Content matching between TV shows and advertisements through Latent Dirichlet Allocation

Tristan Ibáñez Sabran

ANR: 2003020

Supervisor: dr. Grzegorz A. Chrupała

Second reader: dr. Menno M. van Zaanen

Master thesis for Data Science: Business and Governance MSc

Tilburg School of Humanities and Digital Sciences

Tilburg, The Netherlands

January 2018

Abstract

New online TV platforms have generated new competition in content for traditional broadcasters, but also in innovative ways to reach the users. While the offline media companies knowledge of the audience was limited to techniques such as surveys, online platforms have access to a plethora of data about their users, which allows them to create highly effective target marketing campaigns. Competing for a fragmented and evasive market, traditional TV companies need to find strategies to cope with this and provide their potential advertisers with competitive edge using as much of their data as possible.

While targeting in video media, as well as text mining in online text platforms, have been investigated before, there is little research on applying text mining as a content targeting technique in the context of video media. This project explores the possibilities of using an established text mining approach to create a targeting algorithm that suggests the shows that match an advertisement content.

LDA, an established technique used in unsupervised text mining, is used on the data that TV companies have about their content: subtitles. Advertisers are described using a systematic approach, and both subtitles and advertisers are processed by LDA to find similar distribution of topics. This method is applied to the subtitles of 492 different shows specifically provided by media company RTL, and 41 advertisements. The accuracy of the model is measured using a set of 119 manually labelled matches. compared to a baseline method. The coherence of topics as well as the number of coherent topics discovered by the model are found to be dependent on parameters such as the overall topic number setting and the filtering parameters. This has, in turn, an effective impact on the recommendations produced, which are also affected by the filtering of context specific terms. Overall, the effectiveness of the model is found to be limited.

Preface

This thesis, done in collaboration between the Tilburg University and RTL, marks the end of my studies of Master of Science degree in Data Science for Business and Governance. I would like to use this lines to quickly acknowledge those who helped me along the way. I am very grateful to Dr. Elske van der Vaart for helping me getting in touch with RTL. I would like to thank Maurits van der Goes, Hajo Wielinga and Menno van Tienhoven of the Consumer Intelligence Department of RTL for their help with the data and their interest in the project, as well as the manager of the department Geoffrey van Meer, for the support and constant flow of ideas. My master colleagues and good friends Carina Jordens and Raymond Hardij for their support and care. My supervisor dr. Grzegorz Chrupała for his patient guidance and advice. Constance de Quatrebarbes for all the debugging sessions, ideas and help. And above all, my parents, and Gladys for helping me think like a programmer and inspiring me to persevere in the face of adversity with her example.

Contents

1. Introduction

It's a marketer's dream – the ability to develop interactive relationships with individual customers. Technology, in the form of the database, is making this dream a reality (Sloan Management Review, 1991).

In the 27 years that have passed since this sentence was used to introduce an article about interactive marketing in the Sloan Management Review, a lot has changed in the way that companies interact with their customers. This is especially true in the case of media companies.

The landscape was then dominated by mass media, with TV as the king, and information about customers was limited to a few demographic indicators collected via expensive surveys. The eruption of the Internet and multiple ways of collecting user information has since made real the "dream to develop interactive relationships with individual customers", probably well beyond what the author of the quote had in mind in 1991. According to marketing magazine *AdAge,* 2017 has been the first year where advertisement investment was bigger on online platforms than on traditional TV.

One of the main advantages of new, online advertisement over traditional media indeed is its ability to target (Bergemann & Bonatti, 2011). Quantitative marketing techniques in display advertisement (the branch of online marketing that includes video content and would be the online equivalent to traditional TV advertisement) include targeting algorithms that match advertisers to users based on their individual characteristics such as age, gender, internet behaviour and the content of the website or platform they are visiting using information such as text, product descriptions, metadata like genre, etc. (Bucklin & Hoban, 2017).

On the domain of online written media, examples of algorithms that connect users to content include clustering and text mining that news aggregator Flipboard uses to recommend their users a group of articles that deal with the same topic of the article they are reading. In this case, Latent Dirichlet Allocation, an algorithm capable of extracting topics out of unlabelled data, is used to cluster similar articles based on their topics. Another example includes the use of text mining techniques by streaming media platform Netflix on TV series scripts, in order to identify what is going on when many users abruptly stop watching a show at the same time.

While TV media companies are making efforts to adapt to the digital revolution by creating online platforms, there are challenges that they are still facing. Advertisement income is shrinking, and TV companies have to slowly start looking for revenue in the Internet ecosystem through their media platforms, fighting for a smaller piece of the cake in a market dominated by search engine giants and fragmented by the rise of new content by user-generated platforms such as Twitch, Internet video on demand services such as Netflix, and even platforms that do both, like Youtube.

Advertiser customisation in television is beneficial for both the advertisers, who can spend their budget wisely, and for media companies, who can increase the price of advertisement space (Adany et al., 2013). However, little research has been conducted in using text mining content matching techniques such as Latent Dirichlet Allocation, mentioned in the Flipboard example, in the context of video media. Online platforms can store more information about their users, but TV companies can mine the information they have available about their shows, such as the subtitle data, to provide advertiser customisation that is beneficial for them and for the advertisers. This project investigates the possibility of using text mining on subtitles to recommend advertisement space by answering the following research question:

RQ: *How will Latent Dirichlet Allocation perform on the task of matching television advertisements to television shows based on their content, using descriptions of the advertisements and subtitle data of the shows?*

To answer this research questions, we briefly describe the data available and the state of the art in text mining for this kind of data in section 2. The goal of the methods, discussed in section 3, is to find matches between television show content and television advertisement content. An advertisement about food and a cooking show would be a good match, and an advertisement that features sports would be matched to a football program, for instance. In the first place, advertisement text descriptions are built using a systematic approach based on previous research. Based on this information, topics are extracted out of subtitles first and of advertisement text descriptions later using an algorithm suited for this task, Latent Dirichlet Allocation. LDA is able to identify topics in each document of a collection of texts that are assumed to share some of this topics. By applying it to our data, we expect it to discover topics in the content of each of the shows and in the descriptions of the advertisements. We will then compare the topics found in each advertisement to those found in each show, and select the most similar, based on similarity metrics, as the best match for that advertisement.

In order to evaluate if these matches are correct, using the information gathered during the process of generating descriptions of the advertisements, and descriptions of the shows collected using multiple sources, a list of matching combinations of shows and advertisements will be generated. The accuracy of the LDA model compared to a baseline method will be measured in the task of finding these combinations based on subtitles and advertisement descriptions, providing an answer to the research question. Matches based on content are not easily quantifiable, therefore, the intent of this work pursues mainly exploratory purposes. All the steps and the tools involved are discussed, as well as the methods used to evaluate the impact of the decisions made on the result of the model.

This document is structured in the following way: in the next section we examine the related literature relevant to the kind of data and task at hand. In section 3, the selected method is described. Section 4 contains the results of the experiment, while conclusions are discussed in section 5.

2. Theoretical framework

2.1 Collecting TV advertisement data

Advertisement dataset is not directly available for us and we will therefore need to build one based on that contains information about the content of the TV advertisements and is suitable for the task of being processed via text mining techniques and compared to the show subtitles. In section 3 we will describe the process that was followed to collect this data. Here, we will briefly take a look at the techniques that have been used in advertisement studies to accomplish this task.

Traditionally, the branch of social sciences that investigates mass media and communication has used content analysis techniques for the purpose of systematically describing audiovisual advertisement. While the task of identifying the amount or quality that a certain concept is represented (such as race in Mastro & Stern (2003) or gender in Bell & Milic (2002)) is quite frequent, examples of systematic methods for translating to written words audiovisual advertisement are not abundant.

One of the concepts that we are interested in because of its applicability in our problem is the "informative cue". In its introduction, Resnik and Stern (1977) acknowledged that not all TV advertisement conveys the same amount of information.

In order to study this, they created the measure of "informative cue" that represents the amount of information about the product that an advertisement communicates, split in categories that include "price", "quality" or "performance" among others. The authors found that advertisement are heterogeneous in the amount of units of this kind that they communicate: some of the advertisers were found to only provide 1 or even none, while others had more than 3.

Although this is a quite dated research, similar conclusions have been reached in more recent studies (Frosch et al., 2007). As we will see in section 3, this methodology will help us build a dataset of advertiser descriptions that are ready for analysis using data mining techniques. The resulting data will contain sentences that describe 41 randomly chosen advertisers of different categories.

2.2 Brief data description

In order to chose the appropriate approach, we will first describe the goal of the setup and the data that will be used. The main goal is predicting good matches between the advertisers and the shows, and for that, we will use – including the aforementioned advertiser descriptions – three different datasets:

1) **Descriptions of advertisements.** The systematic method used to gather this dataset will be explained in more depth in section 3. The resulting data consists of a list of written descriptions of around 20 words of 41 different advertisements recorded in Dutch television.

2) **Show subtitle data.** 1943 files belonging to 492 shows were provided by media company RTL. The most important elements of the subtitle data are the following:

1. *There is no labelled subtitle data*, or, in other words, there is not a set of "golden subtitle files" with a matching "best advertiser" assigned to it. Therefore, unsupervised techniques will have to be used.
2. *Not domain-specific:* the shows are not a collection of documents that have a base of possible topics such as a collection of medical papers or technology related articles.
3. *Oral text transcriptions*: the documents don't have the formal elements of written language, but they are literal transcripts of what was said on screen. Even in the case of fiction shows where dialogues are scripted, scripts mimic the elements of oral language. This is sometimes referred as "structured dialogue" (Howes et al. 2013)
4. *Dutch language*: All shows subtitles are written in Dutch.

3) **Matches between advertisers and shows.** Descriptions of a portion of the shows featured in the subtitles will be gathered using different sources. Using the information of the advertisements descriptions, a list of shows that match the content of an advertisement will be created. This will allow measuring the model's ability to find matches between advertisements and shows.

Therefore, once the data collection process is concluded, we will be left with three datasets that contain text data. We will start by examining the possibilities and state of the research in similar text data.

2.3 Natural language processing in dialogue transcription

Subtitles have been used in machine translation. For instance, Sivic, Everingham and Zisserman's (2006, 2009) research involves word extraction from subtitles in the area of face labelling in TV material, but no processing is applied to the words.

