What's on tonight — User-centered and Situation-aware Proposals for TV Programmes

Bernd Ludwig, Stefan Mandl, Sebastian von Mammen
Chair for Artificial Intelligence, University Erlangen-Nuremberg
Am Weichselgarten 9, 91058 Erlangen, Germany
Bernd.Ludwig@informatik.uni-erlangen.de,Stefan.Mandl@informatik.uni-erlangen.de,Sebastianvonmammen@web.de

ABSTRACT

This paper presents an approach to exploit free text descriptions of TV programmes as available from EPG data sets for a TV recommender system that takes the content of programmes into account. The paper focuses on the natural language understanding problem underlying the analysis of free text descriptions and on methods of classifying free text descriptions with respect to a natural language user query. We close with an evaluation of user acceptance and a discussion of future work.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Communications Applications; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Natural Language*

General Terms

Algorithms, Human Factors

1. INTRODUCTION

Electronic Programme Guides (EPG) provide an enormous amount of information about TV programmes. Viewers are overwhelmed by the huge number of channels and programmes. User models in current recommender systems allow to search for certain features of programmes, like genre, starting time, and other information – these features are quite easy to retrieve and compare by database queries.

However, viewers would prefer to know more about the content of a programme when deciding whether or not to watch it. This requires that a recommender system has information about the content available and is capable of performing a semantic analysis to compare the information with the viewer's interests.

The paper gives an overview of the natural language component of our recommender system. At the beginning of the paper, we compare our work with previous research.

In Section 1.2 we report on a user study that shows how viewers select programmes. In Section 2 we explain our approach to analyze free text descriptions. We conclude the paper with a report on a first evaluation of the system in Section 3 and some remarks on future work.

IUI'06, January 29–February 1, 2006, Sydney, Australia.
ACM 1-59593-287-9/06/0001.

1.1 The State of the Art

The design and implementation of TV recommender systems has attracted great interest in different research groups already. They use sophisticated user models such as the one presented in [1]. In order to allow for default reasoning, stereotypes for users are applied which are based on the analysis of the average user's lifestyle (see [6]). In order to increase the user's confidence in a system's proposals, the generation of trust-worthy suggestions that consider programmes watched earlier has been studied in detail (see [3]). All these research directions deliver valuable contributions to building recommender systems. But they do not cope with the question how a system could take the contents of TV programmes into account.

1.2 How Do Viewers Select TV Programmes?

A user study [9] conducted as part of the research project EMBASSI (see [7, 8]) revealed a number of interesting facts about how users like to select TV programmes. In a Wizard-of-Oz experiment, candidates were situated in front of a computer display that suggested an automatic recommender system to be at work. Actually, in a room nearby, a human person monitored the candidate and responded according to the information available from a TV magazine. In the experiments, the users were allowed to ask arbitrary questions about the currently available TV programmes. On the display a list of proposals was presented and users could ask more specific questions on certain proposals or start a new search if they wanted to.

The experiments showed that almost all test persons ask questions about the *content* of a TV programme:

> *Ich will was Lustiges oder Informatives sehen. (I want to see something funny or informative.)*
>
> *Beziehung (Relationships)*
>
> *Spannung, Magie, Fantasie, Fabelwesen (Thrill, magic, fantasy, mythical creatures)*
>
> *Liebe, Romantik (Love, romance)*

Users often express emotional attitudes they want the programme to have, or even their own emotions hoping the system would come up with proposals that match their mood:

> *Entspannen (to relax)*
>
> *Ich möchte gern etwas Spannendes sehen. Humor sollte auch dabei sein. (I'd like to watch something thrilling. It should also be humorous.)*

Kate's boss feels that a married person is the best bet for a promotion because they tend to stay put and enhance the firm. In order to advance her career, Kate must find a way to pose as "attached." She also has a crush on co-worker Sam who is only interested in girls who are spoken for. She fulfills both requirements by hiring Nick, a young man she just met at a friend's wedding, to pose as her beloved. Nick agrees to play the part of the "picture perfect fiance" but soon falls in love with Kate for real. When Nick decides to head home, Kate soon realizes what true love can be.

Figure 1: English Free text description of the movie *Picture Perfect* (from `http://www.hollywood.com`). The actual system uses the German description of the movie. The English text is provided for illustration.

With the data available in standard EPG, it is hard to retrieve the necessary information to answer such complex queries. Current recommender systems use some sort of formal typology for genres. As explained for example in [2], the typology of such recommender systems is based on a sort of specialized ontology that provides a sort of standard for classifying the content of TV programmes[1]. With such a categorical system, many user requests can be satisfied with proposals that match well. However, the examples above require inference capabilities that are beyond the limits of a typology, as more specific information on the content is required to give satisfying answers.

