



Prediction of preferred Advertising Formats in e-Advertisements by using Data Mining Techniques

Shilpi
Ph.D. Scholar
Mewar University,
Chittorgarh, Rajasthan (India)

B.K. Sharma

ABSTRACT

e-Advertisements are extraordinarily cost effective due to targeted and focused marketing, nominal wastage, increasing customer-base and unmatched tracking capabilities. Internet enables instinctive two-way relationship with the larger target customers. The use of IT applications and data mining softwares are providing e-advertisement related information specifically product analysis, product promotion, demand forecasting, trends analysis and new product development.

This paper gives brief view about various advertising format trends. Various Advertisement formats related data of e-advertisement from 2002 to 2014 have been collected and applied the tools and techniques of data mining for finding similar cluster. Subsequently, statistical techniques have been applied on similar cluster for predicting preferred advertising formats in e-ads.

Effective use of data mining plays a vital role in Ads formats like Search, Rich Media Display Banners, Classifieds, Lead Generation & Digital Video, Sponsorship etc in e-advertisements.

Keywords

Data Mining, Clustering, e- advertisement, Ads formats.

1. INTRODUCTION

In Internet era, e-Advertising is a vital activity to a company, needs to promote their products and services and to publicize their name amongst the general public. Marketers need to understand the implication of various dimensions guiding the fragmentation on the basis of consumer behavior. The challenge for a marketer is to have a deeper understanding of impact of every media so that they can selectively target appropriate customers for their message delivery. The Interactive Advertising Bureau (IAB) has launched a detail to suggest standards and best practices for display advertising. It initiates a new set of paradigms for advertising.[1]

Data mining is the assignment of discovering interesting patterns from bulky amounts of data where the data can be stored in databases, data warehouses, and other repositories. Data mining engages an integration of techniques from multiple disciplines such as database technology, statistics, machine learning, high performance computing, pattern recognition, data visualization, information retrieval, neural network, signal and image processing and spatial data analysis. By completing data mining, interesting knowledge, high-level information can be mined from databases and viewed or surfed from different angles. The discovered knowledge can be applied to information management, decision-making, process control, and query processing. It is calculated that the volume of sales generated by Internet is substantial and it may be evaluated by using data mining tools and techniques.[2]

Classification and Cluster analysis are important techniques that partition objects have many points into meaningful disjoint subgroups in sequence that objects in each group are more like to each other in the values of their attributes than they are to objects in other groups.

Clustering can be described as the process of separating a set of objects into several subsets on the basis of their similar properties. The aim is to define clusters that minimize intra-cluster changeability while maximizing inter-cluster distances, i.e. finding clusters, which members are like to each other, but distant to members of other clusters. Clustering is a type of learning by observation rather than learning by examples. Therefore, Clustering methods try to find an approximate or local optimum solution.[3]

2. E-ADVERTISEMENT

Internet display advertisement has regularly formed a sizeable portion of overall Internet advertisement market. It has maintained to be an effective long-term selection for electronic marketers in targeting consumers with specific features. Although the share of other selections such as search engine marketing may suppose to larger proportions over time, electronic advertisement will maintain to command a significant part of this spending

The Internet has altered forever almost every phase of our lives and has drastically effected overall communication. With a potential to target millions across the globe instantaneously, the Internet offers advertisers and communication specialists, opportunities like never before, was the challenge. The challenges we face on the Internet are of precise delivery of message, measurement of effectiveness and focused targeting audience. And these are the challenges if we handle effectively, makes Internet advertising a reality, presenting site publishers and advertisers an opportunity to control the power of the Internet as an effective marketing tool [4].

The changes from traditional advertising to e-Advertisement, Advertiser are facilitated with prospective gains. There are unique characteristics of e-Advertisement which offline advertisement lacks. Internet enables intuitive two-way relationship with the larger target customers. But the advertiser loses some of its ability to control and publicize the information since the customer is in control and gathers the information of his choice just at a mouse click

The major benefit of e-advertising is the immediate publishing of information and content that is not limited by geography or time. To that end, the emerging area of interactive advertising presents fresh challenges for advertisers who should adopted an interruptive strategy.



