Cross-system recommendation: User-modelling via Social Media versus Self-Declared Preferences [PRE-PRINT]

Sultan Alanazi
School of Computer Science
University of Nottingham
psxsa16
@nottingham.ac.uk

James Goulding
Horizon Institute
University of Nottingham
james.goulding
@nottingham.ac.uk

Derek McAuley
Horizon Institute
University of Nottingham
derek.mcauley
@nottingham.ac.uk

ABSTRACT

It is increasingly rare to encounter a Web service that doesn't engage in some form of automated recommendation, with Collaborative Filtering (CF) techniques being virtually ubiquitous as the means for delivering relevant content. Yet several key issues still remain unresolved, including optimal handling of *cold starts* and how best to maintain *user-privacy* within that context. Recent work has demonstrated a potentially fruitful line of attack in the form of *cross-system user modelling*, which uses features generated from one domain to bootstrap recommendations in another. In this paper we evidence the effectiveness of this approach through direct real-world user feedback, deconstructing a cross-system news recommendation service where user models are generated via social media data. It is shown that even when a relatively naive vector-space approach is used, it is possible to automatically generate user-models that provide statistically superior performance than when items are *explicitly filtered* based on a user's self-declared preferences. Detailed qualitative analysis of why such effects occur indicate that different models are capturing widely different areas within a user's preference space, and that hybrid models represent fertile ground for future research.

1. INTRODUCTION

Recommendation Systems are highly prevalent on the Web, covering fields as diverse as movies, books, music and academic references [17]. At their heart, Collaborative Filtering (CF) techniques have become virtually ubiquitous as a means of identifying and serving relevant content to any given user. While CF is an extensively researched topic area, several key issues remain unresolved. These include: 1. how to ameliorate the problem of *sparsity* within datasets [24]; 2. how to handle the issue of *cold starts* [28], where systems attempt to recommend content to users who have had little or no prior interaction with the system; and 3. how to ensure user privacy given personalization can quickly distill into a task of monitoring and tracking across users.

ACM ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

Recent researches have proposed a potential solution to these issues in the form of *cross-system* and *cross-domain user modelling* [13, 8]. Here features generated within one system or domain can be used to effectively bootstrap recommendations in another, providing a promising line of attack to handle both sparsity and cold-starts. In this paper, we provide further empirical evidence of the effectiveness of the approach via a focussed study on direct real-world user feedback. This is achieved via implementation of a cross-system news recommendation service, with user model's being automatically generated via social media data (Twitter).

However, in this paper we also ask the question - can a user's interactions with the Web and social media be leveraged in order to produce a cross-system user model that actually out-performs explicit filtering using *self-declared preferences*? And, if so, why? It will be shown that even using a relatively naive vector-space approach, it is possible to automatically generate user-models that provide statistically superior performance than user model's based on *self-declared* preferences. The reasons for these results are qualitatively examined in order to understand why such effects occur, indicating that different models are capturing widely different areas within a user's preference space.

2. BACKGROUND AND MOTIVATION

Recommendation systems have been extensively studied in the research literature, with a multitude of distinct approaches emerging [12, 15, 27]. Collaborative Filtering has proven particularly effective in wide range of applications [23, 10, 5, 14], leveraging the known preferences of a group of users to make recommendations to those whose preferences are only partially observed [25]. A host of mechanisms have been employed to underpin such functionality ranging from Weighted Nearest Neighbor modeling to Bayesian Matrix Completion [18, 22], and with Deep Learning techniques receiving increasing investigation [8]. While CF systems have shown empirical effectiveness in practical settings, several research issues remain, not least the problem of "cold-start" [28, 19, 21]. Cross-Domain Filtering (CDF) has been proposed as a promising solution, bootstrapping recommendations via a user's transactional history in some external system or domain [8]. CDF can be employed when separate item domains exist that share a common set of users [13] or when domains exist that share the same item-set but where 'ratings' are established in different fashions [16]. The field is highly active, with extensions including cross-domain topic modeling [26], cross-domain triadic factorisation [9] and construction of intermediate topic spaces [20].

