

Priority Recommendation System in an Affiliate Network

Zeeshan Khawar Malik, Colin Fyfe and Malcolm Crowe
University of The West of Scotland, Paisley, Scotland, UK
{zeeshan.malik, colin.fyfe, malcolm.crowe}@uws.ac.uk

Abstract—Affiliate Networks are the main source of communication between publishers and advertisers where publishers normally subscribe as a service provider and advertisers as an employer. These networks are helping both the publishers and advertisers in terms of providing them with a platform where they can build an automated affiliate connection with each other via these affiliate networks. The problem that is highlighted in this paper is the huge gap that exists between the publisher and advertiser in these affiliate networks and a solution is provided by proposing a priority recommendation system based on K-Means clustering algorithm. Every advertiser desires to have that type of publisher who is already practiced in his category of business or at least has the same skills and talent. This paper presents the concept of a recommendation system based on clustering the real-time data of all the existing transactions of publishers and advertisers of an affiliate network and based on the resulting POST-HOC classified data, a new publisher or advertiser will automatically be classified. Real-time data is provided by Affiliate Future a well-known company among all the affiliate networks. After carefully examining the data the most effective attribute is selected as the base attribute for clustering. The data is encoded into binary numbers for the purpose of clustering. More than one distance approaches are used and the most suitable one is selected for classifying the data.

Index Terms—Affiliate Marketing, Clustering, Publisher, Advertiser, Sammon Mapping

I. INTRODUCTION

Affiliate Networks like Linkshare, E-Junction and Affiliate Future are becoming the key platforms for all e-business people who want to search for a highly ranked and most effective publisher to market their product or service. Similarly on the other hand, publishers also use these affiliate networks to get connected with their choice of product so that they can better perform in terms of marketing and generating revenue. These affiliate networks have boosted the process of affiliate marketing.

Today affiliate marketing has become a key technique to market a product or service and to generate revenue in the shortest time possible [1]. Affiliate marketing has also shown a great impact on other ecommerce strategies as well in terms of generating revenue by making referrals in an n-tier commission-based mechanism [2]. More than one model has been introduced in affiliate marketing

for the purpose of generating revenue that includes primarily percentage of sales model, pay per lead model, flat referral rate model, pay per email model, cost per view model and cost per click model [3]. The basic working in affiliate marketing is a process in which a publisher gets commission for selling an advertiser's product through its own platform and the advertiser confirms its sale by checking the backlink coming from publisher's own platform [4, 5, 6, 7] and [8]. The three oldest affiliate networks are

- 1) Linkshare
- 2) Be Free and
- 3) Cyberotica [18, 36]

Clustering is one of the most popular methods [9] for exploratory data analysis. The prime purpose is to group similar data into one cluster in such a way that data within the cluster are as similar as possible while data in different clusters are as dissimilar as possible. This technique is continuously being refined and considered as a highly studied area of AI, machine learning and statistics. Clustering is always a very important problem for marketing researchers as well in terms of grouping of persons, products, or occasions which may act as a basis for further analysis [10].

K-means clustering [12] is one of the most widely used clustering techniques in commercial environments and works very efficiently on high dimensional data as well [11]. It clusters the data based on initially defined prototypes for each cluster and based on the calculation of the sum of square distances of each point with all the defined prototypes assigns the point to that cluster where the distance is minimum. Then the prototype's positions are updated to be the average of all the data which has been assigned to that cluster.

II. AFFILIATE MARKETING

The concept of affiliate marketing was first introduced by the pioneering company Amazon, headed by Jeff Bezos in the late 20th century. The concept was so much appreciated that many online companies started adopting this technique of marketing to generate revenue in an n-tier mechanism [13, 14, 15]. Amazon has generated a lot

of income through its Amazon Associates Affiliate Program [17]. There are different methods of pricing proposed by researchers under affiliate marketing but the most common methods that are being preferred by most of the merchants are

- 1) Cost per Click Model and
- 2) Cost per Sale Model [16].

