Behavioural Analytics using Process Mining in On-line Advertising

Maria Diapouli[1], Stelios Kapetanakis[1], Miltos Petridis[2] and Roger Evans[1]

[1]School of Computing, Engineering and Mathematical Sciences, University of Brighton, Watts Building, Lewes Road, Brighton, BN2 4GJ, UK
[2]Department of Computing, University of Middlesex
[1]{M.Diapouli, S.Kapetanakis, R.P.Evans }@brighton.ac.uk, [2]M.Petridis@mdx.ac.ukuk

Abstract. Online behavioural targeting is one of the most popular business strategies on the display advertising today. It is based primarily on analysing web user behavioural data with the usage of machine learning techniques with the aim to optimise web advertising. Being able to identify "unknown" and "first time seen" customers is of high importance in online advertising since a successful guess could identify "possible prospects" who would be more likely to purchase an advertisement's product. By identifying prospective customers, online advertisers may be able to optimise campaign performance, maximise their revenue as well as deliver advertisements tailored to a variety of user interests. This work presents a hybrid approach benchmarking machine-learning algorithms and attribute pre-processing techniques in the context of behavioural targeting in process oriented environments. The performance of our suggested methodology is evaluated using the key performance metric in online advertising which is the predicted conversion rate. Our experimental results indicate that the presented process mining framework can significantly identify prospect customers in most cases. Our results seem promising, indicating that there is a need for further workflow research in online display advertising.

Keywords: Process mining, Process Oriented Workflows, Classification, Online Display Advertising.

1 Introduction

According to statistics published by the Internet Advertising Bureau online advertisers in the UK have spent more than 8.6 billion UK-pounds in 2016 on behavioural targeted advertising a figure which grew 16.4% compared to 2015. The estimate represents steady growth rates of about 20% from 2010 through 2016 [3]. Behavioural targeting and customer prospecting are both promising and challenging aspects in display advertising. Promising since the more information of user behavioural activity exists the better targeted advertisements could be delivered to end users and challenging since display advertising is a rather complex ecosystem which involves multiple interested parties such as end users, advertisers, publishers, and ad platforms. The size of data generated and collected from any involved parties is significantly large: Billions of websites requests every day trigger millions of advertisements that are finally displayed to millions of users.

Digital advertisers attract increasing traffic on their websites aiming for certain user marketing actions, more commonly, accomplishing an online purchase. This action is recorded as a conversion. There are two ways for viewing an advert upon arrival on an affiliate ad-friendly website. Firstly, by clicking on the advert and immediately buying and/or by viewing an advert and waiting for a future return and a possible purchase. The journey of a user throughout several websites can be represented as a series of events with intermediate temporal durations. This can be interpreted into a "workflow" of variant length which may or may not convert at its final stages. Petridis et al. [19] have shown that workflow behaviours with such a distinct event-duration coupling can be formalised over a general theory of time [20], be graph-represented, monitored [21] and explained [22] effectively using Case-based Reasoning techniques [23].

Our research questions on top of the online marketing business model are twofold – One: which metric features in terms of evaluating an online campaign performance are mostly important and -Two: based on the set of identified metrics what is the profile of an ad viewer who is keen to make a purchase. In such way by analysing and classifying past behavioural observations among ad viewers, could allow marketers to identify future prospect customers more effectively.

The work we present in this paper handles a challenging area in the online display advertising marketplace, this of customer prospecting. Customer prospecting identifies web users who are likely to purchase a product after seeing an advertisement. We developed a process mining methodology based on an advertising campaign implemented by an ad network provider. We collected and analysed campaign data that contained audience demographic information and audience behavioural segments to predict whether a user who had no previous seen an advert is likely to convert. The goal of this research was to increase an individual advertising campaign performance by augmenting its CPA ratio.

This paper is structured as follows: Section 2 presents the context of search engine advertising and its online display landscape, section 3 will describe our adopted process mining methodology. Additionally, the imbalanced problem of conversion rate will be explained and our approach to the class imbalance problem will be analysed. Section 4 will present a series of empirical experiments for selecting the best performing classification algorithm. Finally, a discussion upon our experiment results will be presented in section 5.