Despite this promise, CDF does not represent a universal panacea. When applied to CF its assumptions are particularly steep: it requires an extensive user base in both of its domains, as well as sufficient intersection of users and/or items across those domains [7] (further exacerbating well-known sparsity problems that many CF-based systems suffer from [24]). In real-world situations high intersections between user-bases are unlikely - unless situated within niche communities or mass user bases (e.g. Amazon, Netflix, etc.). In addition, transferring user information between such domains generates serious confidentiality and privacy concerns. In this paper, we therefore focus on a Content-Based filtering (CBF) approach, which represents a plausible way to ameliorate many of these issues.

In CBF, recommendations are served by calculating the difference between each item's 'content' (often defined mathematically via some feature set) and some *profile* or model of the user [15]. If such user features can be pre-generated via some cross-domain feature set then not only can cold start be avoided, but a vast intersection of extant users in both domains is no longer required. This combination of CBF and Cross-System approaches already has some precedent [3, 24, 1, 2], and we augment such research by focusing on analysis of direct, real world user-feedback rather than in-sample analysis. Moreover, we investigate whether such an approach can compete with explicitly stated user-preferences - and whether each captures a different, distinct area of the user's preference space.

3. EXPERIMENTAL PLATFORM

An experimental platform was developed to investigate the specific research question: 'can a cross-system user model mined from *social media* generate more accurate recommendations than explicitly stated user preferences?'. If true, this would provide contributing evidence as to the effectiveness of cross-system modelling and the potential value of passive mining of web behaviour. The platform was setup to allow:

- Construction of user-models through: 1. an n-gram vector-space representation derived from social media streams; 2. explicitly defined declaration of categorical user interests; and 3. random parameterization (to serve as a baseline for our testing procedure).
- Application of these models to dynamically rate the relevance of articles supplied to it via any RSS feed.
- Delivery to a user of the most relevant articles based on one of the above models, presenting items via a web interface that allowed for relevance feedback ratings.

As illustrated in Figure 1, functionality is split into 5 *modules*. The **content module** retrieves item sets via external RSS feeds and, using appropriate features, constructs a compressed model for each item (see §3.1). In parallel, the **user modeling module** hosts (potentially multiple) user models for the current participant and, in the case of this research, is also tasked with constructing these models (see §3.2). With this data in place, the **recommendation module** then performs similarity calculations between output items and user models (see §3.3), serving feeds to the **presentation module** for user-evaluation to occur. Once this evaluation is provided, an anonymized log is recorded in the **logging module** ready for post analysis.