The production of NLP research is more copious in the field of spoken text, either automatically extracted from audio input or manually annotated. Howes et al. (2013) used machine learning techniques in the context of automatic medical evaluation. In their research, an unsupervised probabilistic topic modeling, Latent Dirichlet Allocation, is used on clinical dialogue between patients and doctors. The patient satisfaction and therapy quality rating are successfully predicted by their model (Howes et Al., 2013).

Howes, Purver and McCabe (2006) applied similar methods to transcription of business meetings dialogue. They proposed a method based on LDA to automatically extract highly coherent topics from a corpus of spoken text. The method is proven to be robust against the specific problems posed by transcripts of dialogue in contrast to

directly written documents, where "the discourse is by nature less tidily structured and less restricted in domain" (Howes et al., 2006).

Still in the field of business data, research by Grafe (2014) is yet another example of LDA applied to transcriptions of oral text, in this case, the quarterly earnings call transcripts of publicly traded companies. The experiment aimed at correctly classifying the industry and specific area of each company using clustering on top of the topics discovered by the LDA analysis, and although the results were described by the author as "mixed at best", the research provides interesting additional insight on the challenges posed by oral text transcriptions in the context of natural language processing (Grafe, 2014).

Lin and Chen (2012) offer a broader overview of topic modeling methods in information retrieval used with this kind of transcript document – more precisely, the documents that were generated using automatic speech recognition technology. The authors mention "speech recognition errors, problems posed by spontaneous speech, and redundant information" as the main challenges of this kind of document, and, while the first does not concern our research since our data was not automatically generated by one of such systems, the latter two are indeed related to it. The authors argue that LDA, together with probabilistic latent semantic analysis introduced by Hofmann (2001), offers benefits over either literal term matching and different approaches to NLP based on unigram distributions, as well as over other concept matching strategies such as Pachinko Allocation model and correlated topic models (Lin & Chen, 2012).

In the following section, we will examine LDA, provide examples of applications of LDA in information retrieval, and reasons why it matches the formal aspects of the data featured in this project.

2.4 LDA

Latent Dirichlet Allocation was introduced by Blei et al. in 2003 and has since grown to be the most commonly used topic modeling method (Zhao et al., 2015). It is an unsupervised probabilistic method that assumes that all documents in a corpus are formed by a certain, manually definable amount of topics. It analyses and extracts statistics based on word frequencies to construct a distribution of topic per document (θ), and then a distribution of words per topic (ϕ):

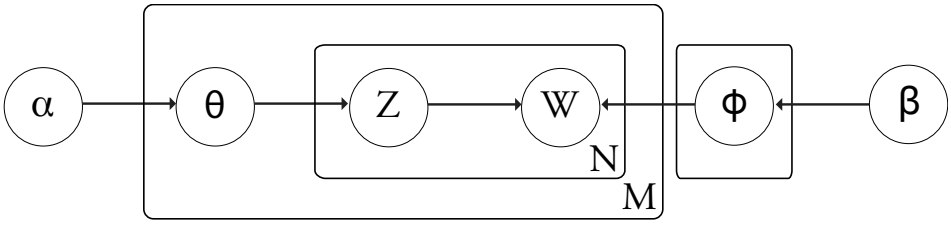


Figure 2.1: Plate diagram of LDA, plates indicate repetition.

Figure 2.1 is a plate diagram that contains the graphical representation of LDA. In a corpus formed by M documents, each of them formed by N words, the larger plate represents the documents, θ is the topic distribution for each of them; while the inner plate represents the choice of topic (Z) and words (W) per document. K is the number of topics, and Φ represents each of the K topics and contains a word distribution. The three plates indicate repetition, α is the parameter of the Dirichlet prior on the topic distributions per document, β is the parameter of the Dirichlet prior on the distribution per topic word.

There are two sides of LDA that need further clarification. Firstly, it is an unsupervised learning process: the model doesn't need a different set of labelled examples. This is appropriate to our dataset of text data that lacks such labels. The model produces for the corpus a number of "word groups", each one of them representing a topic, and an associate probability for each. For each document in the corpus, a different mixture of probabilities associated to each of this topics is produced, which is known as topic probability distribution. Topic distributions from two different documents, such as one from a subtitle file and one from an advertiser description, can later be compared by means of a similarity metric that will be further explained in section 3.

But what do LDA topics represent exactly? To be precise, the full term used in the original LDA introductory paper is "latent topic", and there is no further discussion about its meaning beyond the somehow vague description of "latent variables which aim to capture abstract notions such as topics" (Blei et al., 2003).

Blei went on to co-author a method for interpreting the semantic meaningfulness of LDA generated topics (Chang et al., 2009). Their goal was trying to find a way to evaluate if each topic generated by LDA is capturing indeed groups of words with a coherent meaning (the traditional meaning of the term "topic"), or if it was capturing something else – something such as "opinions and specific (contextual) terms" in the more recent research on this matter by van Atteveldt, Welbers, Jacobi and Vliegenthart (2014).

In the research by van Atteveldt et al. (2014), the authors investigate the exact content captured by LDA. The model was applied on a corpus of political articles from newspapers, the topics – and number of topics – which were known beforehand. When set to discover this exact number of topics, the model was found to be so accurate that the results could be interchangeable with human generated answers ("one can directly use LDA topics in lieu of manual coding"). But the researchers also expressed doubts on whether this results would be consistent when facing a different number of preset topics.

What authors labelled as "opinions and specific terms" are topics that lack the substantive content associated with the traditional term "topic" (such as "international news", formed by the example words "U.S.A.", "Paris", "embassy", etc.) but are instead related to the context of the document. For instance, in an LDA analysis of transcriptions of correspondence, some of the topics discovered will capture the polite set of expressions that are employed in that context, such as "dear", "yours", "truly", "regards", etc.

One can intuitively understand that this will be useful in the context of a classification of documents with the goal of segregating "postal letters" from "e-mails", for instance, since the polite expressions topic will only have high probabilities associated to the former. But if the set of documents is only formed by postal letters, this topic will have little impact on the model's ability to classify the documents based on topics. The task and documents that conform the dataset of this project fall into the latter category. The number of topics and the quality of the unsupervised topics are therefore important elements that will need to be set and explored carefully.

2.5 Topic number selection

A few other authors have expressed similar concern about the quality of LDA topics on situations where there is no *a priori* knowledge on the number of "substantive topics" to be found in the set of documents, and the need to find systematic approaches to discover this number (Zhao et al., 2015, Grant et al., 2013)

In order to select an appropriate number of topics in this situation, the traditional approach was to compare models based on perplexity. This metric measures the ability of the model to predict topics on a held-out test set, and was the metric used in the LDA introduction paper, where it was described as the approach "used by convention in language modeling" (Blei et al., 2003).

However, this de facto measure of LDA topic quality has been questioned in multiple occasions. Zhao et al. (2015) have described it as "not stable", and have proposed a measure based on the rate of perplexity change between different numbers of topics as an alternative metric.

Other authors have expressed similar doubts about the utility of perplexity, especially regarding human interpretation of the resulting topics, which lead them to the exploration of topic coherence measures based on their meaning, including the already mentioned Chang et al. (2009), or on a combination of text mining and scientific philosophy such as C_V, proposed by Röder, Both and Hinneburg (2015), described in section 3.3.3.

2.6 Other aspects

The remaining aspects on topic modeling where there has been recent investigation and are interesting due to their relevance to the problem at hand are those related to preprocessing of topic modeling data. Defined as playing "a very important role" and being "a critical step in text mining" (Vijayarani, 2014; Kannan & Gurusamy, 2014) and investigated in many research papers, some of the traditional techniques that had grown to become standards among NLP practice have been questioned recently. This, together with the peculiarities of our dataset, will be presented in the next paragraphs.

We have divided the tasks in two sections: first, an overview of the usual preprocessing tasks in NLP applicable to our task as observed in multiple papers. The second category is the domain specific problems: how researchers have tackled the problems that rise from working with structured dialogue, and the specific problems of our data that we already mentioned such as the domain specific vocabulary.

2.6.1 General aspects

We will start by quickly enumerating the first steps that traditionally conform the NLP preprocessing toolkit and are applicable to our project.

1. **Stop-word removal:** featured in the LDA introductory paper as well as in a vast amount of papers that include some kind of text information retrieval, the practice of removing a fixed list of words with little semantic content was already defined as "conventional" in a 1986 text (Salton, G. & M. McGill, 1986). Blei et al. (2003) removed a standard list of 50 stop-words in their seminal paper. However, recent research has proposed efficient ways to achieve the same goal, concretely frequently based approaches that remove the terms that are more frequent across the collection of documents. This findings resonate with conclusions from studies on domain-specific words, and will be further discussed in the following subsection.

2. **Decompounding** is the process of separating the words formed in agglutinative languages into single nouns. As we have stated in 2.2, our dataset is in Dutch, a Germanic language that features frequent compound words. The process used to achieve this task is explained in section 3.

2.6.2 Particular aspects

In this section, we will discuss relevant research on papers that involved analysis of data which had similarities to our own data.

Structured dialogue or oral language transcriptions has its own difficulties, in the form of spontaneous speech and redundant information (Lin & Chen, 2012). It is to be expected that in the word distribution per topic of the subtitle files, there will be an important element: the existence of words that are more abundant in the spoken language that the TV scripts capture or emulate – most importantly, the following two kinds:

1. Expressions that belong to the oral language: salutations, good byes, exclamations, etc.
2. Ways of referring to another person: pronouns and person names.

It has been suggested that removing "corpus specific stop words", or terms that have the higher frequency in a corpus, such as the set of given names in our data, has a positive impact on topic coherence, specifically in the context of LDA (Schofield et al, 2017). Similarly, in the context of LDA, it is usual to remove the less frequent words (again, this was done in the introduction of LDA – all words that were featured only once in the collection

of documents were automatically removed (Blei et al., 2003). Those words give little information about the documents they appear in, and hence make document segregation more difficult, and are considered "noise".

In addition to the fixed stop-list approach, other methods based on the notion that both extremely frequent and extremely unfrequent words are uninformative use frequency counts to filter out noise-inducting terms in a way that is automatically adapted to the collection of documents that are being analysed. Examples of these methods include the work by Fan, Doshi-Velez, and Miratrix (2017), who make a distinction between *domain specific* (words that are highly frequent based on the context, such as greetings and names in our case) and *canonical* stop-words, that make reference to the fixed list approach.