2 Related Work

2.1 Search Engine Advertising

The application of statistical algorithms and process mining methodologies is widely applied in search engine advertising. Its outcomes could be observed from user-relevant textual advertisements placed next to search results as they come from several search engines. Choosing the most relevant ad for a user query and the optimal place in which it is displayed could affect significantly the probability for a user to click on that chosen ad.

For any adverts with already known Click-Through Rate (CTR) historical information, CTR could be estimated empirically by dividing the number of impressions over clicks. In any other case when a new ad, with no historical information, is going to be displayed to a random user; a major challenge emerges: This is mainly in terms of identifying a suitable advert where a user would be "tempted to click". Richardson et al. [9] answer this challenge by predicting the CTR for a new ad with no prior historical information. Based on ad text information only (title, body, search keywords, display URL, impressions, clicks, landing URL) logistic regression can produce an accurate model ad CTR prediction. Research in the area has shown that decision rules [7] can be produced for predicting the CTR for unseen ads, from data that contain information regarding advertisements, query terms and URLs. Clustering techniques could also be used to improve the keywords CTR for rare or new keywords [8] based on generation of clusters of related keywords. This can be applied by if different search keywords have a different likelihood of receiving a user click [8].

2.2 Online display advertising landscape

Online Display Advertising is a highly congested and convoluted environment involving an extended range of vendors, services and high volumes of transactions. Its landscape mainly comprises workflows of advertisers, ad agencies, web-users and publishers. Advertisers set up product or service workflows in publisher web sites, also known as inventories, with the aim to attract as many web-users as possible. This is achieved by providing rich and engaging ad-context to the most receptive online audience. In return the end users will click on the ad and will be redirected on the advertiser website to purchase potential product(s).

The complexity of achieving such desired actions has led to the development of new display parties, these of Ad-exchanges and Ad-networks. Both could perform better on accessing and controlling inventories. Ad-exchanges are online auction marketplaces (like e-bay) trading in advertisements as their "products". Ad-exchanges could provide three main services: adverts allocation, prices determination, traffic control [1]. Ad networks provide a variety of excluded services to advertisers such as ad serving, privacy verification, targeting the most suitable audience and advertising campaign reporting. Ad networks could access publishers directly as well as ad exchanges for identifying inventories.

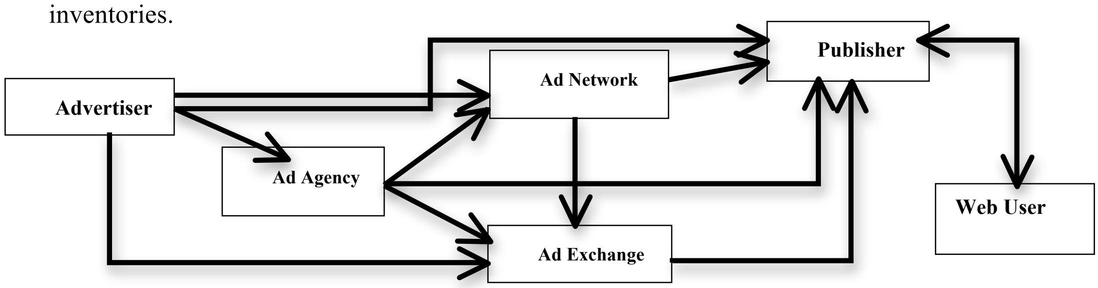


Figure 1: Key stakeholders of an online-advertising workflow

Measuring the revenue and the effectiveness of online display advertising campaigns is achieved through three prevalent pricing models. These are: Cost per impression (CPM), Cost per click (CPC) and Cost per action (CPA). CPM ratio is popularly used for brand recognition campaigns where a fixed cost is charged to advertisers based on the number of displays of advertisements. CPC ratio was introduced to build the advertisers confidence upon their return on investment(s). Advertisers pay publishers when a web user clicks on an advertisement without considering the number of impression displays. However, since most advertisers are retailers the actual advertising benefit derives from the commercial transaction within their websites [2]. CPA metrics is used to serve this purpose. According to the Interactive Advertising Bureau report (2016) in the US market, 65% of online advertising transactions share was attributed on the customer performance models (CPC, CPA), while the second on the list was the CPM model with 33%. Hybrids of impression and performance models reside in 2% of online advertising transactions [3].