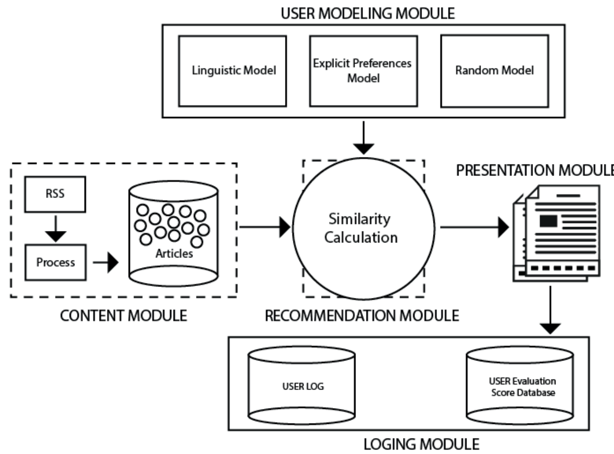

Figure 1: Architecture of Experimental Platform

3.1 Document Modelling

Based as it is on content recommendation, the system currently uses a traditional vector space model (VSM) to represent documents. The content module, first extracts a series of documents from the content source via an RSS feed endpoint transforming each into a corresponding vector space model, $\vec{d}$. For each $\vec{d}$, the VSM iterates through every n-gram it contains, assigning each a weighted importance value, $W_{t,d}$. A range of weighting schemes have been proposed for such this value [4], but we employ the relatively naive *Term Frequency-Inverse Document Frequency* (tf-idf) statistic due to its empirically proven effectiveness [4]. We note that the tf-idf score for any particular n-gram, t, takes: 1. a higher value, if t occurs numerous times within a small number of documents; 2. a low value if t appears fewer times in a document *or* occurs in many documents and 3. a low value, if t occurs in almost all documents. In practical applications the set of features, $\mathcal{F}$, can become extremely large. This leads not only to issues of sparsity and computational efficiency, but also reduce effectiveness of similarity comparisons due to curse of dimensionality. It is therefore desirable to reduce the dimensions of the vector space by removing irrelevant and redundant features [6][1].

3.2 User-Preference Modeling

In order to support comparisons, the experimental system must be able to generate multiple user models for the same individual. While produced via different techniques, the following models are simultaneously constructed for each user when he/she first registers with the system (as represented by the *user-modelling module* in Figure 1):

PASSIVE user-modelling: On registration the user must provide controlled access to an active social media account, which is then mined to generate a linguistic preference model. Users' social media posts are collated, parsed into a bag of words representation (an n-gram frequency model) of their term usage, cleansed and finally encoded as a vector-space model, again using tf-idf.

MANUAL user-modelling: Here user preferences are modelled via presentation of a pre-constructed set of document

[1]In practice this was achieved via functionality available in the python scikit-learn libraries (http://scikit-learn.org).

category and sub-category labels, asking the user to rate their level of interest in each (either selected didactically, or generated by parsing the corpus and modelling the k most common topics [26]. In this case data must be actively supplied by the user through a web interface when they first register with the system.

RANDOM user-modelling: In order to provide a baseline, the system generates a random preference-model for each user. This is a VSM containing the same dimensionality as the passive user model, but with tf-idf scores set randomly for each feature in $\mathcal{F}$. Use of this model should therefore produce random recommendation results.

Once models have been established, the documents in dataset $\mathcal{D}$ are ranked according to their relevance to the user (yielding three document rankings).

3.3 Determining Document Relevance

Recommendation occurs when users begin interacting with the system. Participants are presented with a stream of n documents, each in turn and each generated by *one* of the available user models (which model is selected for each recommendation is specified by the testing regime). To achieve this, prior to presentation the recommendation module must generate multiple rankings of all documents in D, one for each user model that may be used during the experiment. For categorical models, documents are ranked according to the number of labels each individual document is tagged with that match the user's explicitly declared categories of interest. For vector space models (i.e. the Passive and Random user-modelling approaches detailed in §3.2), each document, d is ranked by calculating the similarity between its VSM, $\vec{d}$, and the user's preference-model, $\vec{p}$. The relevance score used here is the traditional *cosine similarity* measure:

$$similarity(\vec{d},\vec{p}) = \frac{\vec{d}.\vec{p}}{||\vec{d}||\vec{p}||} = \frac{\sum_{i=1}^{|\mathcal{F}|} w_{d,i}w_{p,i}}{\sqrt{\sum_{i=1}^{|\mathcal{F}|} w_{d,i}^2}\sqrt{\sum_{i=1}^{|\mathcal{F}|} w_{p,i}^2}} \quad (1)$$

3.4 Presentation and Logging

When a user interacts with the system, articles are presented to them via a web interface as follows: 1. First one of the system's user-models detailed in §3.2 is selected, the distribution of these selections being stochastic and/or defined by the experimental regime used; 2. For the model selected, the highest recommended document not yet viewed is identified, retrieved via its associated URL and presented to the user with a 7-point Likert scale for relevance evaluation; and 3. The user's assessment rating is then collected and stored in a database for analysis, and the process iterates with a new user-model and a new article presented.