The development of affiliate networks is one more reason for the popularity of affiliate marketing and in the absence of affiliate networks, there was no one before who could cross-check the validity of the service provided by the affiliates which resulted in a lot of scams [36]. This is one reason why few empirical researches have been done in this area because of the poor image of affiliate marketing due to inconsistent branding through many non-reliable affiliates and the lack of development of trust due to intrusive mass advertising [20, 4]. The prime objective of affiliate marketing remains the same, that is to generate revenue by selling products or services through additional outlets known as affiliates of the advertisers and in return the affiliates get commission for every sale produced. The pricing model is normally selected depending upon the affiliate model chosen. In the pay per sale model, the commission is given to the affiliate on each sale produced through its platform. In the pay per lead model, the advertisers reward affiliates for each new subscriber coming through their platform. In the pay per click model the advertiser rewards the affiliate for every click or every cost-per 1000 times impressions online users view advertisement [18, 2, 21]. The concept of making affiliates in the real world was first introduced by airlines, hotels and other tourism companies [22].

According to [18] the concept of affiliate marketing first originated in the year 1996 and that is the year when content analysis of 93 articles from three journals most related to marketing was done and the conclusion was that there was very little research of affiliate marketing and most of them failed to meet the current study requirement covering the development of affiliate marketing. The year 1999 is considered to be the year affiliate marketing opened its gates in the UK and resulted in the opening of companies like Commission Junction and Tradedoubler [23]. These affiliate networks are also termed top-tier affiliates which offer key services to merchants such as account tracking and management. Some researchers [24, 2] have also differentiated the affiliates into two types "first tier" and "second tier". The first tier affiliates are large-scale established affiliates having their own brand name and have a relatively large consumer base whereas the second tier affiliates are small-scale individuals who have their own individual platforms through which they offer services to merchants.

III. HARD CLUSTERING

K-means Clustering is

“A major clustering method producing a partition of the entity set into non-overlapping clusters along with within-cluster centroids. It proceeds in iterations consisting of two steps each; one step updates clusters according to the minimum distance rule, the other step updates centroids as the centres of gravity of clusters. The method implements the so-called alternating minimization algorithm for the square error criterion. To initialize the computations, either a partition or a set of all K tentative centroids must be specified” [37].

Let $x_1, x_2, x_3 \dots x_n$ be the number of data points and $c_1, c_2, c_3 \dots, c_n$ be the cluster on the search space, K is assumed as the total number of clusters and n is assumed as the total number of data points. Let $\mu_1, \mu_2, \mu_3, \dots, \mu_k$ be the initially defined prototypes, then clustering of data points into K-means clusters J_k is determined by minimizing the sum of squared errors given in eq.(1).

$$J_k = \sum_{j=1}^k \sum_{i \in c_j} (x - \mu_j)^2 \quad (1)$$

K-means is one of the most popular clustering algorithms among all the algorithms proposed in the literature for clustering: ISODATA [26, 27], CLARA [27], CLARANS [28], Focusing Techniques [29], P-CLUSTER [30], DBSCAN [31], Ecluster [32], BIRCH [34] and GRIDCLUS [33] are all extensions of k-means. Algorithm is shown below which explains the basic working of k-means [12].

DIRECT K-MEANS

Initialize K prototypes ($m_1, m_2, m_3, \dots, m_k$) such that $m_j = i_l, j \in \{1, 2, 3, \dots, K\}, l \in \{1, 2, 3, \dots, n\}$
 Each cluster C_j is associated with prototype m_j
 Repeat
 For each input vector i_l , where $l \in \{1, 2, 3, \dots, n\}$
 Assign i_l to the cluster C_j , with nearest prototype m_j .
 i.e. $|i_l - m_j| \leq |i_l - m_k|, j \in \{1, 2, 3, \dots, k\}$
 For each cluster C_j , where $j \in \{1, 2, 3, \dots, k\}$
 Update the prototype m_j , to be the centroid of all samples currently in c_j so that $m_j = \sum_{i_l \in c_j} (i_l / |c_j|)$

Compute the error function

$$E = \sum_{j=1}^k \sum_{i \in c_j} |i_l - m_j|^2$$

Until E does not change significantly.