But this frequency-based filtering approaches are not routinely extensively documented in research either. In the introduction of the concept, the selection of a cutoff point from which to consider a word "too infrequent" or "too frequent" was described as a "matter of experience" (Luhn, 1958). In recent examples of this approach used in LDA, it is unusual to find mentions of the exact thresholds chosen.

3. Method

In this section, the methodology used to answer the research questions is explained. We will start with an overview of the experimental setup that was prepared based on the information explained in section 2. After that, we will take a deeper view into the process of building the advertiser data in section 3.2, where all three datasets will also be described. Section 3.3 and 3.4 will cover data preparation and the evaluation method respectively.

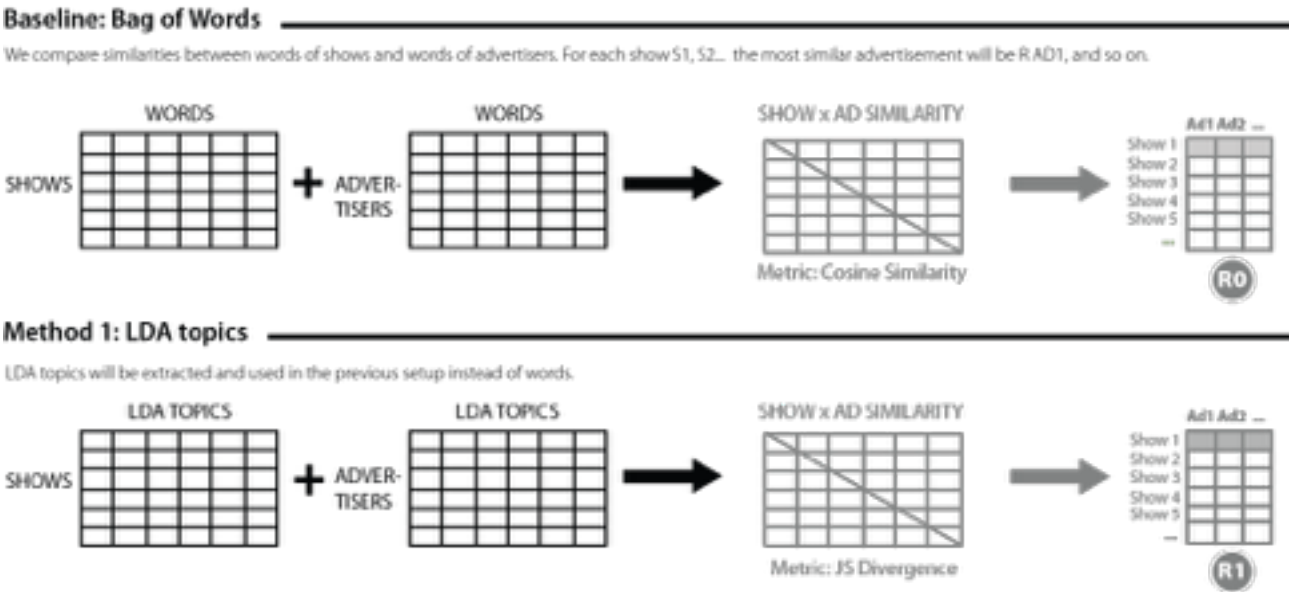
3.1 Method overview

We have devised two methods to provide each show one recommended best matching advertisement, one based on text mining with the text data from advertisers (descriptions) and shows (subtitles) using LDA, and another text mining method based on a less sophisticated approach that will be used as a baseline.

3.1.1 *Bag of words* and Latent Dirichlet Allocation

The main method will be a text mining approach using LDA, an algorithm that is able to work with unlabelled data. The setup of this method, and the baseline method, are depicted in this diagram:

Figure 3.1. Diagram of the baseline and LDA methods



In these two methods, the process can be summarised as follows:

1. **Shows and advertisements are represented in vector space**. In figure 3.1, the two black matrixes of the baseline method represent this step. Each show is assigned a row number, each unique word is assigned a column number. The cell positioned at column W and row S contains the number of times the word W appears in show S. There are two word matrices: one for show subtitles and one for advertisements

descriptions. For the bag of words method, this is the end of step 1.
For the LDA method, words are further preprocessed, and LDA is applied. The matrices contain the resulting topic probability distribution: each column is now an LDA discovered topic, and each cell contains the probability assigned to that topic for the row's show/advertisement. The details are discussed in 3.3.

2. **Similarity between shows and advertisements:** the two matrixes are compared using similarity measures. The left grey matrixes labelled *SHOW x AD SIMILARITY* in each method represent this step in the diagram. For the baseline model, a metric of distance applicable to numeric vectors, cosine similarity, is used. For the LDA method, a measure of similarity between probability distributions, Jensen-Shannon divergence, is used.

3. **Ranking:** in the similarity matrix, each combination of advertisement show with the highest similarity is selected as the recommendation for that advertisement. The right hand grey matrixes in the graph labelled R0 and R1 represent this step. For each method, a final table collects the shows set in order of similarity to each advertisement. The first row, shadowed in the figure, contains the most similar show for each ad, which is the one that will be considered the best match.

3.2 Description of the data and collection process

3.2.1 Collection of advertisement descriptions

There are several goals to accomplish when collecting this data. In the first place, since subtitle data contain text, advertisement need to be in an analogous format, text as well, in order to process it via text mining techniques. Secondly, the subtitles are in Dutch, so the advertisements must be described in Dutch. As we have seen in section 2.1, not all advertisements convey the same amount of information, and there is not an established method to collect written descriptions of TV advertisements.

In order to build a systematic approach to describe each advertisement, with the intention to provide accurate, homogenous information in a format that is usable for text mining algorithms, we used the informative cues concept introduced by Resnik and Stern (1977), information about the product that the advertisements communicate. To tackle the fact discussed in section 2.1 that not all advertisement communicate the same amount of information, we limited the informative cues gathered per ad to 2. All advertisements that had a lesser amount were discarded. The kind of product or service featured in the advertisement was also annotated, and the maximum gathered was limited to 4 advertisements per product or service category (approximately 10% of the total), as to limit the over-representation of a certain industry or kind of service. The additional information that was gathered per advertisement was TV channel, program, brand and name of the product (if applicable).

This method was applied to online TV advertisements recorded between November and December 2017. The audio and video of 41 unique advertisements that fulfilled the above requirement were recorded using

Quicktime. The recordings were analysed using the above procedure. All information was annotated as in this example:

Table 3.1

Examples of informative cues, categories, products and TV stations of the collected advertisers

	Product	**Category**	**Informative cue 1**	**Informative cue 2**	**TV**
Heineken 0,0	Non-alcohol beer	Food	Healthy	Fun, party	RTL
ING	Internet banking	Banking	Many solutions	Easy to use	AT5

A sentence that included both cues, and the kind of product was build for each brand:
Heineken 0,0, a non-alcoholic beer for those who care about their health but still want to have fun.
ING, an internet banking platform for those who want many solutions in a product that's easy to use.

A native Dutch volunteer was provided with the 41 recorded advertisements and the list of sentences in order to translate the sentence and assess the correction of the gathered information. The translations were stored in raw text format for further computer analysis. All the resulting descriptions are featured in the appendix of this document.

3.2.2 Subtitle data

Subtitle data was provided by RTL. The files were extracted to TXT from PAC format, removing the timestamp information. In total 1.943 files that cover a period from 1995 to 2017 were successfully extracted. The first lines of each subtitle file contain the information on the title show, episode if applicable, and original date of air.

3.2.3 Preparation of subtitle files

The preliminary preprocessing of the subtitle dataset consisted in addressing inconsistencies such as different case or punctuation symbols among titles of the same series. In order to achieve this, the following preparation was performed on both datasets after filtering out broken records (files with no information on the title):

1. Each sentence was turned into lowercase
2. All numbers and symbols were removed
3. All words formed by single letter as a result of the above process were removed
4. Dutch and English stopwords were removed (many shows translated in Dutch retain their English title)
5. After this process, titles that still were longer than 2 words were reduced the first 2 words in order to remove uninformative complements frequently found in the titles, especially the season of series stated in incoherent manner such as "Season", "S." or "Se".

As an example, the same show was represented with the title *"THE BOLD & THE BEAUTIFUL S.10"* in the subtitle header of a file, but with the title *The Bold and the Beautiful* in another file. In order to be matched by

software, both need to be represented with the exact same string of characters. Here is a table of steps applied to this example until an homogeneous title is achieved in the shadowed cells:

Table 3.2

An example of cleaning steps taken in order to match two inconsistent codifications of the same show title

	Original	**1**	**2**	**3**	**4**
Example 1	"THE BOLD & THE BEAUTIFUL S.10"	"the bold & the beautiful s.10"	the bold the beautiful s	the bold the beautiful	bold beautiful
Example 2	The Bold and the Beautiful	the bold and the beautiful	the bold and the beautiful	the bold and the beautiful	bold beautiful

After this process was completed, a total of 492 unique shows conformed the final dataset.

3.2.4 Labelled matches

In order to test the accuracy of the model in the task of matching the advertiser descriptions to shows, a separate set of labelled matches was built. A brief description of 110 shows among the 492 present in the subtitle data was manually gathered using information from multiple online sources such as the RTL website, IMDB or Wikipedia. This descriptions were used to create a list of matches between each of the 41 advertiser descriptions and shows that shared similar content based on human judgement. Matches were selected using the data such as informative cues and categories gathered when the advertisements were recorded. Here is a an example of a match based on the "food" category, and a match based on the informative cue "disability":

Table 3.3

Examples of matches between shows and advertiser description

Advertiser	**Description**	**Show**	**Description**
Mora	Cheese balls that are delicious yet easy to cook	**The best**	Cooking show. Every episode focuses on a specific type of dish
Friendship	The association that helps people with disabilities play sports	**Hotel Syndroom**	The conductor runs a hotel with young people who all have mental disabilities

After this process, one advertisement was found to not have a suitable match among the shows and was removed from the data. A total of 119 matches between advertisement description and shows were found using this method.

3.3 Method: LDA

3.3.1 Tools and first steps

For the application of LDA, the Gensim software package that includes the framework introduced in Rehurek and Sojka (2010), was used. This package features Python implementations of LDA and several tools that were used to process and analyse the data.