4. EXPERIMENTAL METHOD

In our experiments passive-user models were generated using a user's *Twitter* stream and the *BBC news RSS feed* provided our output documents. First the system's content module retrieved data from the BBC news feed[2] extracting 2180 articles. Each document was transformed into a corresponding VSM, and stored along with category meta data and the source article's URL. Simultaneously explicit category labels were extracted from the feed: *Technology*, *Science*, *Environment*, *Entertainment*, *Arts*, *Education*, *Family*, *Health*, *Politics*, *Business*, *UK*, and *World*.

40 participants from a wide range of backgrounds were registered with the system. Each participant was required to be an active Twitter user and to have posted a minimum of 150 tweets. The maximum number of tweets of any user since signing up to the service was 46,700, the mean was 6479 and the standard deviation 10,332. On registration each user authenticated the system's Twitter application and explicitly declared categorical interests via the web interface (as selected from the tags used by the BBC news feed). This allowed *Manual*, *Passive* and *Random* Models to then be automatically constructed for each user. In order to construct the passive model, the user's 150 most recent tweets were extracted via the Twitter API[3].

All participants then engaged in a lab-based task experiment where the system presented them with a total of 45 articles[4], whose relevance they were asked to evaluate in sequence. To choose the next article to be presented, the system selected a random user-model from the three available, determined that model's highest ranked unseen document, and presented it to the user along with a Likert evaluation scale. In cases of a tie between a set of documents in any ranking, one was selected at random. Each user was ultimately presented with an equal number of articles for each model. Thus while ranking is deterministic for all models used, each experimental run would still be stochastic in nature in terms of the ordering of articles presented. For each document presented the system recorded the participants *user-id*, the presented article's *document-id* of the article presented, the user's *evaluation score*, and an identifier of the model used to select that article. At the end of this experiment, we were able to test three hypotheses against the 1800 data points produced (40 people × 45 ratings). These hypotheses were: H1. MANUAL vs. RANDOM (i.e. a hypothesis aimed at determining that any sort of filtering is better than simply serving random BBC news articles); H2. PASSIVE vs. RANDOM; and H3. PASSIVE vs MANUAL (i.e. aimed at determining if an implicit filtering approach is superior to one based on explicit statements of preference).

5. EXPERIMENTAL RESULTS

Results of experiments indicated that the PASSIVE model generated the most relevant news item recommendations for users in comparison to both MANUAL and RANDOM models. The mean relevance scores recorded for our baseline model (RANDOM) was 3.81 points. Recommendations generated via automated preferences models (MANUAL) were rated at an average of 4.13 across 600 evaluations. The mean rating of implicit/linguistic filtering model (PASSIVE) was 4.30. In 85% of cases, a user-model improved over the baseline random recommendation. Standard deviation of evaluations was relatively low for all models at 0.912, 0.850 and 0.814 for RANDOM, MANUAL and PASSIVE models respectively. In general, most of the articles selected by the PASSIVE model were scored highest by users - however this was not the case across the board (and in rare cases random selection was favoured). Also of note, was a surprising lack of correlation ($r = 0.05$) between model performance and the number of tweets the participant had posted over

[2] via http://feeds.bbci.co.uk/news/rss.xml

[3] https://dev.twitter.com/rest/public

[4] undertaken in 3 rounds for a more palatable experience.

their lifetime. Long term Twitter activity (which we view as a proxy for Twitter experience) did not seem to have any impact on the performance of the PASSIVE model.

5.1 Wilcoxon Signed Ranks Test

A Wilcoxon Signed Ranks Test [11] was run to determine whether there was a statistically significant mean difference between models (full details are provided in Table 1). Participants logged a mean evaluation of 4.13 for MANUAL and 3.81 for RANDOM. Test results produced a value of $p = 0.002$, indicating that the 0.32 increase was statistically significant (Z = -3.127, p = 0.002). Thus we were able to conclude from hypothesis H1 that the MANUAL model was producing superior performance. Similarly, for hypothesis H2 participants expressed preference for PASSIVE models (4.30) as opposed to the RANDOM baseline (3.81); a statistically significant increase of 0.486 was discovered (Z = -4.098, p = 0.000042, p < .05). Finally, we were also able to show a statistically significant preference for the use of passively mined personal information via the PASSIVE model (4.30) as opposed to the MANUAL filtering (4.13), with an improvement of 0.17 (Z = -2.045, p = 0.041), p < .05.