IV. DESCRIPTION

In this paper the k-means clustering algorithm [38] is selected for a real-time affiliate network data provided by a company named as "Affiliate Future" in the form of excel files. The data is related to the individual profiles and transactional information of advertisers and publishers.

Publisher's Profile Data

- 1) AffiliateID {Unique ID for each publisher}
- 2) AffiliateSiteID {More than one SiteID for single affiliate ID}
- 3) SiteName {Site Description}
- 4) SiteAddress {Site URL}

Publisher's Category Data

- 1) AffiliateSiteID {More than on SiteID for single affiliateID}
- 2) Category {Category Name}

Advertiser's Profile Data

- 1) MerchantID {Unique ID for each advertiser}
- 2) MerchantSiteID {More than on SiteID for single merchantID}
- 3) SiteName {Site Description}
- 4) SiteAddress {Site URL}

Advertiser's Category Data

- 1) MerchantSiteID {More than one SiteID for single merchant ID}
- 2) Category {Category Name}

Publisher and Advertiser's Transactional Information

- 1) MerchantID {Unique ID for each advertiser}
- 2) AffiliateID {Unique ID for each publisher}
- 3) MerchantSiteID {More than one SiteID for single merchant ID}
- 4) AffiliateSiteID {More than one SiteID for single affiliateID}
- 5) LogDate {Temporal information for each transaction}

There are a total of 46 categories for which all the publishers and advertisers are associated with :- 1) Adult 2) Arts and Craft 3) Auctions 4) Baby Gear 5) Books, Catalogues & Magazines 6) Business Services 7) Clothing & Accessories-Men's 8) Clothing & Accessories - Women's 9) Competitions, Freebies & Discounts 10) Computers 11) Dating 12) DVDs, Videos & Games 13) Eco-Friendly 14) Education 15) Experience 16) Financial & Legal 17) Food 18) Gadgets 19) Gaming 20) Gaming & Gambling 21) Gifts & Flowers 22) Health & Beauty 23) Home & Garden 24) Insurance 25) Internet Services 26) Jewellery 27) Latest Merchants 28) Loans 29) Mobile Phones & Accessories 30) Motoring 31) Music 32) Office Equipment 33) Outdoor Equipment 34) Pets 35)

Posters & Memorabilia 36) Seasonal 37) Sports & Fitness 38) Telecommunications 39) Telecom 40) Toy Shops 41) Travel-Accommodation 42) Travel-Essentials 43) Travel-Flights 44) Travel-Holidays 45) Wedding & Celebrations 46) Wine & Drinks.

The data for both the publishers and advertisers are clustered on the basis of the above mentioned categories. A single publisher can be associated with more than one category and similarly with the advertiser's data. There is more than one publisher and advertiser associated with each category. The same approach for clustering has been used for both publisher's and advertiser's data. Further the clustering of only advertiser's data is explained. The total number of prototypes taken for clustering advertiser's data is 46 i.e. one for each category. The reason for taking 46 prototypes with each prototype belonging to each category is to classify the data in a way that each cluster resembles a category which are 46 in total. The data is first encoded into binary numbers. Total 46 bits are associated with each advertiser. Out of these 46 bits, '1' is for On and '0' is for off. A single advertiser can have more than one '1' in their total bits. This means that a single advertiser can be associated with more than one category. The initial values of half of the prototypes are shown in Table 1.

IV. FINAL PROTOTYPES RESULT

More than one distance approach are used to cluster data using the k-means algorithm shown in Table 2. The most suitable proved to be the euclidean distance approach which is shown to be the most compatible distance approach for the k-means algorithm [12]. Some of the final values of the 46 dimensional prototypes is shown in the tables below.

