Real Time Video Content Based Contextual Advertisement

Tarun Sharma *
P.E.S Institute Of Technology
100 Feet Ring Road, Banashankari Stage III
Bangalore, India
tarunsharma.pes@gmail.com

Susheel S *
P.E.S Institute Of Technology
100 Feet Ring Road, Banashankari Stage III
Bangalore, India
susheel.suresh@gmail.com

ABSTRACT

With the advent of intermediary commercial ad-networks in charge of optimizing ad selection with a twin goal of increasing revenue and improving user experience, it is preferable to have ads relevant to the web content than generic ones. This method of advertising is popularly known as *Contextual Advertising*. Advertising has become ubiquitous in the internet community and more so in the ever-growing and popular online video delivery websites (e.g. YouTube, Vimeo). Video advertising is becoming increasingly popular on these websites as it has the most user engagement levels.

Our novel method automatically associates ads from an advertisement database and seamlessly recommends them at the appropriate time within each individual video in real-time. If a context switch occurs in the video a different relevant ad is placed. Ads are in the form of banners (containing visuals and text) and are placed beside a playing video. Unlike most video sites which treat video advertising as general text advertising by displaying general, contextually irrelevant video ads at the beginning, end and beside a playing video, our approach aims to recommend contextually relevant ads for a video stream taking into account the audio/speech aspect. Specifically, given a Web page containing an online video, our method is able to extract keywords/text in the form of phrases from the video segment, perform entity extraction on the phrase, detect and resolve ambiguous entities and finally place relevant ads.

Categories and Subject Descriptors

H.3.3 [[**Information Storage and Retrieval**]: Selection process

Keywords

Contextual Advertising, Computational Advertising, Video Advertising

---

*These two authors contibuted equally

1. INTRODUCTION

Digital screens are an indisputable part of consumer's lives. Interactive advertising is taking on an even more vital role in today's marketing mix and the $27.5 billion Internet Ad revenue during the first half of 2015 according to IAB [2] is a testimony. This figure is a 19% increase over 2014. The annual revenues for 2014 totaled at $49.5 billion, $6.7 billion (or 15.6%) higher than in 2013. These raising figures confirm its growing importance and has made it clear that internet ads are imperative in reaching key demographics with creative and tailored messaging.

Driven by the coming age of the Internet and the advent of near-ubiquitous broadband Internet access, online delivery of video content has surged to an unprecedented level. Today's online users face a daunting volume of video content be it from video sharing or blog content, or from IPTV and mobile TV. YouTube alone has over a billion users (almost a third of all people on the Internet), every day people watch hundreds of millions of hours of YouTube videos which generate billions of views. The number of hours people spend watching videos (aka watch time) on YouTube has increased by 60% y/y, the fastest growth seen in 2 years time. With video content in 76 different local languages, it covers 95% of the Internet population. The mobile revolution has made available video content in the hands of users and 2014 stats [5] show that, the average viewing session is more than 40 minutes, that's an increase of more than 50% w.r.t 2013. The number of hours people spent watching videos on mobile has also increased by 100% y/y. These statistics strongly back the idea that people feel more connected with the content while watching videos as it's more entertaining.

As reported by Online Publisher Association [1], the majority (66%) of Internet users have ever seen video ads, while 44% have taken some action after viewing ads. Accordingly, spending on online video advertising is dramatically increasing. To take advantages of the video form of information representation, video-driven contextual advertising, which associates advertisements with an online video or a Web page containing videos, has become a key online monetization strategy.

In this paper we provide a high level design approach with implementation for real time contextual advertisements in videos, making use of existing infrastructure and technologies. We have experimented our methodology for videos hosted on Youtube and the code is made open source.

2. OVERVIEW OF CONTEXTUAL ADS

Contextual advertising refers to the placement of commercial ads within the content of a generic web page. In contextual advertising usually there is a commercial intermediary, called an ad-network, in charge of optimizing the ad selection with the twin goal of increasing revenue (shared between publisher and ad-network) and improving user experience. Given a page, rather than placing generic ads, it seems preferable to have ads related to the content to provide a better user experience and thus to increase the probability of clicks.

Contextual advertising usually falls into the category of direct marketing (as opposed to brand advertising), that is advertising whose aim is a "direct response" where the effect of a campaign is measured by the user reaction. One of the advantages of online advertising in general and contextual advertising in particular is that, compared to the traditional media, it is relatively easy to measure the user response. Usually the desired immediate reaction is for the user to follow the link in the ad and visit the advertiser's web site and, as noted, the prevalent financial model is that the advertiser pays a certain amount for every click on the advertisement (PPC). The revenue is shared between the publisher and the network.

3. RELATED WORK

A lot of papers have talked about contextual advertising for web pages based on pure text extraction [6, 10], using semantic and syntactic feature of phrases [7], using regression for finding relevancy between page content and ads [8] and by sentiment analysis [9] but only a hand few of them have talked about placing relevant ads for media especially in videos [13, 14, 11, 12]. To make video content more enriching, previous work in literature have attempted to place product advertisement in a specific regions of sports videos. VideoSense [11] is designed for general online videos rather than specific videos and it talks about inserting/placing ads using visual cues from the video. Most of the work done focuses on domain specific ad placement and text-based web page advertising.