Table 1: Wilcoxon Signed Ranks Test Statistics

	M - R	P - R	M - P
Z	-3.127	-4.098	-2.045
Asymp. Sig. (2-tailed)	0.002	0.000042	0.041

model R = RANDOM, M = MANUAL, P = PASSIVE

Table 2: Ranking Results

		N	Mean Rank	Sum of Ranks
M - R	Negative Ranks	11	16.14	177.50
	Positive Ranks	29	22.16	642.50
	Ties	0		
	Total	40		
P - R	Negative Ranks	9	8.89	80.00
	Positive Ranks	28	22.25	623.00
	Ties	3		
	Total	40		
M - P	Negative Ranks	14	21.44	262.50
	Positive Ranks	26	18.75	557.50
	Ties	0		
	Total	40		

model R = RANDOM, M = MANUAL, P = PASSIVE

Results indicate that use of passively mined personal information produced the most effective models in our tests, with Table 2 providing a breakdown of how each model compared in terms of inferred user rankings.

6. DISCUSSION AND POST-ANALYSIS

Results correspond to the intuition that generating a user model, whether based on implicitly or explicitly defined preferences, can play an important role in cross-system recommendations. These improvements illustrate the ability of a model that is using just 20,000 *characters*, based on relatively straight-forward VSM techniques and drawn from a completely different domain to obtain positive results. In this section we qualitatively investigate the reasons for this apparent effectiveness - *why* is it producing the results it is?

With the PASSIVE model having produced the highest evaluation scores, we first examined how it considered articles that were recommended by its rivals (in order to shed light on its selections). On average the cosine score it gave its own recommendations was 0.09642, but for MANUAL recommendations it would have given 0.00881 and for RANDOM 0.00155 (for an example of these differences see Figure 2). This means that, as far as the *Twitter generated VSM model* was concerned, those were performing at an order of magnitude worse than itself. Given that MANUAL achieved statistically significant improvements over RANDOM, we infer from this that the preference information that the PASSIVE model is able to capture via Twitter is wholly *distinct* in nature to that established via an explicit statement of categorical preferences - which strongly suggests that improved recommendations might be achieved by combination of explicitly declared and passively mined preferences.

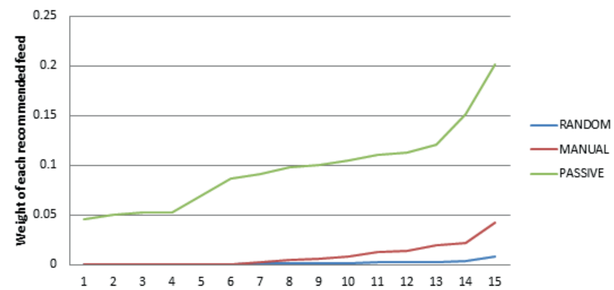

Figure 2: Interpretation of recommendation relevance for all models, based upon the Twitter extracted VSM for *Participant 38* (n.b. articles have been ordered with respect to their cosine value).

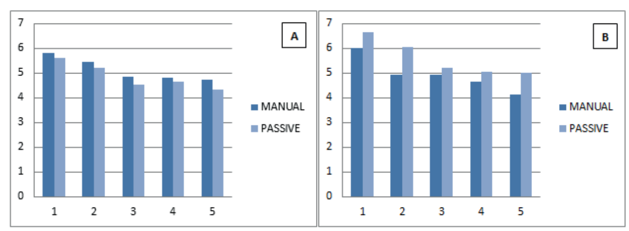

Figure 3: Mean evaluation scores for the [A] Top 5 users favouring MANUAL models and [B] Top 5 users favouring PASSIVE models.