According to Lewis and Reiley (2014) [16] the effect of online advertising on sales is not fully associated with CTR. Their collaboration with Yahoo!Research and an eminent retailer had reported that 78% of the lift in retailer sales was originated from users who had viewed ads but had not clicked them, while only 22% was attributed to those who had clicked. In our research work we found that the online advertising campaign had substantial impact on the users who merely viewed the ads. Based on thorough analysis we identified that impressions are more strongly correlated to conversions than clicks. Most interestingly, clicks had a very trivial correlation (correlation = 0.00000115) with conversion. These findings suggest that the most meaningful metric for evaluating campaign performance is conversions instead of clicks.

3 Methodology

Customer prospecting is related with predictive modelling in process mining terms. Predictive modelling, also referred as supervised data mining, aims to predict the probable future event based on previous historical knowledge [10]. The appropriate selection of data samples is important for effective analysis and prediction based on underlying patterns [13]. In this work decision trees and kNN were preferred over the commonly used logistic regression and collaborative filtering classification methods. Decision trees have been shown as effective in building profiles for the web users who have converted in the past and then predict whether a new web-user is likely to convert [17]. Decision trees although powerful in expressing continuous and categorical inputs, they seem to fall out when there is a mixture of continuation and categorical type data. Thus, kNN has seemed more appropriate since it performs better with continuous data [18]. For predictive modelling decision trees seem most powerful and prevalent tools [15] compared to logistic regression and collaborative filtering since:

-Logistic regression could be used to examine the model's exponentials of the coefficients to explore which user attributes affect the likelihood of conversion. In such way, we would be able to explore the necessary coefficients but be unable to explore the underlying ruleset which could indicate and predict online “target” users willing to convert.

-Collaborative Filtering was also considered since it has been proven effective in finding prospect customers based on past customer behaviours (training samples) [14]. In such way, a successful model would be retrained on a regular basis to include recent user activity information. However, in our investigated dataset such information was not available and this research was not able benefit from “live ad-feeds” including: Ad ids, ad-width, ad-height, visibility time viewed, format, etc.

Therefore, based on the limitations of the dataset and the model target audience, our selection methodology was based on data-mined patterns for ruleset generation to understand and predict successful (or not) online user-conversions.

3.1 The Data

The dataset used in this work was retrieved from real advertising campaigns as conducted by a UK-based advertising company. The company was using a No-SQL distributed database management system (DDMS) based on Apache Hadoop. Any marketing data was extracted from DDMS and stored in tab separated files (tsv) textual format for further processing and analysis. During a one-week campaign 20 million impressions were displayed to web users approximately, a figure which was increased exponentially over more campaigns and longer campaign times or series of campaigns.

The used dataset comprised three distinct types of ad-logs, described as: impressions, clicks, and conversions. Any available log data were organised and aggregated based on the user id feature on a row-per-record basis. A sample of the feature description and example for each column of the ad log data are presented in Table 1. Each record contained three information types: (i) Behavioural data (columns 1,2,6,7,8,9) (ii) Interest profiles (columns 1,2,4,6,7,8,9,10,11,12) and (iii) Intent profiles (columns 1,2,3,6,7,8,9).

Column	Description	Example
1	User id	156780d5128b4c1cb1bc5652ebcadd2d
2	Location data	UK, London
3	Interest profiles	shopping
4	Intent profiles	home and garden
5	Previous browsing activity	https://www.gumtree.com/p/dinnerware-crockery/tea-set/1183263978
6	Device	tablet
7	Browser	mobile safari
8	Operating system	android2
9	Language	English
10	Impressions	19
11	Clicks	1
12	Conversions	1

Table 1: Sample of a data ad-log format.