3. SEVERAL ADDRESSING FORMATS

3.1 Display/Banner Advertising

Advertiser offers an Internet company for space to display a static or logo or hyper-linked banner on one or more of the net company's pages.[9]

3.2 Sponsorship Advertising

It shows custom content and/or experiences created for an advertiser which may or may not comprise ad elements such as advertorial or pre-roll video, display advertising, brand logos. Sponsorships fall into numerous categories:

- Spotlights are custom built webpages integrating an advertiser's brand and housing a collection of content usually around a Subject.
- • Advergaming choice from an advertiser buying all the ad units around a game or a "sponsored by" link to generating a custom branded game experience
- • Content & Section Sponsorship is when an advertiser absolutely sponsors a particular section of the website or email (usually existing content) re-skinned with the advertiser's branding
- • Contests and Sweepstakes can range from branded sweepstakes on the website to a full-fledge branded contest with submissions and judging.

3.3 Email Advertising

Advertiser or Links sponsorships, Banner ads that appear in email marketing campaigns, email newsletters and other commercial email communications. It comprises all types of electronic mail [4].

3.4 Search Advertising

Fees Advertisers reimburse Internet companies to list their company site domain name to a specific search word or phrase. Search Categories consist of:

3.4.1 Paid Listings

Text Links show at the top or side of search results for specific keywords. The more a marketer gives, the higher the position it gets. Marketers only reimburse when a user clicks on the text link.

3.4.2 Contextual Search

Text Links show in an article based in the reference of the content, in its place of a user submitted keyword. Payment only occurs when the text link is clicked.

3.4.3 Paid Inclusion

guarantees that a marketer's URL is guided by a search engine. The listing is decided by the engine's search algorithms.[4]

3.4.4 Site optimization

change a website to make it easier for search engines to automatically index the website and optimistically result in better placement in results.

3.5 Lead Generation Advertising

Fees Advertisers reimburse to online advertising companies that submit qualified purchase inquiries or provide consumer information where the consumer selects into being contacted by a marketer. These processes are valued on a performance

basis, and can comprise user applications, contests, surveys or registrations.

3.6 Classified and Auction Advertising

Fees Advertisers reimburse Internet companies to list specific products or services (e.g., online job boards and employment listings, automotive listings, auction-based listings, real estate listings, yellow pages).[4]

3.7 Rich Media Advertising

Display-related ads that incorporate some component of streaming interactivity. Rich media ads often consist of flash or java script, but not text, and can allow users to analysis and interrelate with products or services. Rich Media is most preferred medium. Video commercials that show in video players are considered Digital Video Ads, not Rich Media. "Interstitials" have been merged within the rich media category and represent full- or partial-page text and image server-push advertisements which emerge in the transition between two pages of content. Forms of interstitials can contain a variation of the following terms:

3.7.1 Splash screens e-Ads

A preliminary page that leads the regular home page of a website that usually supports a particular site feature or offers advertising. A splash page is timed to move onto the home page after completing a short period of time.

3.7.2 Pop-up e-Ads and Pop-under e-Ads

An Advertisement that show in a separate window which automatically appears over an existing content window, without a linked banner.

3.7.3 Daughter windows e-Ads

An Advertisement that shows in a separate window associated with a concurrently displayed banner. The content and banner are typically exhibited first, pursued by the daughter window.

3.7.4 Superstitials e-Ads

Ads that are different from interstitials because of the much higher e-Ad quality, and that they play instantly.[4]

3.8 Digital Video Advertising

Advertising that shows before, during or after digital video content in a video player (i.e. pre-roll, mid-roll, post-roll video e-Ads). Digital Video e-Ads comprise TV commercials online and can show in streaming content or in downloadable video. Display-related e-Ads on a page that contain video are categorized as rich media e-Ads. [4]

Video Overlays are also classified as Digital Video Advertising. Video overlays contain small e-Ads that exist on top of digital video content. They can show to be text, display, video, rich media or another ad format but are included within the video player.