Contextual video advertising that associates ads within an online video content has not yet been studied adequately. Compared with web page text, video has distinctive characteristics so text-based advertising cannot be directly applied to video advertising. As video is an information intensive media, embedding multimodal tracks we argue that contextual relevance between ads and videos should not be measured only based on names or description of videos . Instead, such relevance should come from the multimodal relevance, including textual, visual, and aural modalities. There has been considerable research work on taking into account the visual aspect of videos, finding contextual relevance and in turn suggesting ads. In this paper we focus on the speech aspect of videos, because we feel that taking into account only the visual aspect becomes very irrelevant in many cases like for eg:

- In game play videos, the content creator might talk about the various hardware devices being used while the visual clipping just shows the game being played.
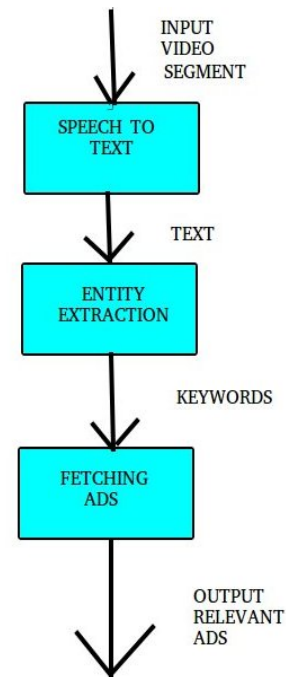

Figure 1: High Level Design For Our Methodology

   Ads about the hardware will be missed out.

- Many channels on Youtube are reviews, video-logs, broadcasts and all of these are speech rich compared to the visuals.

Thus our method will help ad providers recommend better ads for video content which will eventually boost sales.

4. METHODOLOGY

In our approach, shown in **Figure 1**, relevant keywords are extracted from the video content and appropriate ads are selected. This involves three phases, real time speech to text conversion, keyword extraction from raw text, and selecting relevant ads.

4.1 Preprocessing: Speech to Text Conversion

In order to get the context of the video, we use the audio extracted from the video to search for relevant keywords. During cases where video captions are present, we can directly read off the captions file before the video is loaded. Since many multimedia providers like YouTube, Vimeo , DailyMotion , contain millions of videos without captions, a real time speech to text conversion method is required. Google has provided an open source API for this called Google SpeechToText API [4] which takes an audio file encoded in flac format, and returns text in JSON format. Since this does not work in real time, we extract audio every four seconds, encode the audio as a flac file, and send it to the Google SpeechToText API. We then use the text recieved for further processing.

4.2 Real Time Keyword Extraction

Table 1: Phrase in Video and Corresponding Ad

Phrase Used	Generated Ad
I drive a BMW car	BMW product ads, BMW car rental
I like to play basketball	Nike Jordan ad, Adidas basketball
MacBook Air is very light	Apple MacBook Air, MacBook Pro ad
Life in Australia	Australia.com ad, Travel packages for australia
A Logitech mouse is perfect	Logitech mouse ads

The raw text from the video contains words upto four seconds of speech which amounts upto 13-14 words on average. We need to extract important keywords from this text, which would be relevant in terms of advertisements like company names (for other products from same company), product names (for similar products from other companies), places (for tourism packages), names of sports (for sporting items), automobile car companies (for rentals) etc. Manually preparing a dataset for all of these various companies and entities would result in a huge dataset which would constantly need to be updated. Also a search query in a dataset of this size would take a lot of time and this would not be suitable for a real time application like this. It is for these reasons that we have used IBM's AlchemyAPI [3] for this task. AlchemyAPI uses sophisticated statistical algorithms and natural language processing technology to analyze the given text and extract entities from it such as company names, organizations, cities, people etc. To extract context, and to account for nicknames and name variations, AlchemyAPI uses entity disambiguation. When a potentially ambiguous entity is encountered, the surrounding text is examined for contextual cues. AlchemyAPI's disambiguation engine employs tens of millions of hints describing traits of the world's objects, individuals, locations and returns results within microseconds. This API was ideal for the required purpose and hence was used.

4.3 Advertisement Selection

For the final step, we use the entities generated in the previous step to fetch and return relevant ads. We have created a database of advertisements (pictures) and their corresponding description tags. Using the entities generated above, we match the description tags and return the most relevant advertisement. An alternative to this could be to scrape pictures from certain search engines in real time by adding the word 'ad', to the returned entity. As the topic of conversation (in the video) changes, new keywords are generated and this results in different ads. This entire process is done in real time thus providing contextually relevant ads alongside the video, in real time.