Once the final text data was loaded, the process of representing the shows and advertisement descriptions in vector space was done. Texts are converted to lowercase with their symbols removed, separated into tokens, and stop-words are filtered out via a fixed list, using Python functions. Additionally, SECOS, a compound splitter, is used. SECOS is the software implementation of the work by Riedl and Biemann (2016), and it allowed splitting the compound tokens via an unsupervised method. The words were exported from Python, decompounded, and imported again. In the case of the baseline model, the processing stops here: this are the words that will be compared to find similarities between shows and advertisers.

Gensim also provides tools to automatically create a dictionary of words, where every unique word is assigned a number. This is useful when creating matrixes of words, since it helps identifying the words once they have been translated to vector space. A corpus is in turn created: each document is turned into a vector where each position represents a word in the dictionary, and contains the number of times it was featured in that dictionary, ready for the LDA processing.

3.3.2 Word filtering

Following the suggestions of literature featured in section 2.6.1, we used a frequency based approach in addition to removing stop-words out of a fixed list. In that section, we described that both elements of the oral style of speech such as greetings and ways of referring to other characters such as given names or pronouns, were a concern when working with structured dialogue data. Words that are repeated across all documents, such as greetings and pronouns will be eliminated by the frequency based filter.

For the given names, the problem was solved using a fixed list of 5492 given names. The list is a freely available dataset created by technology company Sajari, and it includes Dutch and English given names. Names can convey implicit information about the shows in which they appear such as nationality, but that information is not very valuable in the context of our data. Further impact of this removal is quantified in section 4.

As discussed in the last paragraph of section 2, it is unusual to find explanations of the cutoff settings employed in frequency based filtering of LDA input data. Petterson, Buntine, Narayanamurthy, Caetano, and Smola (2010) used 10% as their higher boundary, and 3 as lower: all words that occurred in at least 10% of the documents, and all of those that only occurred in 3 or less, were removed. This was found to remove 2/3 of the words, which would be useful when a massive vocabulary or a very slow method are used.

As will be discussed in the results section, setting the filter had an impact on the topic coherence using both metrics and human evaluation. However, some advertisement descriptions that featured abundant frequent words were found to be degraded beyond usefulness due to the filtering. To remedy this, all the advertiser words were selected and placed in a *white list* of words that were ignored by the filtering system.

3.3.3 Coherence tests

The most important user set parameter in LDA is the number of topics to be discovered. Since the data is unlabelled, it is not possible to know the exact amount of topics that each show contains beforehand. As a consequence of the reasons specified in section 2, we used Röder et al. C_V measure of topic coherence, one of the many included in Gensim.

In their 2015 paper, Röder et al. present a topic coherence measure that is formed by a combination of text mining and scientific philosophy. Their approach consists in a framework formed by four dimensions: methods of segmentation of documents in smaller pieces, aggregation measures for pairs of those small pieces, ways of compute the word probabilities of those aggregation measures, and finally the methods to aggregate scalar values of those probabilities. This modular system is fit many different combinations that are tested against human evaluation of topic coherence. The best performing measure, called C_V, is shown to obtain superior results than previously established topic coherence methods.

It is also possible to set another parameter, a limitation of the number of iterations the model does over the data before it reaches convergence. Since it is an optional parameter that we don't need to apply in our context, it was not used. After the parameters are set, LDA is trained on the subtitle data. Once the LDA model has discovered the topic distribution for each show, the distribution for each advertisement can be inferred using Gensim: for each ad, the model will assign a probability associated to each one of the topics discovered in the show data.

3.3.4 Similarity metrics

In order to create the matches, two different methods were used. Vector representations of show subtitles and advertisement descriptions of the bag of words method were compared using Cosine similarity, a measure of the angle between two vectors used in Natural Language Processing to judge similarities between vector representations of text (Manning & Schütze, 1999). In order to judge the similarity between the LDA topic distribution of a show subtitle and an advertisement description in the LDA method, Jensen-Shannon divergence, a measure of distance between probability distributions derived from the Kullback-Leibler divergence and that has been used in the context of text mining (Eiron & McCurley, 2003) was used.

3.3.5 Test data and evaluation

In order to test the model ability to match shows and advertisement based on their content, the set of labelled data was used. The ability of the model to find this matches is evaluated using accuracy, a metric used in NLP consisting in dividing the number of correct classifications by the number of overall classifications (Manning & Schütze, 1999). Topic evaluation was also performed, using human judgement and the topic coherence method C_V. Different values in the parameters number of topics and word filtering were tested, and the results were judged using both methods. The results of the topic evaluation are presented in the first place in following section, followed by those of the model accuracy.

4. Results

4.1 Topic analysis

4.1.1 Groups of topics

It was found that the content of the discovered topics differ based on the parameter values that were selected when training the LDA models. Those parameters where the number of topics to be discovered and the filtering thresholds for the dictionary. Human judgement of the coherence and meaning of the topics generated under different parameters allowed for a first classification in two groups. On one side, the topics that have a high coherence and simple interpretation. This kind of topic is discovered by all models set to discover between 20 and 40 topics, and most of them appear also even in models with a smaller topic number. Here is an example of this kind of topics:

Table 4.1

Examples of topic coherent words translated from Dutch to English and a suggested topic description word. These topic words were extracted of the LDA model results set to discover 20 topics.

Description	10 words with the highest probability associated to that topic
Food	meat cheese sugar taste wine vitamin bread pan fat plate sauce pizza vegetable chocolate cooking eggs soup chocolate fruit chicken
Thriller	fbi agent weapon jail million kill murder weapons frank trump murder shoot murdock donald judge drugs penalty new punishment gun
Sports	champion wilfred win combat monroe fight iceman tongue fighting po win jay sport round fort winner train ajax beat right
Military	space team president dr colonel war target army fire radio seconds sir building captain general signal flag soldiers command fire

As the number of topics set to be detected rises, subtopics of this categories appear, and topics more closely related to a certain kind of show begin to flourish. At the same time, the number of very coherent topics such as the ones in table 4.1 start to decrease.

Table 4.2

Examples of subtopics words translated, extracted of the LDA model results set to discover 60 topics.

Description	20 words with the highest probability associated to that topic
Kids / Races	car boys crazy greg thing treasure cute boy baby ready balls team riding pants school dog dick ass
World War 2	fuhrer country russian berlin russian german america bangkok germans artillery russian hitler german maddox phuket europe premier otto family gunther
Navy	captain sharks ship boat sea board water shark meter treasure admiral wind island cady wave coast swimming trip men boating

Royalty / UK	king england queen church majesty prince desi fisher highness hastings country royal bishop yours richard wedding princess brian london head

In the first topic labelled *Kids/Races*, we find words that belong to shows that were either set in the environment of a school or in the environment of car races. Some examples found among the shows with the highest probability associated to this topic include: *Over the top* ("Tough trucker Lincoln Hawk is determined to win back his son and triumph at the world arm wrestling championships", according to the IMDB synopsis); *Born to race: fast track,* a movie about a teenager who competes in car races; and *Bad santa*, a comedy whose main characters include a kid.

The next two topics in the table are related to the topic "military" from the previous table. Spontaneous words from other semantic fields such as "golf" and "caddy" are more frequent among otherwise coherent topics when the number set to be discovered is higher.

On the other side, there are two kind of topics that are harder to interpret. The first would be the ones that show little semantic coherence between the words with the higher probabilities associated. The second is the one formed by names of people:

Table 4.3

Examples of topic uncoherent words, translated from Dutch to English, and a suggested topic description word. This topic words were extracted of the LDA model results set to discover 20 topics.

Description	20 words with the highest probability associated to that topic
Unknown	king church very soon robin power change change gentleman most england price hero deeply fear sky turn only scott marriage
Names	times nina bad ludo lucas sjors bing jack charlie janine maxime danny barbara amy lorena arthur weed sanders sadness in advance

4.2 Role of parameters on topic coherence and accuracy

4.2.1 Impact of word filtering parameters on topic coherence and accuracy

Topics were found to become more obviously coherent when increasing the *above* threshold, or the maximum number of shows that the word could be featured in. 0.80 produced both the greater number of coherent, human interpretable topics. In the *below* threshold, the minimum number of shows where a word must appear, 2 was found to produce topics where certain words are always present. Setting thresholds that are not so strict allow less semantic important words to be filtered among the topics, which has a negative impact on accuracy:

Table 4.4

Impact of word filtering on accuracy using setting LDA to 10 and 20 topics

Filters*	Topics	Accuracy	Topics	Accuracy
0.20	10	0.125	20	0.1
0.50	10	0.1	20	0.2
0.80	10	0.175	20	0.225

Note. **Filters*** makes reference to the filtering threshold: 0.20 means words that occur in at least 20% of the shows are removed

The number of topics populated by given names is consistently abundant in every model, regardless the number of topics selected. Filters based on frequency were able to successfully remove some of this noise, and also had a slight positive impact in the interpretation of topics:

Table 4.5

Impact of given name removal on topic coherence scores using a subset of the data

Names removed	Filter	Topics	Coherent topics	At least 50% names**
Yes	0.80	20	8	0
No	0.80	20	6	2

Note. **At least 50% names**** refers to topics where at least 50% of the 20 words with highest assigned probability were given names

4.2.1 Impact of LDA topic setting on coherence score

35 numbers of topics on the LDA method were tested on the data, which provided the following results:

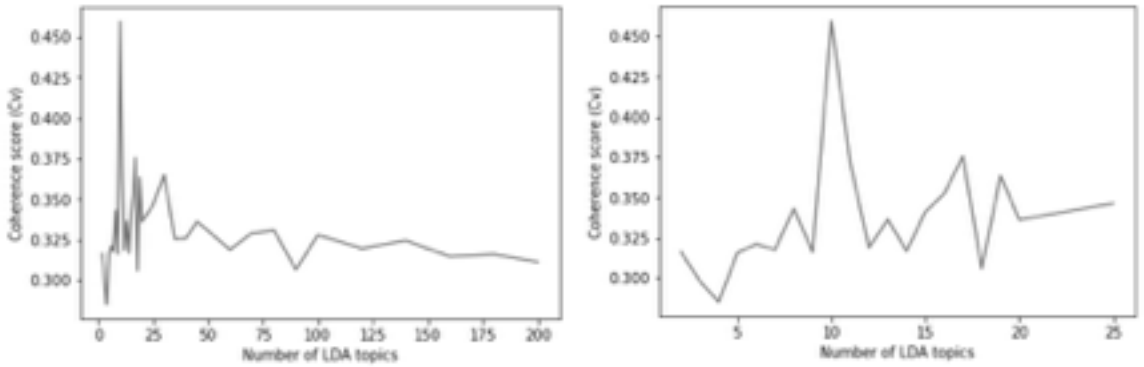


Figure 4.1 Coherence scores and number of topics. The left plot shows values between 1 and 20, from 20 to 50 in steps of 5, from 50 to 100 in steps of 10, and from 100 to 200 in steps of 20. The right plot focuses on the first 25 values in steps of 1.