3.9 Mobile Advertising

Advertising tailored to and carried through wireless mobile devices such as smartphones, feature phones, and media tablets. Typically taking the form of static or rich media display e-Ads, text messaging e-Ads, search e-Ads, or audio/video spots, such advertising generally shows within mobile websites, mobile applications, text messaging services or within mobile search results.



Mobile Advertising formats include: Search, Display (banner ads, digital audio, digital video, sponsorships, and rich media), and other e-advertising served to mobile devices. [4]

3.10 Digital Audio Advertising

It submits to partially or entirely advertising-sustained audio programming available to consumers on a streaming basis, sent via the wired and mobile Internet. This consists of a wide range of services, such as the following: [4]

- Net audio streams of terrestrial radio stations;
- Purely online radio stations, with either professional or amateur DJs;
- Personalized (i.e., without human editors/DJs) and on-demand, streamed audio services that generate playlists based on user preferences of artists, tracks, or genres;

- Music or spoken word audio content sent within a different website or application, e.g., in game music services.

3.11 Social Media Advertising

e-Advertising delivered on social platforms, including social gaming websites and apps and social networking, across all device types, including desktop, smartphone, laptop and tablet.

4. CLUSTERING TECHNIQUE IN e-ADVERTISEMENTS

e-Advertisement plays a vital role to promote by using various types of advertisement formats. According to the IAB reports, the table of percentage of utilization of e- advertisements by various advertising formats is shown. That shows increments annually. 13 years data of various addressing formats have been collected from years 2002 to 2014. The summarized data are given Table-1.[5]

Table 1 Various Advertising Formats

Year	Display/ Banner	Sponsor ships	Classified	Lead Generation /Ref	E-Mails	Search	Rich media	Mobile	Others
2002	29	18	15	1	4	15	5	0	13
2003	21	10	17	1	3	35	8	0	5
2004	38	0	18	3	2	39	0	0	0
2005	34	0	17	6	2	41	0	0	0
2006	22	3	18	8	2	40	7	0	0
2007	21	3	14	7	2	42	9	0	2
2008	21	2	14	7	2	45	7	0	2
2009	22	2	10	6	1	47	7	0	5
2010	24	3	10	5	1	46	6	3	5
2011	22	4	8	5	0	47	6	5	3
2012	21	2	7	5	0	46	3	9	7
2013	19	2	6	4	0	43	3	17	7
2014	16	2	5	4	0	38	3	25	7

4.1 Step 1 & 2

Let the three seeds be the first three years from 2002 to 2004 as Table-2.

Table 2 Initial Seeds

c1	29	18	15	1	4	15	5	0	13
c2	21	10	17	1	3	35	8	0	5
c3	38	0	18	3	2	39	0	0	0

4.2 Step 3 & 4

Now compute the distances using the 9 attributes and using the sum of absolute differences for simplicity by using K-median. The distance values for all the objects are given in

Table-3 Where columns 11, 12 & 13 give the three distances from the three seeds respectively. Based on these distances, every year is allocated to the nearest cluster. We gain the first iteration result as shown in Table -3.[5]



Table 3: First Iteration – allocating each object to the nearest cluster

Code	Display/ Banner	Sponsor ships	Classified	Lead Generation	E-Mails	Search	Rich Media	Mobile	Others	Difference From Clusters			Min. value of Cluste rs
										C1	C2	C3	
Y1	29	18	15	1	4	15	5	0	13	0	50	76	C1
Y2	21	10	17	1	3	35	8	0	5	50	0	48	C2
Y3	38	0	18	3	2	39	0	0	0	76	48	0	C3
Y4	34	0	17	6	2	41	0	0	0	76	48	10	C3
Y5	22	3	18	8	2	40	7	0	0	74	28	32	C2
Y6	21	3	14	7	2	42	9	0	2	74	28	42	C2
Y7	21	2	14	7	2	45	7	0	2	76	32	42	C2
Y8	22	2	10	6	1	47	7	0	5	78	36	50	C2
Y9	24	3	10	5	1	46	6	3	5	75	39	49	C2
Y10	22	4	8	5	0	47	6	5	3	84	44	56	C2
Y11	21	2	7	5	0	46	3	9	7	88	52	60	C2
Y12	19	2	6	4	0	43	3	17	7	95	59	67	C2
Y13	16	2	5	4	0	38	3	25	7	102	66	76	C2

