach uses the utility of a system to users in order to motivate contributions. This type of incentivization is for example, demonstrated by the gamified bookmarking system described above. The system itself has already a utility (link archive, knowledge sharing, information exploration, etc.) for the users, which was supposed to be amplified by applying gamification. In other words, in order to develop a successful gamified system, it is essential to illustrate to the users why they are performing specific tasks such as annotating data. As mentioned above, we aim to study the role of gamification in the context of news recommendation, i.e., the ultimate benefit for users to perform specific tasks is to achieve better news recommendations. Potential tasks that can lead to a better recommendation includes the identification of abstract topics in news articles (i.e., topics that are not directly referred to by the editors), annotation of sentiments in articles (i.e., are topics presented in a positive or negative tone), or can news articles be linked to external knowledge.

The work to be undertaken in Task 4.4 “Recommendation Explanation” will be determined by the requirements elicitation process. CrowdRec holds that explanations for recommendations are essential for crowd engagement because they make it possible to connect with more than one possible motivation influencing user interactions or contributions. The challenge is to create intuitive explanations, which are still rich enough in order to allow users to select recommended items, while at the same time inspiring them in a transparent way to become more engaged with the system.

5 References

[1] Karthik Yadati, Martha Larson, Cynthia C. S. Liem, Alan Hanjalic, Detecting Drops in Electronic Dance Music: Content based approaches to a socially significant music event, The 15th International Society for Music Information Retrieval Conference, ISMIR 2014 to appear.

[2] Anguera, X.; Garzon, A; Adamek, T., "MASK: Robust Local Features for Audio Fingerprinting," Multimedia and Expo (ICME), 2012 IEEE International Conference on , vol., no., pp.455,460, 9-13 July 2012 (Best Paper Award).

[3] M. Larson, A. Said, Y. Shi, P. Cremonesi, D. Tikk, A. Karatzoglou, L. Baltrunas, J. Geurts, X. Anguera, and F. Hopfgartner, “Activating the Crowd: Exploiting User-Item Reciprocity for Recommendation,” ACM RecSys CrowdRec Workshop, Hong Kong, China, 2013.

[4] A. Said, M. Larson, D. Tikk, P. Cremonesi, A. Karatzoglou, F. Hopfgartner, R. Turrin, and J. Geurts, “User-Item Reciprocity in Recommender Systems:Incentivizing the Crowd,” in UMAP Project Synergy (UMAP ProS) Workshop “A forum for UMAP related projects to exchange ideas and practices, 2014.

[5] Jose San Pedro, Alexandros Karatzoglou, “Question Recommendation for Collaborative Question Answering Systems with RankSLDA”, In Proceedings of 8th international Conference on Recommender Systems (RecSys ’14), October 2014, to appear.

[6] Sebastian Deterding, Dan Dixon, R Khaled, and L Nacke. “From Game Design Elements to Gamefulness: Defining Gamification”. Proc. Int. Academic MindTrek Conference, pages 9-15, 2011.

[7] Huotari, K., and Hamari, J. Defining gamification: a service marketing perspective. Proceeding of the 16th International Academic MindTrek Conference (2012), 17–22.

[8] M Meder, T Plumbaum, F Hopfgartner. “Perceived and Actual Role of Gamification Principles”. In Proc. First Workshop on Crowdsourcing and Gamification in the Cloud, pages 488-498, IEEE, 2013.