TABLE 2. DISTANCE APPROACH

Measure	Forms
Euclidean Distance Approach	$D_{ij} = \left(\sum_{i=1}^d x_{ii} - x_{ji} ^{1/2} \right)^2$
Manhattan Distance Approach	$D_{ij} = \sum_{i=1}^d x_{ii} - x_{ji} $
Cosine Similarity Approach	$S_{ij} = \cos \alpha = \frac{x_i^T x_j}{\ x_i\ \ x_j\ }$
Dot Product Approach	$D_{ij} = \ x_i \cdot x_j\ $

In order to further elaborate the Sammon mapping mathematically. Let us suppose the distance between two points x_i and x_j such that ($i \neq j$) in the n-dimensional space be denoted by d_{ij} and the distance between two projected points y_i and y_j in the q-dimensional space be denoted by d'_{ij} , then the mapping from the higher dimension space to the lower dimension space is done by minimization of the Sammon stress function defined in equation (2).

$$E = \frac{1}{\lambda} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{(d_{ij} - d'_{ij})^2}{d_{ij}}, \lambda = \sum_{i=1}^{N-1} \sum_{j=i+1}^N d_{ij} \quad (2)$$

For minimization of E, steepest descent procedure is most commonly used in which the new iteration y_{il} at iteration t+1 is given in equation (3).

$$y_{il}(t+1) = y_{il}(t) - \alpha \left[\frac{\partial E(t)}{\partial y_{il}} \right] \quad (3)$$

where y_{il} is the l-co-ordinate of point y_i in the new space and α is a constant which sammon takes to be 0.3 or \$0.4\$. The partial derivatives in (3) are

$$\frac{\partial E(t)}{\partial y_{il}(t)} = -\frac{2}{\lambda} \sum_{k=1, k \neq i}^N \left[\frac{d_{ki} - d'_{ki}}{d_{ki} d'_{ki}} \right] (y_{il} - y_{ki}) \quad (4)$$

$$\frac{\partial^2 E(t)}{\partial^2 y_{il}(t)} = -\frac{2}{\lambda} \sum_{k=1, k \neq i}^N \frac{1}{d_{ki} d'_{ki}} \left[(d_{ki} - d'_{ki}) - \left[\frac{(y_{il} - y_{ki})^2}{d'_{ki}} \right] \right] \left[1 + \frac{d_{ki} - d'_{ki}}{d_{ki}} \right] (y_{il} - y_{ki}) \quad (5)$$

The graphs in Figure 1 show the visualization of 46 dimensional data before and after implementation of k-means clustering algorithm. Figure 1 top left shows the Sammon mapping of the data using category information supplied to us. The second diagram in that column shows a zoom in projection of the first taking only the central portion. The third diagram in that column shows a further zoom in. Ideally we would like to see groups of projected points which are categorised as the same type of merchant as lying close to each other (all the blue *s close to each other, and separate from all the red +s etc.) but this has not happened.

The right hand side of Figure 1 shows projections when we use categorical data from a prior k-means clustering. The projections are a bit different because a single merchant belonging to more than one categories have more than one projections. Here we do not say that a merchant belongs to a particular type but only a particular cluster. We see in the second and third diagrams (after zooming in) groups of points which are very close to one another and separate from other groups of points.