4.3 Step 5

Table -4 compares the cluster means of cluster found in Table-3 with the actual seeds.

Table 4: Comparing Seeds

Code	Display/ Banner	Sponsor ships	Classified	Lead Generation	E-Mails	Search	Rich Media	Mobile	Others
C1	29	18	15	1	4	15	5	0	13
C2	20.9	3.3	10.9	5.2	1.1	42.9	5.9	5.9	4.3
C3	36	0	17.5	4.5	2	40	0	0	0
A1	29	18	15	1	4	15	5	0	13
A2	21	10	17	1	3	35	8	0	5
A3	38	0	18	3	2	39	0	0	0

4.4 Step 3 & 4

Use the new cluster means to recalculated the distance of each object to each of the means, again allocating each object to the nearest cluster. Table-5 illustrates the second iteration result.

Table 5: Second Iteration – allocating each object to the nearest cluster

Code	Display/ Banner	Sponsor ships	Classified	Lead Generation	E-Mails	Search	Rich Media	Mobile	Others	Difference From			Min. value of Clusters
										C1	C2	C3	
Y1	29	18	15	1	4	15	5	0	13	0	77.4	76	C1
Y2	21	10	17	1	3	35	8	0	5	50	35.6	48	C2
Y3	38	0	18	3	2	39	0	0	0	76	50.6	5	C3
Y4	34	0	17	6	2	41	0	0	0	76	42.2	5	C3
Y5	22	3	18	8	2	40	7	0	0	74	26.4	28	C2
Y6	21	3	14	7	2	42	9	0	2	74	18.4	37	C2
Y7	21	2	14	7	2	45	7	0	2	76	18.6	37	C2
Y8	22	2	10	6	1	47	7	0	5	78	16	45	C2
Y9	24	3	10	5	1	46	6	3	5	75	11.4	44	C2
Y10	22	4	8	5	0	47	6	5	3	84	12.4	51	C2
Y11	21	2	7	5	0	46	3	9	7	88	18.4	55	C2
Y12	19	2	6	4	0	43	3	17	7	95	27.2	63	C2
Y13	16	2	5	4	0	38	3	25	7	102	44	74	C2

Now, the clusters in Table 3 & 5 are same. Therefore, these clusters membership is as follows:

Cluster 1 : Y1

Cluster 2: Y2, Y5, Y6, Y7, Y8, Y9, Y10, Y11, Y12, Y13

Cluster 3 : Y3,Y4.

Through the reports of IAB organization, data has been collected and analyzed for different perspective like usage of e-advertisements of various addressing formats etc with the help of data mining tools and techniques.



5. FINDINGS

The utilization of several addressing formats in e-advertisements is varied continuously as shown in Fig 1. It has to be calculated that what would be the utilization of e-ads in coming years. Here we also applied regression technique for calculating estimated utilization of several advertising formats

in e-advertisements. Linear regression is appropriate for one input attribute. The estimated results of various addressing formats in the year 2018 and 2020 are given in Table-6 and Graphical representation of various addressing formats and estimated results can be displayed as shown in Fig. 1 and Fig 2 respectively.