The dataset used in this experiment was gathered as a day-campaign and it consisted of 3,425,119 impressions that were displayed on 3,407,293 users. Among them 8,082 users clicked on the displayed advert (click response rate 0.24%) and 913 converted (convert response rate 0.03%). Due to the very sparse number of response rate both for click and conversion the data were highly skewed. To overcome this limitation of imbalanced data a sampling technique was adopted and will be discussed in the next section. The size of the data set had 3,407,293 observations. Each observation characterised a web user and was described by 46 independent features and 1 dependent feature (sample of log data are shown in Table 1). The independent features were both categorical and nominal. These features were related to a user's browsing history (URLs that user has visited in the past). The dependent feature was a nominal one-named conversion rate which illustrated whether a user had made a purchase in the past.

3.2 Our Approach

The process of Knowledge Discovery in Databases (KDD) was adopted as methodological approach on this case study, where process mining was the gist in the overall process [11]. The experiment went through all the steps of KDD. The data were extracted from the DDMS based on the process mining project specifications to ensure consistency and completeness. In the data transformation phase, categorical values where transformed on numerical ones to adhere to the algorithm specifications. Additionally, since the response rate for conversion was a small number, only 0.03% of the total dataset, a method that modified data values and derived new fields from response rate for conversion was adopted. A new field was created which contained two values for the conversion field: zero (indicated that a user did not convert) and one (indicated a successful user conversion). However, this new field did not overcome the very low percentage of conversion rate. This was regarded a challenge since a high-performance classifier was required to have an accurate model. In such: training data should be evenly distributed between conversion and no-conversion values. In any other case the classifier could be biased since it would try to achieve the overall classification accuracy by identifying mostly the majority class (conversions) and would overlook the minority one (no-conversion). This would offer little contribution to the model accuracy [4]. Our approach in balancing the classifier will be described in section 4.2.

4 Experiments and Evaluation

4.1 Modelling and imbalanced data sets

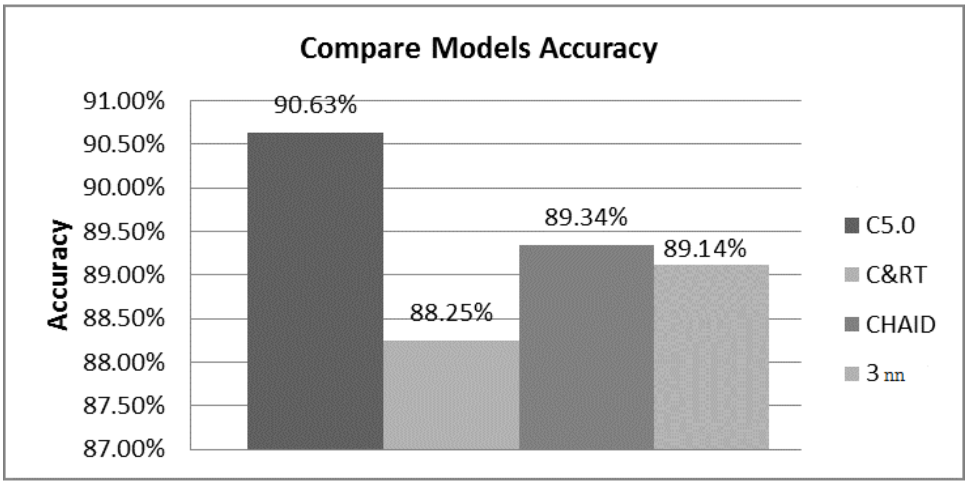
For this work, IBM SPSS Modeler 18.0 was used for all experiments. Different algorithms were assessed to benchmark the most appropriate and accurate dataset for classification and prediction. The data set was separated into training and testing set to build and evaluate our decision tree model with a 70 - 30 split rate respectively. In the training phase, the model was processed by using the training set and then tested to evaluate our model's accuracy

To overcome the problem of heavily imbalanced data two approaches were considered: a) using over-sampling and b) using under-sampling. In the over-sampling approach, the training set was populated with replicated data that belonged to the minority class until the training set was balanced [6]. The information remains the same but the misclassification cost of the minority class is increased.

In the under-sampling approach data from the majority class were removed to balance the training set [5]. For our experiments, the under-sampling approach was used where cases from the dependent feature conversion rate were randomly eliminated.