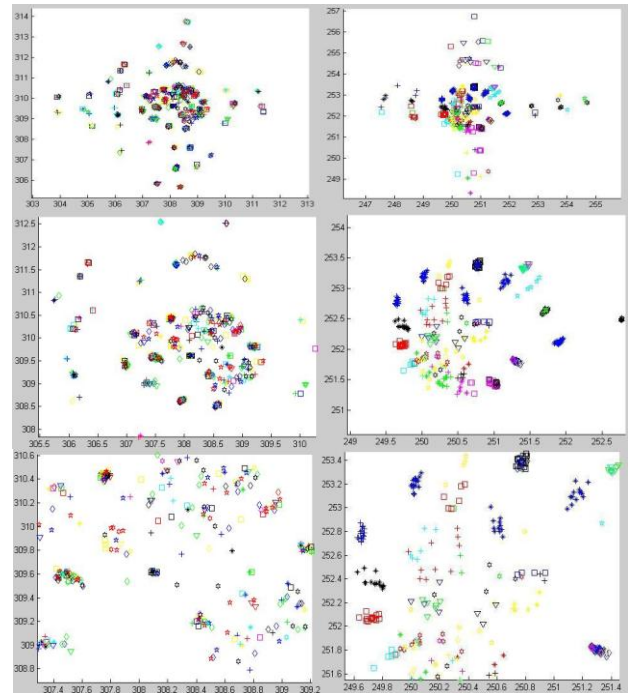


Figure 1. Top left: Visualization using the Sammon Mapping of the data. The class information is given by the pre-defined categories. Top right: the same projection but the class information is taken as the clusters from the k-means algorithm. Middle and bottom diagrams: zooming in on the top diagrams in each column.

Since the data is 46 dimensional and a single merchant is most of the time working in more than one category and also with more than one publisher, the resulting clusters also consist of more than one category of merchants classified in the same cluster. It can be further elaborated as for suppose a single merchant X is working at the same time in "Art & Craft", "Baby Gear" and "Auction" categories and at the same time linked with more than one publishers working in these three categories so this data when classified will give the recommendation of all the publishers working in these three categories to the new merchants working in any one of the above mentioned categories. This is the main reason for the looseness of some of the clusters whereas in some clusters where the merchants are working in one category the data is tightly classified. The samples in cluster 1 are 62 related to "Adult", "Travel-holidays", "Insurance" and "Financial & Legal", cluster 2 are 4 related to "Art & Crafts", "Baby Gear", "Books Catalogues and Magazines", cluster 3 are 3 related to "Auctions", cluster 4 is 1 related to "Baby Gear", "Food", "Gifts & Flowers", "Latest Merchants" and "Wedding & Celebrations", cluster 5 are 7 related to "Baby Gear", "Food", "Gifts & Flowers", "Latest Merchant", "Wedding & Celebrations", "Travel & Essentials", "Home & Garden", "Jewellery" and "Books", "Catalogues & Magazines", cluster 6 are 8 related to "Business Services", "Travel-Essentials" and "Travel-Holidays", cluster 7 are 20 related to "Clothing &

Accessories -Men's", "Adults", "Clothing & Accessories - Women's", cluster 8 are 26 related to "Adult", "Clothing & Accessories - Women's", "Latest Merchants", "Gifts & Flowers", "Wedding & Celebrations", "Travel-Essentials" and "Baby Gear", cluster 9 are 14 related to "Competitions", "Freebies & Discounts" and "Gaming & Gambling", cluster 10 are 10 related to "Computers", cluster 11 are 8 related to "Dating", cluster 12 are 3 related to "DVDs, Videos & Games", cluster 13 are 5 related to "Eco-Friendly", "Education" and "Experience", cluster 14 are 7 related to "Education", "Baby Gear" and "Travel-Holidays", cluster 15 are 9 related to "Experience", "Travel-Essentials" and "Travel-Holidays", cluster 16 are 3 related to "Financial & Legal", "Insurance", "Latest Merchants", "Motoring" and "Travel-Essentials", cluster 17 are 26 related to "Food", "Gift & Flowers", and "Gadgets", cluster 18 are 9 related to "Clothing & Accessories-Women's", "Gadgets", "Gaming", "Mobile Phones & Accessories" and "Jewellery", cluster 19 are 9 related to "Gaming", "Competition Freebies & Discounts", "Gaming" and "Gaming & Gambling", cluster 20 are 30 related to "Games & Gambling" and "Latest Merchants", cluster 21 are 0, cluster 22 are 32 related to "Health & Beauty" and "Adult", cluster 23 are 42 related to "Home & Garden", "Baby Gear" and "Latest Merchants", cluster 24 are 0, cluster 25 are 2 related to "Internet Services", cluster 26 are 2 related to "Jewellery", cluster 27 are 9 related to "Health & Beauty", "Latest Merchant", "Health & Beauty", "Travel & Essentials", "Latest Merchants" and "Motoring", cluster 28 are 10 related to "Loans", cluster 29 are 18 related to "Mobile Phones & Accessories" and "Travel-Essentials", cluster 30 are 0, cluster 31 are 4 related to "Latest Merchants" and "Music", cluster 32 are 2 related to "Latest Merchants" and "office Equipment", cluster 33 are 1 related to "Home & Garden", "Outdoor Equipment" and "Sports & Fitness", cluster 34 are 3 related to "Pets", cluster 35 are 1 related to "Posters & Memorabilia", cluster 36 are 1 related to "Gift & Flowers" and "Seasonal", cluster 37 are 15 related to "Sports & Fitness", "Gifts & Flowers" and "Latest Merchants", cluster 38 are 2 related to "Telecommunications", cluster 39 are 1 related to "Telecoms", cluster 40 are 16 related to "Motoring" and "Mobile Phones & Accessories", cluster 41 are 62 related to "Travel-Accommodation", "Travel-Holidays", "Travel-Essentials", "Travel-Flights", cluster 42 are 0, cluster 43 are 6 related to "Travel-Flights", "Insurance" and "Travel-Holiday", cluster 44 are 0, cluster 45 are 3 related to "Wedding & Celebrations", "Jewellery", "Clothing & Accessories-Men's", "Clothing & Accessories-Women's" and "Jewellery".