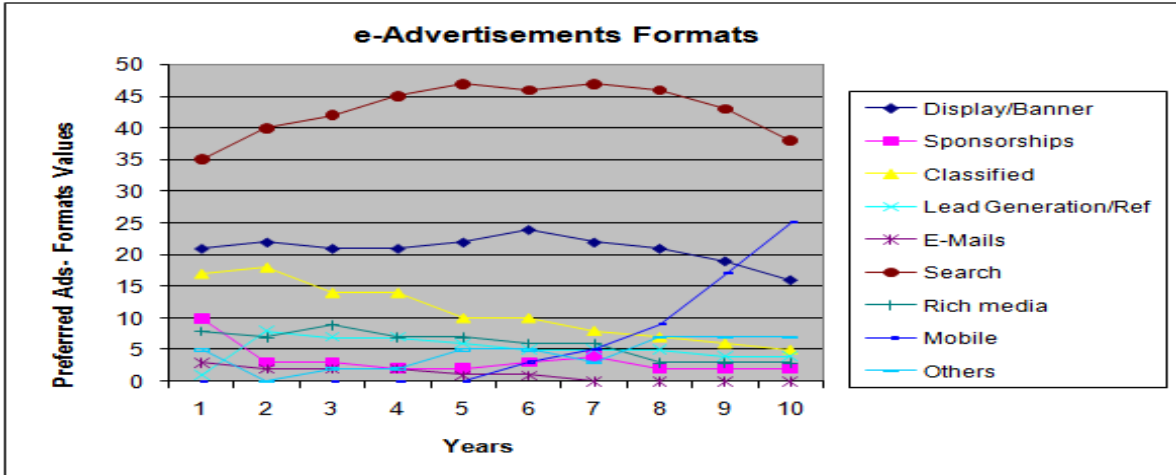


Fig 1: Graphical Representation of Various Addressing Formats

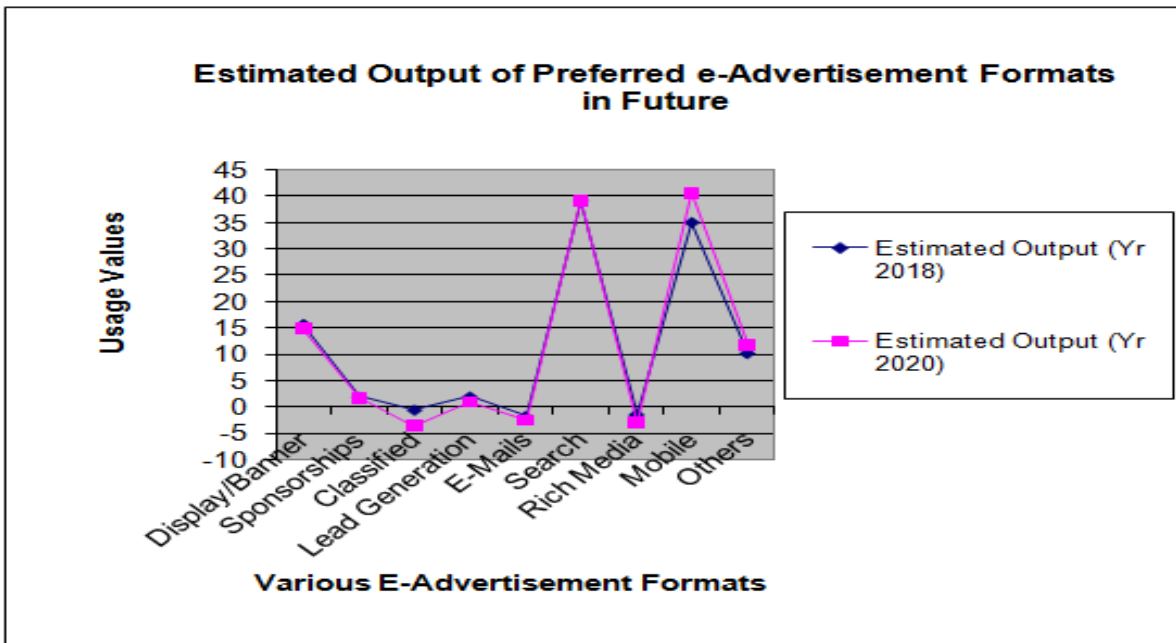


Fig 2: Estimated Output of preferred Advertising Formats in Year 2018 & Year 2020