2. RECOMMENDATIONS ON THE BASIS OF FREE TEXT DESCRIPTIONS

2.1 EPG Data and the Dornseiff-Lexicon

For generating user-tailored recommendations that take the content of programmes into account, we rely on natural language information available in EPG data (which is provided by the TV stations). As an example, in Figure 1 you can see a description of the movie *Picture Perfect* in English.

Our approach to shallow analysis is based on the DORNSEIFF lexicon for German [5]. It groups words according to certain topics, i.e. in each group there are words (even of different word categories) that describe a particular aspect of a certain topic. The DORNSEIFF lexicon[2] is not a synonym lexicon, but a "topic" lexicon. Structured in a two-level hierarchy, the lexicon organizes topics in chapters (e.g. chapter 15 contains subtopic *social life*) and sub chapters (e.g. sub chapter 15.39 is the topic *reward*). If the meaning of a word is ambiguous, it is listed in more than one sub chapter. For the German word *Beförderung* in the text about *Picture Perfect* there are four topics the word is related to:

group id	description
8.5	*Beförderung* (*transport*)
9.33	*Vollenden* (*completion*)
15.39	*Belohnung* (*reward*)
15.62	*Ehre, Ruhm* (*fame, glory*)

As an example for a lexicon entry, we give the translations of some of the nouns in group 15.39.

entry in German	English paraphrase
Beförderung	*promotion*
Belohnung	*gratification*
Ehrenbürgerschaft	*honorary citizenship*
Gehaltserhöhung	*raise of salary*
Prämie	*bonus*

[1]ETSI EN 300707 by the European Telecommunications Standards Institute (available via `http://www.etsi.org`) specifies such a standard

[2]Interested readers can test the online version of the DORNSEIFF lexicon on `http://wortschatz.uni-leipzig.de`.

As English words have connotations different from those of German words, a perfect translation is impossible without context. Therefore, the example is mainly to show that words in a group tend to be related to a common topic.

group	#	group	#	group	#
intensity	1	creation	2	maintenance	1
duration	2	visible	1	stopping	1
steering	1	pull	1	plan	1
random	1	work	2	preparation	1
custom	1	easy	1	high quality	1
improve	1	cooperate	1	help	1
prohibit	1	success	2	wit	1
hope	1	wish	1	love	2
think	1	reason	1	creativity	1
illusion	1	learn	1	insane	1
secret	1	reveal	4	notify	1
advice	1	affirm	1	proof	1
truth	2	pop music	1	family	1
marriage	1	single	2	applause	1
harmony	1	friendship	2	reward	2
unsociable	1	resistance	1	fight	1
victory	1	glory	1	hot, salty	1
sports	1	game	1	reign	2
authority	2	command	1	obligation	1
subserviency	1	imprisonment	1	acquisition	5
grant	3	sell	1	dishonest	1

Figure 2: DORNSEIFF characterization (group numbers omitted) of the free text description in Fig. 1

2.2 Classification by Valence and Arousal

As we are looking for a programme that comes closest to the user query, we need a measure for the distance of programmes. As discussed in Sect. 1.2, viewers tend to choose programmes that are close to their own current mood.

The question now is how to calculate the emotional connotations that come along with the free text description of a programme. This is particularly difficult because words for emotions (e.g. *hate*, *fear*) are sparse in the descriptions.

Our solution is to assume that each content word in a free text description contributes to the overall emotional quality of the text. Therefore, we need to map words to basic emotions. COWIE (see [4]) provides a system of 107 emotional attitudes. They are related to German words as follows (we use the attitude *adventurous* as an example):

In German, *adventurous* means *abenteuerlich*. Its DORNSEIFF groups are *9.72 Gefahr (danger)*, *10.23 Lächerlich (ridiculous)*, *10.38 Tollkühn (daredevil)*, and *11.26 Einbildung, Wahn (illusion)*. Each of these groups comprises words and sometimes even phrases that are used now to indicate the attitude *adventurous* when they appear in a free text description. Some examples for group 9.72 are: *gefährlich (dangerous)*, *tollkühn (daredevil)*, *Hinterhalt (ambush)*.

To each attitude, COWIE assigns a position in a two-dimensional diagram, for *adventurous* the coordinates are (4.2, 5.9)). In his view, any emotional state can be expressed by two values: *valence*, which addresses the quality of an emotion (ranging from very negative over neutral to very positive) and *arousal*, which refers to the (quantitative) activation level of the feeling (from very low to very high).

For the 107 emotional terms and their valence-arousal coordinates (further called VA-coordinates) provided in [4], we searched for the corresponding DORNSEIFF groups as described above. This mapping provides a way to numerically characterize the emotional connotations in each German word in terms of a set of 2D coordinates.