In this paper clustering is being done by using the robust method of k-means experimented on a real-time data provided by "Affiliate Future" and further recommendation process is functioned by an enhanced recommendation algorithm given below.

VI.PRIORITY RECOMMENDATION SYSTEM FOR AN AFFILIATE NETWORK

Function Priority-Recommendation ()

Take N Publishers (P₁, P₂, P₃,....., P_N) as the total number of publishers

Take K Advertisers (A₁, A₂, A₃,....., A_k) as the total number of Advertisers

Step 1

Retrieve the total Number of publishers linked with the input vector consisting of more than one advertiser

Step 2

For each input of vector Advertiser A_k where k ∈ {1, 2, 3,,L} where L<=K

Calculate the Number of Advertisers linked with each publisher in an array linked[J] where J ∈ {1, 2,3,.....N}
End For

/* Sorting the Linked Array using Bubble Sort Procedure from Higher to Lower */

Step 3

```
Repeat
    swapped = false
    For i = 1 to length (linked) - 1 inclusive
do:
    /* if this pair is out of order */
    if linked[i-1] > linked[i] then
        /* swap them and remember something changed */
        swap(linked[i-1],linked[i])
        swapped = true
    end if
end for
until not swapped
```

In this way when the new publisher comes to select an advertiser, the advertiser currently working with higher number of publishers will be ranked the highest and so on. The same procedure will be applied for the advertisers as well looking to work with the most appropriate and trust worthy publishers. The Figure 2 shown below portrays the overall working of the proposed idea of clustering on an affiliate network data and priority recommendation process.

VII.CONCLUSION

In this paper, a recommendation system using k-means clustering is introduced in the new domain of affiliate networks. The experiments demonstrated that the recommendation system proposed using k-means clustering and priority recommendation algorithm will play a significant role in selecting the best suitable

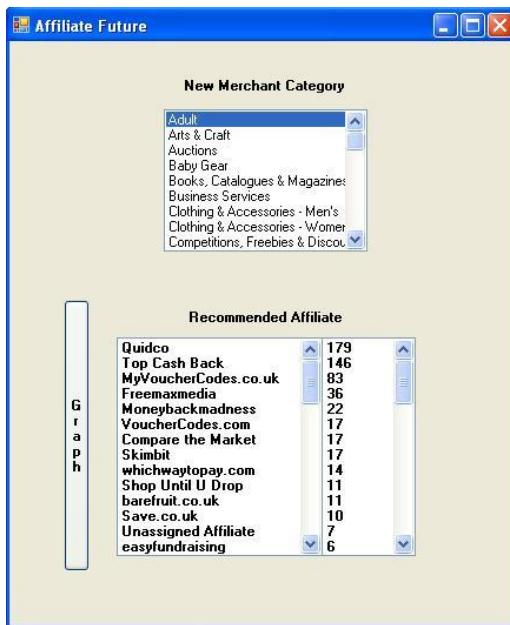


Figure 2. Snapshot of the Experiment

candidate in both the cases of publisher as well as advertisers.