Table 6: Estimated Results for Year 2018 & Year 2020

Ads Formats	Estimated Output (Yr 2018)	Estimated Output (Yr 2020)
Display/Banner	15.72727273	14.90909091
Sponsorships	1.954545455	1.818181818
Classified	-0.5	-3.5
Lead Generation	2	1
E-Mails	-1.727272727	-2.409090909
Search	39.18181818	39.22727273
Rich Media	-1.363636364	-2.954545455
Mobile	35.04545455	40.68181818
Others	10.22727273	11.90909091

Mobile Advertising Format has risen from 5% in FY11 to the current level of 25% for FY14. This is expected to reach 35% and 40% of total advertising formats for FY 2018 and FY 2020 respectively as shown in Fig 2.

6. CONCLUSION

On-line marketers require understanding the implication of various parameters guiding this fragmentation basis of the consumer behavior. The challenge for a seller is to have a good understanding of every medium so that they can selectively target appropriate customers for their message delivery. e- Advertising eliminates the middlemen and provides brands online with the unique opportunity to have a direct relationship with their customers.



Data mining tools calculate future trends and behaviors, allocating businesses to make positive, knowledge-driven decisions. Since Cluster analysis is explanatory, but the results of cluster analysis can be difficult to interpret. As numeric data of several years had been shown according to the IAB Revenue reports, The last 9 years data made a common cluster. Therefore, by applying regression technique on same cluster, the expected results have been found and investments can be made according to these analysis.

7. REFERENCES

- [1] Alex Beron, Stephen J. Smith, Data Warehousing, Data Mining & OLAP, Tata McGraw-Hill Edition pp336, 2004.
- [2] Bryan Wiener, “Interactive Advertising Forecast (U.S.)”, Forrester Research. April 2009.
- [3] G. K. Gupta, “Introduction to Data Mining with Case Studies”, Prentice-Hall of India, ed 2006
- [4] IAB, “Data Usage & Control Primer”: best practices & definitions, May 2010
- [5] Jiawei Han, Micheline Kamber, Data Mining Concepts and Techniques, Morgan Kaufmann Publishers pp28- 30, 451-456,ed 2006.
- [6] Kurt Thearling, An Introduction to Data Mining, CSI Vol 30, Oct 2006, No. 7 pp 4-6.
- [7] PricewaterhouseCoopers, “IAB Internet Advertising Revenue Report” 2013 Full-Year Results, April 2014
- [8] PricewaterhouseCoopers, “IAB Internet Advertising Revenue Report” 2012 Full-Year Results, April 2013
- [9] PricewaterhouseCoopers, “IAB Internet Advertising Revenue Report” 2011 Full-Year Results, March 2012
- [10] PricewaterhouseCoopers, “IAB Internet Advertising Revenue Report” 2010 Full-Year Results, May 2011
- [11] PricewaterhouseCoopers, “IAB Internet Advertising Revenue Report” 2009 Full-Year Results, April 2010
- [12] PricewaterhouseCoopers, “IAB Internet Advertising Revenue Report” 2008 Full-Year Results, March 2009
- [13] PricewaterhouseCoopers, “IAB Internet Advertising Revenue Report” 2007 Full-Year Results, May 2008
- [14] PricewaterhouseCoopers, “IAB Internet Advertising Revenue Report” 2006 Full-Year Results, May 2007
- [15] PricewaterhouseCoopers, “IAB Internet Advertising Revenue Report” 2005 Full-Year Results, April 2006
- [16] PricewaterhouseCoopers, “IAB Internet Advertising Revenue Report” 2004 Full-Year Results, April 2005
- [17] PricewaterhouseCoopers, “IAB Internet Advertising Revenue Report” 2003 Full-Year Results, April 2004
- [18] PricewaterhouseCoopers, “IAB Internet Advertising Revenue Report” 2002 Full-Year Results, June 2003