The best values were found to be between 10 and 20 number of topics, with a peak at 10, with $C_V = 0.459$, the second best value being at 17 with $C_V = 0.375$, and the third at 19 with $C_V = 0.363$. From that point on, coherence gradually degrades, while the less coherent topics mentioned as "subtopics" in the previous section begin to appear.

4.2.3 Impact of LDA topic setting on accuracy

The accuracy of 35 numbers of topics on the LDA model was compared to the baseline bag of words model.

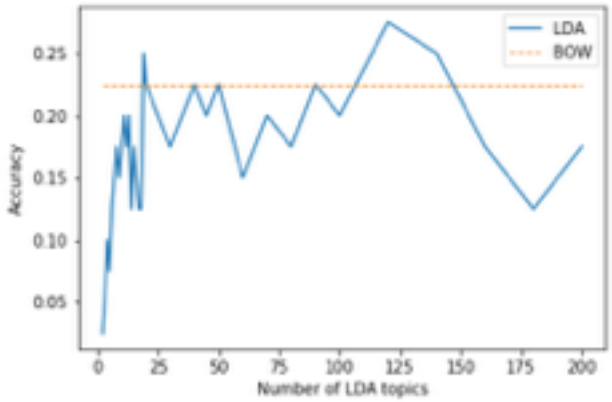


Figure 4.2. Accuracy scores for methods LDA and Bag or words (BOW), and number of LDA topics. Values for number of topics range from 1 to 20, from 20 to 50 in steps of 5, from 50 to 100 in steps of 10, and from 100 to 200 in steps of 20.

The accuracy score of the bag of words model was 0.225. The LDA model was able to only modestly outperform the baseline in a few occasions, particularly with the 120 topics setting, with an accuracy of 0.275. The impact of the selected number of topics on accuracy shows a dissimilar pattern to its impact on the coherence score. The best accuracy results are found when topics are set between 100 and 150, and the worse results are between 2 and 20.

5. Discussion

The results suggest that there are multiple areas where text mining based on LDA lacks the predictive power of similar techniques used in content targeting of online text media when applied to subtitles of TV shows and the systematic description of advertisers proposed in this paper. This lack of accuracy can only provide a negative answer to our research question:

RQ: *How will Latent Dirichlet Allocation perform on the task of matching television advertisements to television shows based on their content, using descriptions of the advertisements and subtitle data of the shows?*

The results of the experiments show that Latent Dirichlet Allocation perform poorly on the task of finding matches between advertisement and television shows using advertisement descriptions and subtitles when measuring accuracy using labelled data in comparison to the Bag of Words approach. Overall, LDA was not found to be a viable text mining method to perform the task. But let's take a moment to examine the causes and the opportunities of research and improvement.

Topics are not always topics

As discussed in section 2, there is ongoing research around the concept of *topic* in the context of LDA. Therefore, the intuition that this method will always provide *topics* equivalent to the traditional meaning of the concept is not completely correct. LDA topics can be formed by contextual terms that depend on the kind of corpus it is used on, and may provide little help in identifying the topics, such as formalities found in legal jargon on a corpus of court recordings. When applied on our data, less than half of the topics were consistently found to be coherent. We have achieved modest improvements in topic coherence measured by both human judgement and a coherence metric through word filtering, but it has shown to have a negligible impact on the accuracy of the model.

Insufficient information in subtitles

After manually inspecting some of the documents and matches, some inherent limitations of the subtitles for the task became obvious. As an example, in the show *AC/DC live concert,* the narrator only introduces the song name and the album it belongs to, and these are the only lines of text present in the subtitle file. Although it is a viable match for advertisement descriptions that feature music related informative cues such as the Philips advertisement (*Philips Hue, light customisable lightbulbs for people who love music*), no matter how the parameters are set, matches with this music related show cannot be made because the lack of music related terms in the subtitles prevents LDA from assigning the show a high probability to topics with words related to music. This may have had a negative impact in the model ability to mimic the matches a human judge produces by comparing the advertiser description to the show description.

Deepness of some topics

Even when lyrics of songs are present in live music shows, such as *New kids on the block live,* where the subtitles contained the lyrics of the songs, the model would produce a match based on the content of the lyrics and it would be difficult for it to identify that the words belong to a music show, therefore making matches between advertisers where music is somehow related to live music shows very difficult. Contextual terms that only occur in the context of live shows could be gathered in one topic by LDA, but the lack of this kinds of words in an advertiser description related to music such as the Philips example could have prevented associating the advertiser with the topic, resulting in a high divergence between the two. Incorporating metadata is a possible solution that would improve the ability of the model to find this kind of matches.

Quality of the match

Applications of content targeting marketing assume the tone of the context where the targeted concepts are found is positive. However, during the construction of the label data, it became apparent that some the matches could be problematic, because the content of the shows match the content of the advertiser description but it may be critical towards the product. To be specific, the aforementioned *Philips hue* lightbulbs advertisement was matched to *Lights out!*, a documentary that investigates the negative effects of artificial lighting on health. So the model can successfully match this show and this advertisement that share a similar topic, but it is unable to filter out those shows where the topic is investigated critically. Unless the shows have been labelled as "critical" in a database, there is little that metadata could do to solve this. Investigating ways to incorporate a form of sentiment analysis could help prevent this kind of matches.

Other opportunities of further research are abundant. Advertiser descriptions could be generated in different fashion, by automatically incorporating information on the brand found on online sources such as Wikipedia, or by applying speech-to-text techniques to the advertisements. Clues from other media such as press or internet advertisement could also be incorporated, and other features such as the industry, target public or features of the product could be used as features beside the description. Similarly, metadata on the show extracted from online sources or from the TV companies archives such as the profile of its users could be incorporated. Sentiment analysis techniques could be applied in parallel to determine if a show is matched to a certain advertisement but it is not critical with the industry or the product. Additionally, better filtering methods could be investigated to try to filter out the noise while preserving valuable information. Finally, methods that allow for less subjective approach of the evaluation of content-based matches would also greatly improve the task of investigating this sort of algorithms.

Content targeting in TV and in video content is an area of research with interesting potential applications for both the conventional TV companies and the emerging Internet platforms. LDA is an established approach which has taken a prominent position in the landscape of text mining algorithms used in content matching and semantic classification of documents. In this exploratory work, it has shown multiple specific needs when addressing the task of matching advertisers to shows based on content. When systematic solutions to this needs are developed, there is no reason to believe that LDA won't be an appropriate solution to this task.

Bibliography

Adany, R., Kraus, S., & Ordonez, F. (2013). Allocation algorithms for personal TV advertisements. *Multimedia systems*, *19*(2), 79-93.

van Atteveldt, W., Welbers, K., Jacobi, C., & Vliegenthart, R. (2014). LDA models topics… But what are 'topics'?

Bergemann, D., & Bonatti, A. (2011). Targeting in advertising markets: Implications for offline versus online media. *The RAND Journal of Economics, 42*(3), 417-443. Retrieved from http://www.jstor.org/stable/23046807

Bhadury, A. (08/02/2017) Clustering Similar Stories Using LDA. *Flipboard engineering*, retrieved from http://engineering.flipboard.com/2017/02/storyclustering

Blattberg, R. C., & Deighton, J. (1991). Interactive marketing: Exploiting the age of addressability. *Sloan management review*, *33*(1), 5.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993-1022.

Bucklin R.E., Hoban P.R. (2017) Marketing Models for Internet Advertising. In: Wierenga B., van der Lans R. (eds) *Handbook of Marketing Decision Models*. Springer

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In *Advances in neural information processing systems* (pp. 288-296).

Eiron, N., & McCurley, K. S. (2003, July). Analysis of anchor text for web search. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval* (pp. 459-460). ACM.

Everingham, M., Sivic, J., & Zisserman, A. (2006). Hello! My name is... Buffy – automatic naming of characters in TV video.

Everingham, M., Sivic, J., & Zisserman, A. (2009). Taking the bite out of automated naming of characters in TV video. *Image and Vision Computing*, *27*(5), 545-559.

Fan, A., Doshi-Velez, F., & Miratrix, L. (2017). Prior matters: simple and general methods for evaluating and improving topic quality in topic modeling.

Frosch, D. L., Krueger, P. M., Hornik, R. C., Cronholm, P. F., & Barg, F. K. (2007). Creating demand for prescription drugs: a content analysis of television direct-to-consumer advertising. *The Annals of Family Medicine*, *5*(1), 6-13.

Govind, N. (11/06/2014) Optimizing the Netflix Streaming Experience with Data Science. *Medium*, retrieved from https://medium.com/netflix-techblog/optimizing-the-netflix-streaming-experience-with-data-science-725f04c3e834

Grafe, P. Topic Modeling in Financial Documents. Department of Computer Science Stanford University.

Grant, S., Cordy, J. R., & Skillicorn, D. B. (2013). Using heuristics to estimate an appropriate number of latent topics in source code analysis. *Science of Computer Programming*, *78*(9), 1663-1678.

Hofmann, T. (2001). Unsupervised learning by probabilistic latent semantic analysis. *Machine learning*, *42*(1), 177-196.

Howes, C., Purver, M., & McCabe, R. (2013). Investigating topic modelling for therapy dialogue analysis. Association for Computational Linguistics.

Kannan, S., & Gurusamy, V. (2014). Preprocessing Techniques for Text Mining.

Lin, S. H., & Chen, B. (2012). A Comparative Study of Methods for Topic Modeling in Spoken Document Retrieval. *Computational Linguistics & Chinese Language Processing*, 65.

Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of research and development*, *2*(2), 159-165.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT press.