ACKNOWLEDGMENT

The authors would like to gratefully acknowledge the careful reviewing of an earlier version of this paper which has greatly improved the paper

REFERENCES

- [1] B., C., Brown, "The Complete Guide to Affiliate Marketing on the Web: How to Use and Profit from Affiliate Marketing Programs," *Atlantic Publishing Company*, 2009.
- [2] D., L., Duffy, "Affiliate Marketing and its impact on e-commerce" *Journal of Consumer Marketing*, vol. 22(3), pp. 161-163, 2005.
- [3] S., Bandyopadhyay, J., Wolfe and R., Kini, "A Critical Review of Online Affiliate Models" *Journal of Academy of Business and Economics*, vol. 9(4), pp. 1, 2009.
- [4] F., M., Del and M., Paul, "Reevaluating Affiliate Marketing" *Journal of Academy of Business and Economics*, vol. 9(4), pp. 1, 2009.
- [5] Clare, G., "Affiliate Marketing [Electronic Version]" *New Media Age*, vol. 11, 2006.
- [6] S., Goldschmidt, S., Junghagen and U., Harris "Strategic Affiliate Marketing" *Edward Elgar Publishing*, 2003.
- [7] M., Haig, "The e-marketing handbook: An indispensable guide to marketing your product and services on the internet" *Kogan Page Limited*, 2001.
- [8] D., Tweney, "Affiliate Marketing: the future of e-commerce or another hard sell" *Info World Electronic*, 1999.
- [9] R., Xu and W., D., II "Survey of Clustering Algorithm" *Neural Networks, IEEE Transactions*, vol. 16(3), pp. 645-678, 2005.
- [10] P., Girish and D., W., Stewart, "Cluster Analysis in Marketing Research: Review and Suggestions for Application", *Journal of Marketing Research*, pp. 134-148, 1983.
- [11] C., Elkan, "Clustering with K-means: faster, smarter, cheaper", *SIAM International Conference on Data Mining*, 2004.
- [12] J., Macqueen, "Some methods for classification and analysis of multivariate observations", *Proceedings of the fifth Symposium on Math, Statistics and Probability*, 1967.
- [13] J., Dysart, "Click Through Customers", *Bank Marketing*, 2002
- [14] L., Forex, "Affiliate Marketing Makes Headway", *Upside*, vol. 12(4), pp. 176, 2000.
- [15] S., Oberndorf, "Get yourself affiliated", *Catalog Age*, vol. 16(9), pp. 63-64, 1999.
- [16] B., Libai, E., Biyalogorsky and E., Gerstner, "Setting referral fees in affiliate marketing", *Journal of Service Research*, vol. 5(4), 303-315, 2003.
- [17] A. Dhesikan, "Exploring Internet Marketing: A Whole New World of Opportunity", *Bachelor Project Report Worcester Polytechnic Institute*, 2012.
- [18] S. Collins, "History of Affiliate Marketing", *Successful Affiliate Marketing for Merchants*, 2000.
- [19] M., Anastasia, D., Roberto and B., David, "Unintended consequences in the evolution of affiliate marketing networks: a complexity approach", *The Service Industries Journal*, vol. 30(10), pp. 1707-1722, 2010.
- [20] M., D., Paula, "The Web Generation Calls for Extra Image Vigilance", *Bank Technology News*, vol. 17(10), pp. 40-42, 2004
- [21] G., Helmstetter and P., Metivier, "Affiliate Selling: Building Revenue on the Web", *John Wiley & Sons, Inc.*, 2000.
- [22] C., Dale, "The Competitive Network of Tourism emediaries: New strategies new advantages", *Journal of Vacation Marketing*, vol. 9(2), pp. 109-118, 2003.
- [23] J., Wallington and D., Redfeam, "IAB Affiliate Marketing Handbook", *Internet Advertising Bureau*, 2007.
- [24] C., Dorobantescu, "6 Tips for Affiliate Managers", *Avantgate*, 2008.
- [25] R., Leslie, "Clustering for data mining: A data recovery approach", *Psychometrika*, vol. 72(1), 2007.
- [26] A., Jain and R., C., Dubes, "Algorithm for clustering data", *Prentice-Hall, Inc.*, 1988.
- [27] L., Kaufman and P., J., Rousseeuw, "Finding groups in data: an introduction to cluster analysis", vol. 344, 2009.
- [28] R., T., Ng and J., Han, "Efficient and Effective Clustering Methods for Spatial Data Mining", *Proceedings of the 20th VLDB Conference*, pp. 144-155, 1994.
- [29] M., Ester, H., Kriegel and X., Xu, "Knowledge Discovery in Large Spatial Database: Focusing Techniques for Efficient Class Identification", *Proc. 4th Int. Symp. On Large Spatial Databases*, pp. 144-155, 1995.
- [30] Dan, J., P.,K., McKinley and A.,K., Jain, "Large-scale parallel data clustering", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 871-876, 1998. [33]
- [31] M., Ester, H., Kriegel, J., Sander and X., Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise", *AAAI Press*, pp. 226-231, 1996.
- [32] J., A., Garcia, J., A., Garcaposa, J., Fdez-valdivia, F., J., Cortijo and R., Molina, "A Dynamic Approach for Clustering Data", *Signal Processing*, 1994.