5. EXPERIMENTS AND RESULTS

5.1 Speech to Text

Since the ads to be generated in real time next to a video, a speech to text api supporting many different web browsers is required. There are many speech to text convertors available like IBM Speech to Text, AT&T Speech to Text, Wit.ai etc. We have decided to use Google SpeechToText API as it is open sourced and gives good results. Also the response time is in microseconds and is hence suitable for a real time application. Google SpeechToText API also supports multiple versions of popular web browsers like Google Chrome

and Mozilla Firefox . It requires a FLAC encoder which is inherently present in all Windows, OSX and Linux systems based on a i385 architecture. The API is written in PHP and is hence ideal to be used with webpages and servers. The API requires an api key, which can be obtained from google developers website for free. It returns a JSON file which contains the converted text along with the confidence.

5.2 Keyword extraction and Ad selection

For keyword extraction we have used IBMs AlchemyAPI. We have used AlchemyAPI's PHP API in order to make the entire framework run easily on any browser as a web page. In order to test the entire framework, we created a webpage where the user could entire the video ID of any YouTube video and ads would be generated in real time next to the video. We also created our own 30 second video where we use phrases like "I like to play basketball", "I own a HTC phone" etc. Different ads are generated based on the current topic in discussion in real time fashion. The results of phrase spoken against the top two matched ads are shown below in **Table 1**. The ads generated here are ads which are part of our collected dataset and more accurate results can be obtained by using a larger dataset or by using web scraping. The link[1] to the video demo showing real time contextual ad generation is given below. Here we show the generated ads next to the video and also show the entities extracted. We have also open sourced the entire project code [2] on GitHub for others to use and modify.

6. CONCLUSION

In this paper, through our experiments we have shown that using existing infrastructure for speech to text conversion and keyword extraction we can suggested relevant ads for videos in a web page. Our approach combined with the VideoSense [11] methodology will cover all aspects (visual and aural) for understanding a video and thus can significantly increase the relevance of an ad being placed for a video. Our implementation provides a good starting point for developers to further improve ad relevance.

User-targeted advertising is another key for online advertising in addition to contextual advertising. Targeted video advertising means that video ads will reach specified target audiences by leveraging user-provided demographic profiles. To support such a advertising framework, our future work will include collecting user profiles from browser history and social media, study how to deliver personalized video ads based on user interests, locations, past and current behaviors and implement the future study with the model presented in this paper so as to create a robust advertisement

---

[1]Youtube Video: youtube.com/watch?v=OdpZLuuF3Zc.
[2]github.com/susheels/contextual-advertisement-generator

selection framework which can garner higher click rates and create an ecosystem where in ads can co-exist with content without any resentment from users.

Ads selected with a mix of features like context and user profiles may be expected to be relevant to the web page containing the source videos rather than the source videos themselves.

7. REFERENCES

[1] Digital content next, previously *Online Publisher Association*. `https://digitalcontentnext.org/`. Accessed: 2015-10-22.

[2] Interactive advertising bureau. `http://www.iab.com/`. Accessed: 2015-10-22.

[3] *Alchemy API*. `http://www.alchemyapi.com/`. Accessed: 2015-10-22.

[4] *Google Speech-To-Text API*. `https://developers.google.com/apis-explorer`. Accessed: 2015-10-22.

[5] Youtube statistics. `www.youtube.com/yt/press/statistics.html`. Accessed: 2015-10-22.

[6] A. Addis, G. Armano, A. Giuliani, and E. Vargiu. A recommender system based on a generic contextual advertising approach. In *Computers and Communications (ISCC), 2010 IEEE Symposium on*, pages 859–861. IEEE, 2010.

[7] A. Broder, M. Fontoura, V. Josifovski, and L. Riedel. A semantic approach to contextual advertising. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 559–566. ACM, 2007.

[8] D. Chakrabarti, D. Agarwal, and V. Josifovski. Contextual advertising by combining relevance with click feedback. In *Proceedings of the 17th international conference on World Wide Web*, pages 417–426. ACM, 2008.

[9] T.-K. Fan and C.-H. Chang. Sentiment-oriented contextual advertising. *Knowledge and Information Systems*, 23(3):321–344, 2010.

[10] T.-K. Fan and C.-H. Chang. Blogger-centric contextual advertising. *Expert Systems with Applications*, 38(3):1777–1788, 2011.

[11] T. Mei, X.-S. Hua, and S. Li. Videosense: A contextual in-video advertising system. *Circuits and Systems for Video Technology, IEEE Transactions on*, 19(12):1866–1879, 2009.

[12] S. H. Sengamedu, N. Sawant, and S. Wadhwa. vadeo: video advertising system. In *Proceedings of the 15th international conference on Multimedia*, pages 455–456. ACM, 2007.

[13] B. Wang, J. Wang, S. Chen, L.-Y. Duan, and H. Lu. Semantic linking between video ads and web services with progressive search. In *Data Mining Workshops, 2009. ICDMW'09. IEEE International Conference on*, pages 196–201. IEEE, 2009.

[14] K. Yadati, H. Katti, and M. Kankanhalli. Cavva: Computational affective video-in-video advertising. *Multimedia, IEEE Transactions on*, 16(1):15–23, 2014.