Mastro, D. E., & Stern, S. R. (2003). Representations of race in television commercials: A content analysis of prime-time advertising. *Journal of Broadcasting & Electronic Media*, *47*(4), 638-647.

Petterson, J., Buntine, W., Narayanamurthy, S. M., Caetano, T. S., & Smola, A. J. (2010). Word features for latent dirichlet allocation. In *Advances in Neural Information Processing Systems* (pp. 1921-1929).

Purver, M., Griffiths, T. L., Körding, K. P., & Tenenbaum, J. B. (2006, July). Unsupervised topic modelling for multi-party spoken discourse. In Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics (pp. 17-24). Association for Computational Linguistics.

Rehurek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. In *In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*.
Note: Gensim was released under GNU LGPLv2.1 license and is available freely for academic use. It is available at https://radimrehurek.com/gensim/index.html

Resnik, A., & Stern, B. L. (1977). An analysis of information content in television advertising. *The Journal of Marketing*, 50-53.

Riedl, M., & Biemann, C. (2016). Unsupervised Compound Splitting With Distributional Semantics Rivals Supervised Methods. In *HLT-NAACL* (pp. 617-622).

Röder, M., Both, A., & Hinneburg, A. (2015, February). Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining* (pp. 399-408). ACM.

Rose, J., Mackey-Kallis, S., Shyles, L., Barry, K., Biagini, D., Hart, C., & Jack, L. (2012). Face it: The impact of gender on social media images. *Communication Quarterly*, *60*(5), 588-607.

Sajari (2014). *First names CSV* [Data file]. Retrieved from https://www.sajari.com/free-data/CSV_Database_of_First_Names.csv

Salton, G., & McGill, M. J. (1986). Introduction to modern information retrieval. McGraw-Hill, Inc.

Schofield, A., Magnusson M., Mimno, D. (2017). Pulling Out the Stops: Rethinking Stopword Removal for Topic Models. *EACL 2017*, 432.

Slefo, G. (26/04/2017) Desktop and mobile ad revenue surpasses TV for the first time. *AdAge*, Retrieved from http://adage.com/article/digital/digital-ad-revenue-surpasses-tv-desktop-iab/308808/

Vijayarani, S., Ilamathi, M. J., & Nithya, M. (2015). Preprocessing techniques for text mining-an overview. *International Journal of Computer Science & Communication Networks*, *5*(1), 7-16.

Zhao, W., Chen, J. J., Perkins, R., Liu, Z., Ge, W., Ding, Y., & Zou, W. (2015). A heuristic approach to determine an appropriate number of topics in topic modeling. *BMC bioinformatics*, *16*(13), S8.

Appendix

Advertiser descriptions and TV show matches descriptions

This appendix features the description sentence for each of the 40 advertisements that were found to have at least one match among the show descriptions. All advertisements are also presented together with the description of the shows that were labelled as good matches in the dataset used to evaluate the accuracy of the methods.

Mora, cheese balls that are delicious yet easy to cook
Eetclub. Series, reality show, cooking. A group of friends take turns inviting each other to a home made dinner. Each episode covers the process of preparing the meal: deciding the menu, getting the ingredients, cooking and preparing the decoration. At the end of each episode, the guests rate the dinner and give advice to the cook.
The best. Cooking show. Every episode focuses on a specific type of dish, including sweet breakfast, Eggs & Lamb, Chocolate & Fish, Sandwiches & steaks.
You are what you eat. Reality TV, the show often uses shock tactics to get the participants to lose weight. In each episode, all food eaten in one week by the person(s) taking part is placed on a table to highlight problem areas of their diet.
In defense of food. "Eat food. Not too much. Mostly plants." With that seven-word maxim, US-based journalist Michael Pollan (The Omnivore's Dilemma) distills a career's worth of reporting into a prescription for reversing the damage being done to people's health by today's industrially driven Western diet. In Defense of Food debunks the daily media barrage of conflicting claims about nutrition.

Hansaplast, band-aids that for the athletes that don't want to be stopped by bruises
Centraal medisch centrum. Drama series that takes place in a hospital. The stories of CMC are based on realistic medical themes and ethical dilemmas and are hung around romances and intrigue between the doctors and nurses. The series also includes patients who suffer from diseases that require hospitalization.
Coma. A mistery movie. When a young female doctor notices an unnatural amount of comas occurring in her hospital she uncovers a horrible conspiracy.
Voetbal Inside. Sports, talk: Dutch football talk show that has aired on TV channel RTL 7. In a studio, host Wilfred Genee discusses the latest developments in Dutch and international football with Johan Derksen, René van der Gijp and a guest.
David Beckham: Into the Unknown. Travel, football, sports: After 22 years of playing for the world's greatest football teams, David Beckham has retired and he has the freedom to do whatever he wants. To mark the occasion he's going on an adventure. He's chosen Brazil. Starting with a beach foot-volleyball game in Rio, they travel deep into the Amazon, ending up with the remote Yanonami tribe, with David desperately trying to explain the beautiful game.
Tiki Taka Touzani Sports. The world's best freestyle football player Soufiane Touzani speaks in this second season of Tiki Taka Touzani with both active and retired dutch football players who play home or abroad. He takes a unique look into the life of the (former) football players, has surprising conversation with them and presents them with exciting new challenges.
In oranje. The 12-year-old Remco is a talented football player. It is his big dream to ever be allowed to play in the Dutch national team. When his father Erik, his coach and biggest supporter, suddenly dies, Remco's world collapses.

Gall & Gall, a liquor store where you can find gifts at good prices
Eetclub. Series, reality show, cooking. A group of friends take turns inviting each other to a home made dinner. Each episode covers the process of preparing the meal: deciding the menu, getting the ingredients, the wine, cooking and preparing the decoration. At the end of each episode, the guests rate the dinner and give advice to the cook.
The best. Cooking show. Every episode focuses on a specific type of dish, including sweet breakfast, Eggs & Lamb, Chocolate & Fish, Sandwiches & steaks. Ingredients include wine.
You are what you eat. Reality TV, the show often uses shock tactics to get the participants to lose weight. In each episode, all food eaten in one week by the person(s) taking part is placed on a table to highlight problem areas of their diet.

Maggie, a vegetable soup that is still delicious and for active people
Eetclub. Series, reality show, cooking. A group of friends take turns inviting each other to a home made dinner. Each episode covers the process of preparing the meal: deciding the menu, getting the ingredients, cooking and preparing the decoration. At the end of each episode, the guests rate the dinner and give advice to the cook.
In defense of food. "Eat food. Not too much. Mostly plants." With that seven-word maxim, US-based journalist Michael Pollan (The Omnivore's Dilemma) distills a career's worth of reporting into a prescription for reversing the damage being done to people's health by today's industrially driven Western diet. In Defense of Food debunks the daily media barrage of conflicting claims about nutrition.
You are what you eat. Reality TV, the show often uses shock tactics to get the participants to lose weight. In each episode, all food eaten in one week by the person(s) taking part is placed on a table to highlight problem areas of their diet.

The best. Cooking show. Every episode focuses on a specific type of dish, including sweet breakfast, Eggs & Lamb, Chocolate & Fish, Sandwiches & steaks.

Etos, a beauty store where you can find many gifts at good prices.
Roy Donders Stylist van de straat. Roy Donders is the stylist for women from popular neighborhoods and caravan camps and now has his own reality series! The Tilburg stylist, who still lives with his mother, in his reality series gives the viewer a glimpse into his own world, which is connected to glitters, stones and sequins.
Iliza Shlesinger War Paint. War Paint is Iliza Shlesinger's first standup comedy special. Shlesinger's material is highly relatable to women, as she discusses dating, friendship and girls' nights out, but it is just as interesting for men as she provides insight into a number of things women do: wearing make-up, ordering salad on a date, and pretending to like hiking.
Moordwijven. Comedy, film. Three millionaire best friends, Kitty, Estelle and Nicolette, spend their days having botox injections, liposuction and other plastic surgery – all to keep up their looks for their husbands. After witnessing the murder of their favorite plastic surgeon by his wife, the three ponder the idea of similarly killing their husbands should they sleep with someone else.

Elmex sensitive repair and prevent, a toothpaste that protect the gums and has been scientifically tested
Galileo. Galileo is a popular science TV program that has been broadcast on RTL 5 since 2016. It is based on a German format with the same name. In the items remarkable experiments, places and people are discussed, often with a scientific explanation.
Coma. A mistery movie. When a young female doctor notices an unnatural amount of comas occurring in her hospital she uncovers a horrible conspiracy.
Hospital. Medical documentary that tells the story of the British public healthcare near breaking point and how staff are coping in these unprecedented times in five different London hospitals.
Centraal medisch centrum. Drama series that takes place in a hospital. The stories of CMC are based on realistic medical themes and ethical dilemmas and are hung around romances and intrigue between the doctors and nurses. The series also includes patients who suffer from diseases that require hospitalization.

Colgate visible effect, a toothpaste that helps that improves mouth health in a visible way.
Coma. A mistery movie. When a young female doctor notices an unnatural amount of comas occurring in her hospital she uncovers a horrible conspiracy.
Hospital. Medical documentary that tells the story of the British public healthcare near breaking point and how staff are coping in these unprecedented times in five different London hospitals.
Centraal medisch centrum. Drama series that takes place in a hospital. The stories of CMC are based on realistic medical themes and ethical dilemmas and are hung around romances and intrigue between the doctors and nurses. The series also includes patients who suffer from diseases that require hospitalization.

Hubo and Multimate, building material for home for those who like do it yourself and for those who need professional technicians.
The renovation. Film, drama. Tessa opens her luxurious clinic with a big party. She is smart, beautiful, savvy, and happily married with an intelligent adolescent son. But appearances are deceptive. In reality, her marriage is falling apart, her house renovation is a disaster and the finances of her clinic are a mess.
Furry Vengeance. Comedy, family - In the Oregon wilderness, a real-estate developer's new housing subdivision faces a unique group of protestors: local woodland creatures who don't want their homes disturbed.

Playstation 4 Pro, a video game console with better graphics for dedicated players
Ender's game. In the near future, a hostile alien race (called the Formics) have attacked Earth. Arriving at Battle School, Ender quickly and easily masters increasingly difficult war games.
De unieke wereld van Lego. Documentary in which a unique look behind the scenes is given at the toy empire LEGO in the run-up to the festive season. They are working hard to create the largest lego Christmas show ever.
All-In Kitchen. Good friends Tjé and Hein are asked to save the restaurant of their friend Marcel, where guests play poker at the end of a dinner.