4.2 Comparing the performance of different algorithms

Different decision tree algorithms have been selected for searching patterns in data as well as kNN with k =3 for more accurate classification of continuous data attributes. This process included deciding which algorithm provided the lowest average classification error. The selected algorithms were: classification and regression tree (C&RT), Chi-squared automatic interaction detector (CHAID), C5.0 and 3NN. The performance results of these algorithms as applied on the test set are shown in Graph 1.



Graph 1: Summary of results

	C&RT	Chaid	C5.0	3NN
Converter True Positives (Hit) Rate	93.36%	96.48%	99.02%	91.23%
Non-Converter True Positives (Hit) Rate	88.24%	89.34%	90.62%	89.32%
Accuracy	88.25%	89.34%	90.63%	89.14%
Sensitivity	93.36%	96.48%	99.02%	90.15%
Specificity	88.24%	89.34%	90.62%	88.14%

Table 2: Accuracy Measures and Alternatives

Table 2 illustrates a higher likelihood (90.63%) for C5.0 to predict the event for someone converting on an advertisement compared to the other baselines (88.25% for C&RT, 89.34% for CHAID and 89.14% for 3NN). In Table 2 C5.0 is showing 99.02% sensitivity (the portion of users that were correctly predicted to convert) and 90.62% specificity (the portion of users that did not convert and were successfully predicted) which accounts for the overall accuracy of 90.63%. The sensitivity and specificity measures are used to ascertain the model validity and accuracy [12].

4.3 Comparing the performance of the different data sets

Typically, the performance of machine learning algorithms is evaluated using predictive accuracy. The evaluation and interpretation of the mined patterns in terms of reliability and accuracy of the derived rules have taken place in the evaluation phase.

We performed our experiments using 10-fold cross validation. The original dataset was randomly divided into ten (10) subsets. Each time, one of the 10 subsets was used as the test set and the other 9 subsets were combined to form the training set. For the conversion field, there were two classes, the positive class, assigned as 1, that comprised "converted" users and the negative class consisted of "no converters", assigned as 0.

Partition	Training Set	Percentage Training Set	Testing Set	Percentage Testing Set
Correct	1,482	93.09%	3,088,032	90.67%
Wrong	110	6.91%	317,669	9.33%
Total	1,592		3,405,701	

Table 3: Model Accuracy

Table 3 illustrates the correct and wrong prediction of our model.. The rows defined by actual values and columns defined by predicted values, with the number of records having that pattern in each cell.

4.4 Evaluating the model performance with bootstrap aggregation

The data for building decision trees with C5.0 algorithm models were re-sampled using a bootstrap aggregation technique to form several pairs of training and testing data sets. A decision tree model was developed for each pair of training and testing data sets. Predictions from any individual decision trees were merged via a voting system which led to the highest possible accuracy for the final model (ensemble) predictions. It was observed that similar models were generated throughout ensemble learning. This was evidence that the chosen algorithm was stable throughout the dataset.

5 Conclusions

In this paper, we demonstrated process mining as an effective tool for direct marketing which can improve online marketing campaigns. Most the existing research in this area so far focuses on computational and theoretical aspects of direct marketing though little efforts have been put on technological aspects of applying process mining in the process of direct marketing. The complexity of process mining models makes it difficult for marketers to use it, hence we outlined a simplified framework to guide marketers and managers in making use of process mining methods and focus their advertising and promotion on those categories of people to reduce time and costs. We explained the steps and tasks that are carried out at each stage of the process mining framework and showed some examples of the type of predictive efficiency that can be achieved using the proposed approach. This has shown that substantial gains can be achieved by adopting this pragmatic and exploratory approach to predict user behaviour in on-line advertising.

This work has shown capable of dealing with the uncertainty underlying within behavioural data as on-line advertising experts have noted that user behaviours can vary significantly. Our suggested approach seems capable of dealing with more complex online advertising models and thus our future directions will focus on more complex, variant and fuzzy attributes.

The results obtained so far, are promising and encourage us to continue experimentation with more sophisticated models or other algorithms to further improve the performance of the system. It seems sensible to experiment with the following settings in future work:

– introducing the temporal dimension to our model to apply time series analysis techniques to build the model
– combining the model with content-based approach
– additional category-based and continuous-based attributes specifying the times spent on each of the categories, with possible division into work-days and week-days, for example a different choice of categories.