It was noted that there were 5 cases where participants exhibited an overall preference for recommendations made by the MANUAL model. This could be for several reasons: perhaps the labels offered to them during explicit user modeling better expressed their interests than for other participants. It may have been the converse, with their tweets not sufficiently expressing the diversity of their interests. Alternatively it could simply be because their preferences were very focused (e.g. to one category). Closer investigation of the users who most favoured explicitly generated recommendations, however, showed both multiple category selection and ratings statistically indistinguishable from those they gave the PASSIVE model (this proximity is illustrated in Figure 3). From this it seems most likely that these are indeed cases of participants being well matched to the ontology of labels supplied rather than passive models failing them.

This contrasts, however, with the 5 users who most favoured passively generated recommendations, where the difference between that model and their MANUAL evaluations was indeed significant (again see Figure 3). We infer from this that even though, in general, Twitter modelling is giving us different informational value, the information it is producing is still sufficient to overcome anything lost by not incorporating explicitly stated preferences.

Investigation of the articles missed by some models showed that, in particular, the Twitter model could make extremely niche recommendations that cut across broad categories. For example, *Participant 4* declared interest in *Science & Environment*, *Politics* and *Tech* tags, evaluating those MANUAL recommendations at 4.1 (compared to 3.1 for RANDOM). However, a third of his/her PASSIVE model's recommendations did not include these labels - and yet were still awarded an even higher evaluation of 4.5. We infer from this that unlike explicit filtering, which is necessarily bound to some pre-defined ontology of labels, passive filtering was detecting more personalised, specific article recommendations that cut across categories.

For many participants MANUAL preference selection appears consistently too coarse to reflect the subtleties of individual user preferences. *Participant 19* serves as an example of this. He/she was presented with several news articles tagged with the 'UK' label by both PASSIVE and MANUAL models. Yet, those served by the MANUAL model received ratings of only 3.66, compared to 5.50 for those served by the PASSIVE model. This represented a commonly identified theme where a participant was indeed accurately identifying an interest in UK articles, but was unable to specify that it was a specific subset of these that held most interest.

Missing articles due to not manually selecting the superset that a label represented was another common theme. An example of this occurred in *Participant 30*, who highly rated two medical articles recommended by the PASSIVE model (evaluating them both with a score of 6), despite stating no general interest in *health* categorized items. Because this pattern was so frequent, we present some specific examples in closer depth. *Participant 22* stated that items tagged as 'World' (i.e. non-UK) events were not of interest him/her. However, when the PASSIVE model created a VSM via his social media posts, it found several hits to items with the 'World' tag. Two illustrative examples, which the user evaluated as having high relevance to them, were:

news item 1: *"The moment Nepal's earthquake hit my home"*

news item 2: *"The day my generation will talk about for the rest of our lives"*

Investigating the participant's VSM indicated that the similarity was being expressed due to a high tf-idf score for the features 'Nepal' and 'aid', and this was corroborated by the detection of posts in his Twitter timeline expressing empathy for the region following recent natural disasters. A similar situation occurred for a user who indicated a high evaluation for a 'UK' tagged news item (which referenced the UK soccer team, *Chelsea*), despite stating that he no preference for UK specific stories. That article was:

news item 3: *"Why Chelsea won the league, by Alan Shearer"*

Exploration of the participant's VSM initially drew a blank, showing no indication of a high expression for any terms related to *Sport*, *Football* or *Soccer*. However, further investigation identified a tweet on the participant's timeline that referenced "Mourinho" (the coach of the Chelsea Football team at the time of the experiment). This term did indeed have a high tf-idf expression in the document's vector-space model as well as a high activation in the participant's PASSIVE model, consequently resulting in its recommendation. We note that this granularity would be impossible to achieve via explicitly stated preferences. However, we also noted that for some participants their Twitter hashtags (parsed as n-grams by the PASSIVE model) themselves served as forms of folksonomic tagging and expressions of highly granular categorical interest. An example of this was *Participant 32* who frequently used the hashtags: *#bigdata, #datascience, #analytics and #IoT*. As a result the participant was recommended the following article via the PASSIVE model:

news item 4: *"Why measure feet with iPads?"*

This article discussed how a shoe retailer had introduced tablet devices to automate measure and capturing of invaluable data about their customers' feet. Despite no apparent relevance to any of their other interests, the user assigned the item a score of 6. Because the story was tagged with the label 'Health' it was overlooked by the MANUAL model.

