OAUC's participation in the CLEF2015 SBS Search Suggestion Track

Joachim Fugleberg[1] and Michael Preminger[2]

[1] The National Archives of Norway
[2] Oslo and Akershus University College of Applied Science

Abstract. In this article we describe the OAUC's participation in the CLEF 2015 SBS Search Suggestion track. We are trying to represent appeal elements, used in readers' advisory theory and practice, to see if they can be used in an automatic retrieval and recommendation context. We are starting out with the pace appeal element, used on fiction to representing how quickly a buildup of the story is. The results so far indicate that much tuning is needed when building models that can represent pace.

1 Introduction

There are many qualities to books besides their formal characteristics, such as title, author and subject (the latter being examples of metadata). Books, particularly fiction, also evoke the readers' emotions, which is arguably their major mission. This article explores how emotions and other subtle qualities can be discovered in user generated data and subsequently used in a system for automatic classification of books, as a part of an automatic recommender system. For this year's task we try to measure the performance of book suggestion based on a number of emotion evoking characteristics. The challenge is twofold: try to identify certain emotion waking characteristics of books, and measure whether identification of such characteristics helps us match readers' wishes based on similar characterization of their recommendation requests. We see it as a start of a process where we try to operationalize diverse subtle properties (appeal elements) known from literature promotion, so that they can be used as extra evidence in recommender systems.

2 Theoretic Approach

Emotion evoking characteristics are properties of books that are not usually a part of the metadata, technically, because they are difficult to trace back. Even though most people might agree on one or the other subtle property of a book, there is potential for dispute, because they characterize the relation between a book and a reader [1], and readers are different.

2.1 Saricks framework of appeal

[2] has developed a framework / terminology that enables librarians, or other reading-promotors, to discuss books through short excerpts, user reviews and the like, boiling down to "appeal". Appeal has a number of elements

Pace According to Saricks, *pace* is the most important appeal-element, and has the best potential of distinguishing potential readers. Pace has to do with the build up of the story / plot in a book, and how quickly the reader is drawn into it. Some readers (in some situations) will prefer fast paced books, other will rather endeavour on a slow-paced book. [3] Also have a

Characterization This element has to do with the introversy or extroversy of the characters in the book. Readers often remember the characters in the book more easily than they remember the plot. Alas, the conception of a well developed character varies greatly among users[3], making this element less hospitable to analysis of appeal than the case is for the pace element.

Frame The frame is about the tone of a book (melancholic, positive), its feeling (funny or romantic), and its atmosphere (menacing or elevating). though difficult to define, this element is often decisive for the reader's choice.The book can be amusing, bleak, bittersweet[4]

Storyline The storyline is of course dependent on the previously discussed elements. But typical values[5] will be Issue-oriented, Nonlinear or Open-ended.

2.2 Representing and modelling appeal elements

The appeal elements are not directly manifest in the book text, let alone its metadata, and we need to find some representation so that a recommender system can take them into account. To this end we need to find some manifest indicators that can automatically match a recommendation request and a book using the appeal elements as evidence (in addition to other evidence), when recommending a book based on this recommendation request.

Finding and using such indicators is a challenge, which character differs among the elements. Being metadata of different kinds rather than full content, the texts we have are sparse, but on the other hand (for a portion of the books) include reader reviews, which should be a condense summarization of the book done by readers, the target group of a recommender system.

---

[3] [3] lists 30 types of characters that can appear in fiction
[4] [3]lists 58 categories of "Tone".
[5] [3] lists 9 types of storyline

One way of representing an appeal element (element-name element-value) is, using occurrences of sentences that are characteristic to some value of an appeal element. Feeding these to an NLP-system, The NLP system may identify functionally similar sentences in any analyzed book-review to use in the classification of the books. In our implementation, a model of an appeal element is a summary of sentences that are likely to appear in a review of a book that has this value (or valence) of this element. This method has a potential for accuracy, but needs quite a large set of reader reviews given to books with known values (and valences) of the appeal element, and is extremely prone to overfitting. A simpler but less exact method will be identifying single words or word combinations used by readers when reviewing books of different values / valences of appeal elements. Such words need somehow to be classified, so that a system looking for appeal elements in reviews has a broader repertoire of words to look for than the one occurring in the training set.