- [33] E., Schikuta, "Grid-Clustering: An efficient hierarchical Clustering method for very large data sets", *Pattern Recognition, IEEE Proceedings of the 13th International Conference on*, vol. 2, pp. 101-105, 1996.
- [34] Z., Tian, R., Raghu and L., Miron, BIRCH: An efficient data clustering method for very large databases, *Proceedings of the 1996 ACM SIGMOD international conference on Management of data*, 1996.
- [35] S., L., Brown and K., M., Eisenhardt, "The art of continuous change: linking complexity theory and timepaced evolution in relentlessly shifting organizations", *Administrative Science Quarterly*, pp. 1-34, 1997.
- [36] D., Samosseiko, "The Partnerka – What is it, and Why should you care?", *Virus Bulletin Conference*, 2009.
- [37] B., G., Mirkin, "Clustering for Data Mining, A Data Recovery Approach", *Chapman & Hall/CRC Taylor & Francis Group*, 2005.
- [38] A., Enright, J., V., Dongen, S. and A. Christos, "An efficient algorithm for large-scale detection of protein families." *Nucleic acids research*, vol. 30(7), pp. 1575-1584, 2002.
- [39] D., de Ridder *, R., P., W., Duin, "Sammon's Mapping using Neural Networks: A Comparison", *Pattern Recognition*, vol. 18, pp. 1307-1316, 1997.
- [40] J., W., Sammon, JR., "A Non-Linear Mapping for Data Structures Analysis", *IEEE Transactions on Computers*, vol. 18(5), 1996.



Colin Fyfe is a Personal Professor at The University of the West of Scotland. He has published more than 350 refereed papers and been Director of Studies for 23 PhDs. He is on the Editorial Boards of 6 international journals and has been Visiting Professor at universities in Hong Kong, China, Australia, Spain, South Korea and USA.



Malcolm Crowe obtained his D.Phil (Oxon) in Mathematics in 1978 and was appointed a Professor of Computing in 1985. He maintains a number of research interests including Enterprise Computing and Business Intelligence, and is Director of Studies or second supervisor to over a dozen PhD students at the University of the West of Scotland.



Zeeshan Khawar Malik is currently a PhD Candidate at the University of The West of Scotland. He received his MS and BSCS (honors) degree from University of The Central Punjab, Lahore Pakistan, in 2003 and 2006, respectively. By profession he is an Assistant Professor in University of The Punjab, Lahore Pakistan currently on

Leave for his PhD studies.