Mazda CX-5, a Japanese car that is very easy to drive
's werelds mooiste auto's. Documentary about a the Pebble Beach Concours d'Elegance in California, where the world's most exclusive cars are featured.
Million pound motors. Series, sports. This First Cut documentary explores the exclusive world of the vintage motor trade, meeting one of Britain's most eccentric car salesmen and his super-rich clients
De slechtste chauffeur van Nederland. Reality show, driving. This program is about drivers with a driving license B who have problems with driving a passenger car.

1: Life on the limit. A documentary film that traces the history of Formula One auto racing from its early years, in which some seasons had multiple fatalities, to the 1994 death of Ayrton Senna, the sport's most recent death at the time of production.

Oscillococcinum, a cough medicine for those who want to remain active in the winter
Voor ik het vergeet. Reality TV, science. Angela Groothuizen follows people aged between 31 and 65 who have different forms of dementia for a year. A diagnosis with a lot of impact as there is no cure for this disease. Angela gets an intimate look into the lives of people who until recently were still alive.
The visit. Documentary. The story of the family of Idris, who was diagnosed when he was a kid of Duchenne's disease.

Swisse, vitamin pills with more ingredients that helps you feel good
De Waarheid Over Vitaminen. Documentary, health, nutrition. Filmmaker and health freak Bryce Sage travels through America in search of the answers to his burning questions. We all know that we need them, but why? When will we get enough of it? And why don't we make them all ourselves?

Kruidvat, a health and beauty products store that offers free jewellery and is always advantageous
Roy Donders Stylist van de straat. Roy Donders is the stylist for women from popular neighborhoods and caravan camps and now has his own reality series! The Tilburg stylist, who still lives with his mother, in his reality series gives the viewer a glimpse into his own world, which is connected to glitters, stones and sequins.
Iliza Shlesinger War Paint. War Paint is Iliza Shlesinger's first standup comedy special. Shlesinger's material is highly relatable to women, as she discusses dating, friendship and girls' nights out, but it is just as interesting for men as she provides insight into a number of things women do: wearing make-up, ordering salad on a date, and pretending to like hiking.
Moordwijven. Comedy, film. Three millionaire best friends, Kitty, Estelle and Nicolette, spend their days having botox injections, liposuction and other plastic surgery – all to keep up their looks for their husbands. After witnessing the murder of their favorite plastic surgeon by his wife, the three ponder the idea of similarly killing their husbands should they sleep with someone else.

Braun series 9, 7 and 5, electric shaver that incorporates technology to be efficient and gentle
Super factories. Documentary, engineering. The program explores the inner workings of factories worldwide. Each episode profiles the machinery and manpower behind each factory's main product.

Gillet Fusion 5, a razor used by football players that provides a precise shave
In oranje. The 12-year-old Remco is a talented football player. It is his big dream to ever be allowed to play in the Dutch national team.
Voetbal Inside. Sports, talk: Dutch football talk show that has aired on TV channel RTL 7. In a studio, host Wilfred Genee discusses the latest developments in Dutch and international football with Johan Derksen, René van der Gijp and a guest.
David Beckham: Into the Unknown. Travel, football, sports: After 22 years of playing for the world's greatest football teams, David Beckham has retired and he has the freedom to do whatever he wants. To mark the occasion he's going on an adventure. He's chosen Brazil. Starting with a beach foot-volleyball game in Rio, they travel deep into the Amazon, ending up with the remote Yanonami tribe, with David desperately trying to explain the beautiful game.
Tiki Taka Touzani Sports. The world's best freestyle football player Soufiane Touzani speaks in this second season of Tiki Taka Touzani with both active and retired dutch football players who play home or abroad. He takes a unique look into the life of the (former) football players, has surprising conversation with them and presents them with exciting new challenges.

Always My Fit, sanitary pads for woman that adapt to each body and provide total protection
Iliza Shlesinger War Paint. War Paint is Iliza Shlesinger's first standup comedy special. Shlesinger's material is highly relatable to women, as she discusses dating, friendship and girls' nights out, but it is just as interesting for men as she provides insight into a number of things women do: wearing make-up, ordering salad on a date, and pretending to like hiking.

Nivea Care and Hold, a hairspray for styling that also protects the hair
Roy Donders Stylist van de straat. Roy Donders is the stylist for women from popular neighborhoods and caravan camps and now has his own reality series! The Tilburg stylist, who still lives with his mother, in his reality series gives the viewer a glimpse into his own world, which is connected to glitters, stones and sequins.
Iliza Shlesinger War Paint. War Paint is Iliza Shlesinger's first standup comedy special. Shlesinger's material is highly relatable to women, as she discusses dating, friendship and girls' nights out, but it is just as interesting for men as she provides insight into a number of things women do: wearing make-up, ordering salad on a date, and pretending to like hiking.

Moordwijven. Comedy, film. Three millionaire best friends, Kitty, Estelle and Nicolette, spend their days having botox injections, liposuction and other plastic surgery – all to keep up their looks for their husbands. After witnessing the murder of their favorite plastic surgeon by his wife, the three ponder the idea of similarly killing their husbands should they sleep with someone else.

Pampers, diapers that offer the best protection while being comfortable
What to Expect When You're Expecting. Five expectant couples learn that having a baby is anything but predictable in this uplifting romantic comedy based on Heidi Murkoff's ubiquitous best-seller.
Het beste voor je kind. Series, education - Program in which parents with a clear vision on the education of their child are followed. They use special parenting styles from a vision of life, belief or from their choice for a particular lifestyle.
Jo Frost Extreme Parental Guidance. Over the course of the series, Jo also takes a look at what she believes to be the single most important cause of parenting problems: families not spending enough time together.

Veva, a military school where you can learn useful abilities
Hummingbird. Action crime movie. Homeless and on the run from a military court martial, a damaged ex-special forces soldier navigating London's criminal underworld seizes an opportunity to assume another man's identity.
Bloodsport. Action, biography, drama. Follows Frank Dux, an American martial artist serving in the military, who decides to leave the army to compete in a martial arts tournament in Hong Kong where fights to the death can occur.

MasterCard, a credit card for important purchases that keeps your purchases safe
De Psychologie Van Geld. Documentary in which scientific research is done into the way crucial financial decisions are made. Why could not economists predict the 2008 gigantic financial crisis?

Ikea family, a discount card for furniture that provides good prices on kitchen appliances
Eetclub. Series, reality show, cooking. A group of friends take turns inviting each other to a home made dinner. Each episode covers the process of preparing the meal: deciding the menu, getting the ingredients, cooking and preparing the decoration. At the end of each episode, the guests rate the dinner and give advice to the cook.

ABN AMRO Financialfocus, the assets platform that has many options and is also inspirational
De Psychologie Van Geld. Documentary in which scientific research is done into the way crucial financial decisions are made. Why could not economists predict the 2008 gigantic financial crisis?
Turnaround king. Sales expert Grant Cardone shows his TurnAround skills off in a gym. Cardone known for his brash 21st century sales skills demonstrates on national tv how to make a sale, how to increase pricing, and how to turnaround the morale of the staff and more importantly turnaround the profits of the company.

Friendship, the association that helps people with disabilities play sports
Hotel syndroom. Hotel SynDROOM is a Dutch television program that is broadcast by RTL 4. The presentation is in the hands of Johnny de Mol. De Mol runs a hotel with young people who all have mental disabilities, for example Down's syndrome or autism.
Malaika. Zoë, Anna and Robin are three trainees who start their internship at the Malaika care center. They have to stand up in a heavy job within the care center, while in their private life they also have tension.
In oranje. The 12-year-old Remco is a talented football player. It is his big dream to ever be allowed to play in the Dutch national team.
Voetbal Inside. Sports, talk: Dutch football talk show that has aired on TV channel RTL 7. In a studio, host Wilfred Genee discusses the latest developments in Dutch and international football with Johan Derksen, René van der Gijp and a guest.
David Beckham: Into the Unknown. Travel, football, sports: After 22 years of playing for the world's greatest football teams, David Beckham has retired and he has the freedom to do whatever he wants. To mark the occasion he's going on an adventure. He's chosen Brazil. Starting with a beach foot-volleyball game in Rio, they travel deep into the Amazon, ending up with the remote Yanonami tribe, with David desperately trying to explain the beautiful game.
Tiki Taka Touzani Sports. The world's best freestyle football player Soufiane Touzani speaks in this second season of Tiki Taka Touzani with both active and retired dutch football players who play home or abroad. He takes a unique look into the life of the (former) football players, has surprising conversation with them and presents them with exciting new challenges.

Uit de eenzaamheid, an association that helps lonely elderly people through literature.
Kees de Jongen. Film, drama. Filming of the classic book about the Amsterdam schoolboy Kees who lives in two worlds: the harsh reality and his fantasy
Malaika. Zoë, Anna and Robin are three trainees who start their internship at the Malaika care center. The trainees of the care center have to stand up in a heavy job within the care center, while in their private life they also have the necessary tension

Staxi, an app where you can book and pay a taxi ride online.
Stolen. Action movie. A former thief frantically searches for his missing daughter, who has been kidnapped and locked in the trunk of a taxi.
Collateral. A cab driver finds himself the hostage of an engaging contract killer as he makes his rounds from hit to hit during one night in Los Angeles.

This is Holland, a flying experience that not only shows the Netherlands but also really lets you experience it.
In oranje. The 12-year-old Remco is a talented football player. It is his big dream to ever be allowed to play in the Dutch national team.
Flying with Qantas. Documentary series in which the Australian airline Qantas is followed by the eyes of the enthusiastic employees, who give everything for the success of this iconic society.