Matching can thereafter be done by attempted applying the same, or a slightly different model to the recommendation request, assuming that a recommendation request and a review belong to the same genere. Here we have several options:

- Retrieving books by a traditional retrieval model (using text-based metadata elements for matching) and then reranking so that books with matching appeal element rank ahead of other books
- Weighing up books with matching appeal at retrieval time
- Traditional retrieval accompanied by pseudo relevance feedback based on the appeal models.

As our current main experimentation line is around pace, we will be more detailed when suggesting a model for pace, than for the other elements.

We will use Wordnet to expand our model with different speach forms / styles / synonyms that may appear in the reviews. A similar procedure will be applied to the topic queries, and matching will consider match in appeal.

2.3 Pace

Pace can be seen as a binary variable, either ”low” or ”high”, making it the easiest element to model and represent, but at the same time less controllable. Saricks poses some questions the answers to which may provide clues as to the pacing:

- Is the book densely written?
- Are there short sentences / short paragraphs, short chapters?
- Is there a straight line plot

High pace	Low pace
Character reveals fast	Character reveals over time
Character reacts to external events	Character only psychologically involved in events
Plot exposed early	Plot exposed over time
Book oriented towards an end	Open ended
Dialog-intensive	Detalied descriptions
Short sentences, paragraphs, chapters	

Fig. 1. Characteristics of the pace appeal elements

3 Related work and our approach

Our work belongs in the realm of content based recommender systems, like for example [4]. The main advantage of such systems is their independence of users and their history of reading and recommendations. As such, these system have a better ability to recommend items not yet recommended to anyone, thereby better supporting serendipity. They are also less likely to serve very close material to what a reader already has read, thereby supporting novelty.

Saricks' framework is reportedly being extensively used in libraries, and in recent years it is starting to gain more systematic use, prominently in a Reader's Advisory resource like Novelist™. Novelist™[6] is a paid service by Ebsco™, marketed towards Reader Advisory (RA) services of libraries active since the late 90's. Among other book characteristics used for recommendation, They have, since 2010 also been recording and utilizing Saricks appeal elements.

In a more research related context, [5] has developed a conceptual approach (guiding the current research), of using Saricks elements in book recommendations. As a part of a Phd-work, [6] have experimented with automatic extraction of appeal elements from reviews using rules related to occurrences / co-ocurrences of types of words from reader reviews. Both the design approach and the evaluation approach are quite straight forward. The appeal element extraction is a combination of a finite list of words (mostly adjectives) expanded by wordnet-extracted synonyms, and rules for these words' occurrence in the sentences of a review. The rules analyse governor - subordinate relations between pairs of words.

Interestingly, [6] have assessed the quality of their ABET extractor by directly comparing its performance to NoveList's recommendations using appraisals by Amazon Mechanical Turk workers as the gold standard, finding ABET more accurate. They also compared the performance their entire system Rabbit (of which ABET is a component) to other recommendation services by using Mechanical Turk appraisers as gold standard when choosing new books that "best relate" to each one from a sample of ten books. The evaluation strategy taken here is very practical, and the results certainly promising. Still we feel that our challenge here is different, as we wish to match books with recommendation

[6] `https://www.ebscohost.com/novelist`

requests (not having other books to relate our recommendations to), and we therefore feel that we need to take a slightly more general approach, which is based on a broader classification of Parts of Speech, particularly adjectives.

Resembling [6], we will also need to take a part / whole approach, trying to see (a) whether our POS-classification has the potential to elicit individual appeal elements (b) whether it is possible to classify recommendation requests the same way as user reviews (whether or not those two types belong to the same genre) and (c) whether correct identification indeed gives us better recommendations on the basis of textual recommendation requests.

4 Data, Experiments and Results

As we were preparing this year's experiments, based on the approach we have taken, we have seen that the crunching of the data is extremely time-consuming and at the moment of writing the data are still in the preparation stage.

4.1 Our Data

The SBS Suggestion Track's (SST) data consist of metadata drawn from LibraryThing and Amazon, describing about 2,8 million books, keyed by their ISBN (meaning the number of distinct works is somewhat lower, as ISBN keys manifestations of works). About half of these, (over 1.3 millions) have non-vacuous reader reviews as a part of their metadata. It is these reviews (free texts) that constitute the data of this paper.

In order to analyse this data, we have so far been taking the following steps:

- POS-tagging of all free texts of the reviews using the Apache OpenNlp[7]
- Collecting all adjectives, basic ($< JJ >$), comparative ($< JJR >$) and superlative ($< JJS >$)
- normalizing the adjective-forms captured by the POS-tagger, and linking each review to the normalized forms of the adjectives.