These fine grained investigations drew us to several conclusions concerning the behaviour of the PASSIVE approach, and its divergence from self-declared preferences: 1. Participants did indeed Tweet about things they were interested in reading about, allowing the PASSIVE model to pick up true positives; 2. The dynamic nature of both Social Media posting and News articles meant that collation of data from a constrained time window was appropriate; 3. highly relevant content-based recommendations identified by the PASSIVE model can be easily missed by the MANUAL model if tagged with an over-generalized label; 4. any universal taxonomy for explicit statement of preferences appears unfeasible; 5. while PASSIVE modelling via Twitter produced statistically superior results to MANUAL models, the two approaches appear to be capturing different forms of preference information. From this we concluded not only that a hybrid model would produce improved results, but that generating a user model from numerous combined domains (Web search logs, Twitter posts, Facebook usage, etc.) would like produce even more effective functionality.

7. CONCLUSION

In this paper, we have investigated via direct user-feedback, the effectiveness of a recommendation system based on personal data stream information that combines the advantages of both Cross-System and Content-Based Filtering. A cross-system user model was constructed by mining Twitter data streams, and its performance corroborated via real world user assessments of BBC news recommendations. We showed not only 1. the viability of harnessing linguistic vector-space user models generated from social media data, but also 2. that this automated cross-domain approach can actually be superior to explicit filtering using self-declared preferences. However, post-analysis also indicated that these two approaches were capturing *different* information and there is fertile ground in combining the two mechanisms. While there is much opportunity to improve the complexity of the linguistic model used to represent user preferences, there is equal potential in integrating passive models generated from different data streams (e.g. Facebook, Web Search logs, Product purchase descriptions, etc.), each with its own window into users' interests and preferences.

8. ACKNOWLEDGMENTS

This work was jointly supported by the RCUK Horizon Digital Economy Research Hub grant, EP/G065802/1 and the EPSRC Neodemographics grant, EP/L021080/1.

9. REFERENCES

[1] F. Abel, Q. Gao, G.-J. Houben, and K. Tao. Twitter-based user modeling for news recommendations. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, pages 2962–2966. AAAI Press, 2013.

[2] F. Abel, E. Herder, G.-J. Houben, N. Henze, and D. Krause. Cross-system user modeling and personalization on the social web. *User Modeling and User-Adapted Interaction*, 23(2):169–209, 2013.

[3] A. Ahmed, A. Das, and A. J. Smola. Scalable hierarchical multitask learning algorithms for conversion optimization in display advertising. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, WSDM '14, pages 153–162, New York, NY, USA, 2014. ACM.

[4] A. Aizawa. An information-theoretic perspective of tf–idf measures. *Information Processing & Management*, 39(1):45–65, 2003.

[5] R. M. Bell and Y. Koren. Improved neighborhood-based collaborative filtering. In *KDD-Cup and Workshop*, pages 7–14. ACM press, 2007.

[6] G. Costa and R. Ortale. Xml document co-clustering via non-negative matrix tri-factorization. In *Tools with Artificial Intelligence (ICTAI), 2014 IEEE 26th International Conference on*, pages 607–614, Nov 2014.

[7] P. Cremonesi and M. Quadrana. Cross-domain recommendations without overlapping data: myth or reality? In *Proceedings of the 8th ACM Conference on Recommender systems*, pages 297–300. ACM, 2014.

[8] A. M. Elkahky, Y. Song, and X. He. A multi-view deep learning approach for cross domain user modeling in recommendation systems. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15, pages 278–288, Republic and Canton of Geneva, Switzerland, 2015. International World Wide Web Conferences Steering Committee.