Black angus, delicious bitter ballen that are made from exclusive ingredients
In defense of food. "Eat food. Not too much. Mostly plants." With that seven-word maxim, US-based journalist Michael Pollan (The Omnivore's Dilemma) distills a career's worth of reporting into a prescription for reversing the damage being done to people's health by today's industrially driven Western diet. In Defense of Food debunks the daily media barrage of conflicting claims about nutrition.
The best. Cooking show. Every episode focuses on a specific type of dish, including sweet breakfast, Eggs & Lamb, Chocolate & Fish, Sandwiches & steaks.
You are what you eat. Reality TV, the show often uses shock tactics to get the participants to lose weight. In each episode, all food eaten in one week by the person(s) taking part is placed on a table to highlight problem areas of their diet.
Eetclub. Series, reality show, cooking. A group of friends take turns inviting each other to a home made dinner. Each episode covers the process of preparing the meal: deciding the menu, getting the ingredients, cooking and preparing the decoration. At the end of each episode, the guests rate the dinner and give advice to the cook.

Dier & Zorg, catfood that tastes good and creates a bond with the owner
Life stories. Documentary series about the natural behavior of wild animals. Fighting, communicating, loving, hunting, eating, procreating: all aspects of animal life are covered. We see special life stories of all kinds of animals, from jellyfish to elephants and from ants to lions.
Wild animal reunions. Special format, nature. Wild Animal Reunions compiles the internets most moving animal reunions with actual reunions as they happen in the wild. Keepers who raised orphan elephants return to the wild to find them and see whether their bond has stood the test of time.
Shark tale. Animation adventure comedy. When a son of a gangster shark boss is accidentally killed while on the hunt, his would-be prey and his vegetarian brother decide to use the incident to their own advantage.

Simyo, the telecom provider that takes care of their customers every day
Liefde Via Een App. Documentary that explores the world of mobile dating apps. In the present time there is fast and intensive living, leaving little time for the search for love. The solution is the use of mobile dating apps that enable fast partners in the vicinity
App. Film, thriller. A young psychology student is drawn into the dark and fearful world of a diabolic and mysterious App that starts to terrorize her, distributing compromising photographs, videos and text messages about herself and delves deeper and deeper into her personal life, flawlessly exposing all of her deepest secrets.
Peter R. Internet pesters. Reality TV. In every episode, De Vries helps people who are bullied or threatened via the internet. The program tries to find out the often anonymous internet testers and confront them with their actions.

Nowgo, travel insurance that can be bought anywhere and quickly
Richard Hammond's Jungle Quest. Reality show, travel, nature. Jungle quest features the presenter travelling to the Amazon to photograph animals including three-toed sloth, pink river dolphins and harpy eagles. The episodes, were produced in association with Sky Rainforest Rescue, Sky's partnership with WWF.
David Beckham: Into the Unknown. Travel, football, sports: After 22 years of playing for the world's greatest football teams, David Beckham has retired and he has the freedom to do whatever he wants. To mark the occasion he's going on an adventure. He's chosen Brazil. Starting with a beach foot-volleyball game in Rio, they travel deep into the Amazon, ending up with the remote Yanonami tribe, with David desperately trying to explain the beautiful game.

Philips Hue, light customisable lightbulbs for people who love music
Lights out! Documentary, industry, health. Lights Out! joins leading scientists in the lab and in the field to discover how much harm light at night may be causing people and to learn about the ground-breaking steps being taken to protect us.
Gabbers. A documentary about the gabber music scene. It includes images from the 1995 loladamusic's gabber documentary and interviews with former gabbers in their forties who tell us their experiences in the clubs, with the drugs and fashion.

New kids on the block - Live. Music - Live show. Live show from the boys band New Kids on The Block recorded in 1990 in Providence

Concentrate Summer flight, a travel company for young people that makes booking easy
Costa. Series. Romance. Attractive Dutch and Flemish youngsters are 'proppers', which means a nightclub on the Spanish coast pays them to lure other young tourists by their natural -sexy- charms as well as handing out fliers, e.g. They cohabit in free accommodation, so there's ample opportunity for conflict, pranks and love affairs among each-other as well as affairs with 'clients'.
Richard Hammond's Jungle Quest. Reality show, travel, nature. Jungle quest features the presenter travelling to the Amazon to photograph animals including three-toed sloth, pink river dolphins and harpy eagles. The episodes, were produced in association with Sky Rainforest Rescue, Sky's partnership with WWF.
David Beckham: Into the Unknown. Travel, football, sports: After 22 years of playing for the world's greatest football teams, David Beckham has retired and he has the freedom to do whatever he wants. To mark the occasion he's going on an adventure. He's chosen Brazil. Starting with a beach foot-volleyball game in Rio, they travel deep into the Amazon, ending up with the remote Yanonami tribe, with David desperately trying to explain the beautiful game.

Vrijbuiter, clothes and accessories made for adventures outside even in bad weather and rain
Richard Hammond's Jungle Quest. Reality show, travel, nature. Jungle quest features the presenter travelling to the Amazon to photograph animals including three-toed sloth, pink river dolphins and harpy eagles. The episodes, were produced in association with Sky Rainforest Rescue, Sky's partnership with WWF.
David Beckham: Into the Unknown. Travel, football, sports: After 22 years of playing for the world's greatest football teams, David Beckham has retired and he has the freedom to do whatever he wants. To mark the occasion he's going on an adventure. He's chosen Brazil. Starting with a beach foot-volleyball game in Rio, they travel deep into the Amazon, ending up with the remote Yanonami tribe, with David desperately trying to explain the beautiful game.

Andrélon Silver Care, champoo for mature people that makes you look more attractive
The fall and rise of sex. Following the development of Viagra, scientists are now on a quest to help the sexually dysfunctional to maintain a perfect sex life. For baby-boomers reaching retirement, it means that they will be able to remain as sexually active as they were in their youth, thanks to such innovations as the re-engineering of penile tissue.
How to have sex after marriage. Reality TV, romance, sex. Hosted by writer Catherine Townsend, three experts take a couple with a non-existent sex life and try to put some passion back in their bedroom.

Centraal beheer, the damage insurance company for demanding users that also covers your car
's werelds mooiste auto's. Documentary about a the Pebble Beach Concours d'Elegance in California, where the world's most exclusive cars are featured.
Million pound motors. Series, sports. This First Cut documentary explores the exclusive world of the vintage motor trade, meeting one of Britain's most eccentric car salesmen and his super-rich clients
De slechtste chauffeur van Nederland. Reality show, driving. This program is about drivers with a driving license B who have problems with driving a passenger car.
1: Life on the limit. A documentary film that traces the history of Formula One auto racing from its early years, in which some seasons had multiple fatalities, to the 1994 death of Ayrton Senna, the sport's most recent death at the time of production.

Head and shoulders lavander, a shampoo for woman with a new lavander smell that takes care of your dandruff
Roy Donders Stylist van de straat. Roy Donders is the stylist for women from popular neighborhoods and caravan camps and now has his own reality series! The Tilburg stylist, who still lives with his mother, in his reality series gives the viewer a glimpse into his own world, which is connected to glitters, stones and sequins.
Iliza Shlesinger War Paint. War Paint is Iliza Shlesinger's first standup comedy special. Shlesinger's material is highly relatable to women, as she discusses dating, friendship and girls' nights out, but it is just as interesting for men as she provides insight into a number of things women do: wearing make-up, ordering salad on a date, and pretending to like hiking.
Moordwijven. Comedy, film. Three millionaire best friends, Kitty, Estelle and Nicolette, spend their days having botox injections, liposuction and other plastic surgery – all to keep up their looks for their husbands. After witnessing the murder of their favorite plastic surgeon by his wife, the three ponder the idea of similarly killing their husbands should they sleep with someone else.

jandoets.nl Canada, travels to Canada with beautiful landscapes and wild animal spotting.
Richard Hammond's Jungle Quest. Reality show, travel, nature. Jungle quest features the presenter travelling to the Amazon to photograph animals including three-toed sloth, pink river dolphins and harpy eagles. The episodes, were produced in association with Sky Rainforest Rescue, Sky's partnership with WWF.

David Beckham: Into the Unknown. Travel, football, sports: After 22 years of playing for the world's greatest football teams, David Beckham has retired and he has the freedom to do whatever he wants. To mark the occasion he's going on an adventure. He's chosen Brazil. Starting with a beach foot-volleyball game in Rio, they travel deep into the Amazon, ending up with the remote Yanonami tribe, with David desperately trying to explain the beautiful game.

Knorr tacos, food that is healthy and easy to make

In defense of food. "Eat food. Not too much. Mostly plants." With that seven-word maxim, US-based journalist Michael Pollan (The Omnivore's Dilemma) distills a career's worth of reporting into a prescription for reversing the damage being done to people's health by today's industrially driven Western diet. In Defense of Food debunks the daily media barrage of conflicting claims about nutrition.

The best. Cooking show. Every episode focuses on a specific type of dish, including sweet breakfast, Eggs & Lamb, Chocolate & Fish, Sandwiches & steaks.

You are what you eat. Reality TV, the show often uses shock tactics to get the participants to lose weight. In each episode, all food eaten in one week by the person(s) taking part is placed on a table to highlight problem areas of their diet.

Eetclub. Series, reality show, cooking. A group of friends take turns inviting each other to a home made dinner. Each episode covers the process of preparing the meal: deciding the menu, getting the ingredients, cooking and preparing the decoration. At the end of each episode, the guests rate the dinner and give advice to the cook.

Lidl, a supermarket chain with happy and effective workers

Een Kijkje In De Keuken Bij Domino's Pizza. Documentary in which a look in the kitchen is given at one of the most popular fast food chains in the world: Domino's Pizza. A very popular brand among consumers and employees, the so-called 'Dominoids'.

Flying with Qantas. Documentary series in which the Australian airline Qantas is followed by the eyes of the enthusiastic employees, who give everything for the success of this iconic society.

Hospital. Medical documentary that tells the story of the British public healthcare near breaking point and how staff are coping in these unprecedented times in five different London hospitals.

Vacansoleil, holiday campings in the nature with social gatherings.

Richard Hammond's Jungle Quest. Reality show, travel, nature. Jungle quest features the presenter travelling to the Amazon to photograph animals including three-toed sloth, pink river dolphins and harpy eagles. The episodes, were produced in association with Sky Rainforest Rescue, Sky's partnership with WWF.

David Beckham: Into the Unknown. Travel, football, sports: After 22 years of playing for the world's greatest football teams, David Beckham has retired and he has the freedom to do whatever he wants. To mark the occasion he's going on an adventure. He's chosen Brazil. Starting with a beach foot-volleyball game in Rio, they travel deep into the Amazon, ending up with the remote Yanonami tribe, with David desperately trying to explain the beautiful game.