We have also extended a request to Ebsco to obtain the basic data of the NoveList appeal terms so that we can analyse our methods' ability to extract appeal terms against their data. This will hopefully give us a better possibility to analyse the net-contribution of identifying appeal elements to the ranking of recommendations.

4.2 Our Purpose and Overall Research Design

As an overall, guiding design principle when approaching this issue, we intend to assign values or valences of appeal elements to unseen books (represented by their respective review texts), based on intellectually assigning such values

[7] `https://opennlp.apache.org/`

to chosen books, and building models based on reviews of the latter ones. The most straight-forward way of doing so based on existing NLP-tools, is using reviews of books with known values to build document-categorization models that can use reviews of unseen books to classify those into appropriate categories (low vs. high valence, different intervals of element values a.s.o). Classifying the recommendation requests (topics) in the same manner, can provide us with an additional piece of evidence when matching requests to books.

The problem is that such models, done the ideal way are bound to be week and (as already apparent in the current results, see below) tend to overfit. To combat this problem we need to look at auxiliary procedures mostly based on occurrences of different types of words or word categories. Such procedures are, in their nature simpler than the former, but, as we see it, have the potential of complementing these, hopefully with better results. Here we can hopefully utilize work done on different Parts of Speech around Wordnet[8], and possibly other semantic and lexical resources ([7], [8]).

4.3 Procedure

We start out experimenting with pace as the simplest and (according to [2] the most prominent appeal element. We employ a two-step procedure ranking documents with traditional retrieval methods first, reranking afterwords by matching the paces of the recommendation request and the book reviews. We also experiment with weighing up books based on their pace-match with the recommendation request.

Apeal / Measure	Baseline	Pacing rerank	Pacing upwtd 1.2	Pacing upwtd 1.3
NCDG@10	0,082	0,08	0,012	0,012
MRR	0,0182	0,0182	0,042	0,043
MAP	0,052	0,051	0,009	0,009
R@1000	0,341	0,341	0,254	0,254

Fig. 2. Results of runs with and without incorporating appeal elements

5 Conclusion

As already explained, we see this as a beginning of a long term research endeavor, where the purpose is to directly utilize appeal elements in generating better recommendations based on recommendation requests. We have started out trying to model the pace appeal element in both books' reader reviews and

[8] `https://wordnet.princeton.edu/`

the topics (recommendation requests), trying to see if matching those can give better recommendations. We use two strategies, one based on weighing up relevant terms at retrieval time, and the other one re-ranking results of traditional retrieval based on

The current results seem to suffer from an insufficient pace model, that we obviously need to work more on, tuning particularly the wordnet expansion. The weighing up strategy seems to perform much worse than the re-ranking strategy.

References

1. Syversen, P.C.B.: Anbefalingssystemer for litteratur i den digitale hverdagen. Master's thesis, Oslo University College (2011)
2. Saricks, J.: Readers' Advisory Service in the Public Library. ALA editions. American Library Association (2005)
3. Victoria Caplinger, Elizabeth Coleman, D.C.L.G.L.K.C.K.A.M.E.R.R.Y.: The secret language of books, a guide to appeal. `http://www.ebsco.com/promo/novelist-the-secret-language-of-books` (2015) Promotion Brochure by Ebsco. Accessed: 2015-07-11.
4. Aciar, S., Zhang, D., Simoff, S., Debenham, J.: Informed recommender: Basing recommendations on consumer product reviews. Intelligent Systems, IEEE **22**(3) (May 2007) 39–47
5. Fugleberg, J.R.: Automatisk klassifikasjon av bker basert p brukeranmeldelser: Et konsept. Master's thesis, Hgskolen i Oslo og Akershus (2014)
6. Pera, M.S., Ng, Y.K.: Automating readers' advisory to make book recommendations for k-12 readers. In: Proceedings of the 8th ACM Conference on Recommender Systems. RecSys '14, New York, NY, USA, ACM (2014) 9–16
7. Tsvetkov, Y., Schneider, N., Hovy, D., Bhatia, A., Faruqui, M., Dyer, C.: Augmenting english adjective senses with supersenses. In: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014), Reykjavik, Iceland, May 26-31, 2014. (2014) 4359–4365
8. Kamps, J., Marx, M., Mokken, R.J., de Rijke, M.: Using WordNet to measure semantic orientation of adjectives. In: LREC 2004. Volume 4. (2004) 1115–1118