[9] L. Hu, J. Cao, G. Xu, L. Cao, Z. Gu, and C. Zhu. Personalized recommendation via cross-domain triadic factorization. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, pages 595–606, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee.

[10] P. Knees, D. Schnitzer, and A. Flexer. Improving neighborhood-based collaborative filtering by reducing hubness. In *Proceedings of International Conference on Multimedia Retrieval*, ICMR '14, pages 161:161–161:168, New York, NY, USA, 2014. ACM.

[11] L. M. LaVange and G. G. Koch. Rank score tests. *Circulation*, 114(23):2528–2533, 2006.

[12] J. Lee, S. Bengio, S. Kim, G. Lebanon, and Y. Singer. Local collaborative ranking. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, pages 85–96, New York, NY, USA, 2014. ACM.

[13] B. Li. Cross-domain collaborative filtering: A brief survey. In *Tools with Artificial Intelligence (ICTAI), 2011 23rd IEEE International Conference on*, pages 1085–1086. IEEE, 2011.

[14] H. Liu, J. Goulding, and T. Brailsford. Towards computation of novel ideas from corpora of scientific text. In *Machine Learning and Knowledge Discovery in Databases*, pages 541–556. Springer, 2015.

[15] Manisha Hiralall. *Recommender systems for e-shops.* Msc dissertation, Vrije Universiteit, 2011.

[16] W. Pan, E. W. Xiang, N. N. Liu, and Q. Yang. Transfer learning in collaborative filtering for sparsity reduction. In *AAAI*, volume 10, pages 230–235, 2010.

[17] D. H. Park, H. K. Kim, I. Y. Choi, and J. K. Kim. A literature review and classification of recommender systems research. *Expert Systems with Applications*, 39:10072–10059, 2012.

[18] J. D. M. Rennie and N. Srebro. Fast maximum margin matrix factorization for collaborative prediction. In *Proceedings of the 22Nd International Conference on Machine Learning*, ICML '05, pages 713–719, New York, NY, USA, 2005. ACM.

[19] Y. Rong, X. Wen, and H. Cheng. A monte carlo algorithm for cold start recommendation. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, pages 327–336, New York, NY, USA, 2014. ACM.

[20] S. D. Roy, T. Mei, W. Zeng, and S. Li. Socialtransfer: cross-domain transfer learning from social streams for media applications. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 649–658. ACM, 2012.

[21] S. Sahebi and P. Brusilovsky. Cross-domain collaborative recommendation in a cold-start context: The impact of user profile size on the quality of recommendation. In *User Modeling, Adaptation, and Personalization*, pages 289–295. Springer, 2013.

[22] R. Salakhutdinov and A. Mnih. Bayesian probabilistic matrix factorization using markov chain monte carlo. In *Proceedings of the 25th International Conference on Machine Learning*, ICML '08, pages 880–887, New York, NY, USA, 2008. ACM.

[23] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pages 285–295. ACM, 2001.

[24] B. Shapira, L. Rokach, and S. Freilikhman. Facebook single and cross domain data for recommendation systems. *User Modeling and User-Adapted Interaction*, 23(2-3):211–247, 2013.

[25] X. Su and T. M. Khoshgoftaar. A Survey of Collaborative Filtering Techniques. *Artificial Intelligence*, pages 1–19, 2009.

[26] J. Tang, S. Wu, J. Sun, and H. Su. Cross-domain collaboration recommendation. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '12, pages 1285–1293, New York, NY, USA, 2012. ACM.

[27] F. Xia, N. Y. Asabere, A. M. Ahmed, J. Li, and X. Kong. Mobile Multimedia Recommendation in Smart Communities: A Survey, 2013.

[28] D. Zhang, C.-H. Hsu, M. Chen, Q. Chen, N. Xiong, and J. Lloret. Cold-start recommendation using bi-clustering and fusion for large-scale social recommender systems. *Emerging Topics in Computing, IEEE Transactions on*, 2(2):239–250, June 2014.