For each free text description, a set of DORNSEIFF groups is computed (e.g. the one in Fig. 2). The set is mapped onto a corresponding set of VA-coordinates as described above. This geometric interpretation (the frequency of a group gives information in the third dimension) allows to interpret the distribution of DORNSEIFF groups as a kind of density distribution (higher frequencies weighting more than lower ones), and the center of gravity is computed as a VA-coordinate. In this way, two descriptions can be compared by computing the EUCLIDean distance of their centers of gravity.

In the sense of the distance definition above, the best proposal for a user query is the description whose center of gravity is closest to that of the query.

3. EVALUATION AND CONCLUSIONS

One long-term research goal that accompanies the development of the described system is to understand how users select TV programmes when they are allowed to interact intuitively with user interface by natural language.

When we want to develop a system that is considered to offer more comfort and flexibility to the users. The most important figure to evaluate is how good do proposals match the user query in the user's view. Users rate the quality of the system high, if the system generates suggestions that seem plausible to the user on the basis of how he/she understands the query.

For a first preliminary evaluation, in a public presentation of the demonstrator system people of different sex, age, education and interest could test the system as long as they wanted to. Then they filled in a questionnaire in order to rate how good the programme descriptions met their expectations, how helpful the recommender was for choosing a TV programme, and whether they would like to have such a system at home.

60 questionnaires are being evaluated at the time of this writing. More than 90 percent of the users said, the proposals were very good, good, or fair, the rest rated them as not matching well or inappropriate. For about 75 percent the system was helpful, and an equal number of persons would buy such a system for a price comparable to a state-of-the-art TV set and recommend the system to a friend. These results are very promising and provide good motivation for further work.

Firstly, we are aware of the fact that the center of gravity approach delivers quite imprecise results. Therefore, we are experimenting with more sophisticated distance measures. for DORNSEIFF characterizations.

Secondly, to further improve the computation of the VA-coordinates, we are currently working on applying GERMANET[3], the German counterpart of WORDNET[4] to compute distances of entries in a DORNSEIFF group by using algorithms for word sense disambiguation. In this way, the can increase the accuracy of VA-coordinates as we can capture the meaning of individual words, not only the meaning of a complete DORNSEIFF group.

4. ACKNOWLEDGMENTS

The research presented in this paper is sponsored by *Software-Offensive Bayern* (`http://www.software-offensive-bayern.de/english.xml`). Ferdinand Hermann, Heike Ott, and Kristina Makedonska provided valuable help implementing major parts of the presented system.

5. REFERENCES

[1] L. Ardissono, C. Gena, P. Torasso, F. Bellifemmine, A. Difino, and B. Negro. User modelling and recommendation techniques for personalized electronic program guides. In L. Ardissono, A. Kobsa, and M. T. Maybury, editors, *Personalized Digital Television – Targeting Programs to Individual Viewers*, volume 6 of *Human-Computer Interaction Series*, chapter 1, pages 3–26. Springer, 2004.

[2] L. Ardissono, F. Portis, P. Torasso, A. Chiarotto, and A. Difino. Architecture of a system for the generation of personalized electronic program guides. In *Proc. UM2001 Workshop on Personalization in Future TV (TV01)*, Sonthofen, July 2001.

[3] A. L. Buczak, J. Zimmerman, and K. Kurapati. Personalization: Improving ease-of-use, trust, and accuracy of a tv show recommender. In *Proceedings of the TV'02 workshop on Personalization in TV*, Malaga (Spain), 2002.

[4] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. Taylor. Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, 18:32 – 80, January 2001.

[5] F. Dornseiff. *Der deutsche Wortschatz nach Sachgruppen*. Walter de Gruyter, 8 edition, 2004.

[6] C. Gena. Designing tv viewer stereotypes for an electronic program guide. In *Proc. UM2001 Workshop on Personalization in Future TV (TV01)*, Sonthofen, July 2001.

[7] T. Herfet, T. Kirste, and M. Schnaider. Embassi – multimodal assistance for infotainment and service infrastructures. *Computers and Graphics*, 25(4):581–592, August 2001.

[8] T. Kirste and T. Heider. Supporting goal-based interaction with dynamic intelligent environments. In F. van Harmelen, editor, *Proceedings of the 15th European Conference on Artificial Intelligence*, pages 22–26, Lyon (France), July 2002. IOS Press.

[9] J. Nitschke and M. Hellenschmidt. Design and evaluation of adaptive assistance for the selection of movies. In *Proceedings of IMC 2003 "Assistance, Mobility, Applications"*, Rostock, June 2003.

[3] http://www.sfs.nphil.uni-tuebingen.de/lsd/

[4] `http://wordnet.princeton.edu`