As future work, we will incorporate more user and publisher information obtained from third party media providers into data hierarchies to improve model prediction.

Acknowledgement

This work was funded by Innovate UK (formerly known as Technology Strategy Board).

References

1. Balseiro, S., Feldman, J., Mirrokni, V., Muthukrishnan. S., Yield optimization of display advertising with ad exchange. In ACM Conf on E-Commerce, (2011)
2. Moona, Y., Kwonb, C.: Online advertisement service pricing and an option contract, Electronic Commerce Research and Applications, Elsevier (2011)
3. Interactive Advertising Board (IAB). Internet Advertising Revenue commissioned report, (Updated June 2016), https://iabuk.net/, accessed May 2017
4. Visa, S., Ralescu. A., Issues in mining imbalanced data sets - a review paper. Midwest AI and Cognitive Science Conference, 67-73 (2005)
5. Kubat, M., Matwin, S.: Addressing the curse of imbalanced training sets: one-sided selection, International Conference on Machine Learning, 179–186 (1997)
6. Ling, C., Li, C.: Data mining for Direct Marketing: Problems and Solutions Proc. Of the Fourth Inter. Conf. on Knowledge Discovery and Data mining, AAAI Press, 73-79, (1998)
7. Kotlowski, W., Debmbsczynski, K., Weiss, D.: Predicting ads click through rate with decision rules. In Workshop on Target and Ranking for Online Advertising, WWW 08, (2008)
8. Regelson, M., Fain D. C.: Predicting click-through rate using keyword clusters. In Electronic Commerce (EC). ACM, 2006.

9. Richardson, M., Dominowska, E., Ragno, R.: Predicting clicks: estimating the click-through rate for new ads. International conference on World Wide Web, 521–530, (2007).
10. Apte, C. V., Hong, S. J., Natarajan, R., Pednault, E. P. D., Tipu, F. A., Weiss, S. M.: Data-intensive analytics for predictive modelling, IBM journal of research and development , 47(1) (2003)
11. Chapman, P., Clinton, J., Kerber, R., Khabaza T., Reinartz, T., Shearer, C., Wirth, R., 1999. CRISP-DM 1.0 Step-by-step data mining guide. Crisp DM Consortium, (Updated 2010), https://www.the-modeling-agency.com/crisp-dm.pdf, accessed April 2016
12. Han, J., Kamber, M., Kaufmann, M.: Data mining: Concepts and Techniques (2nd Edition) (2006)
13. Kim, Y.: Toward a Successful CRM: Variable Selection, Sampling, and Ensemble, Decision Support Systems, 41(2), 542-553 (2006).
14. Anastasakos, T., Hillard, D., Kshetramade, S., Raghavan, H.: A collaborative filtering approach to ad recommendation using the query-ad click graph, ACM conference on Information and knowledge management, 1927-1930 (2009).
15. Reddy, C.S., Vasu, V., Kumara Swamy Achari B.: Effective Decision Tree Learning, International Journal of Computer Applications, 82(9), (2013).
16. Lewis, R.A., Reiley, D.H.: Online ads and offline sales: measuring the effect of retail advertising via a controlled experiment on Yahoo!, Quantitative Marketing and Economics, 12(3), 235-266 (2014)
17. Diapouli, M., Petridis, M., Evans, R., Kapetanakis, S.: An Industrial Application of Data Mining Techniques to Enhance the Effectiveness of On-line Advertising. In Proceeding of Thirty-sixth SGAI International Conference on Artificial Intelligence, AI 2016, pp.395- 401 (2016)
18. Shah, Z., Mahmood, A. N., Orgun, M. A., Mashinchi, M. H.: Subset Selection Classifier (SSC): A Training Set Reduction Method. In proceedings of the 16th International Conference in Computational Science and Engineering (CSE), 2013 IEEE, pp. 862-869 (2013)
19. Petridis, M., Ma, J., Knight, B.: Temporal model for business process. Intelligent Decision Technologies. 5 (4), pp. 321 – 332 (2011)
20. Ma, J., Knight, B.: A General Temporal Theory, the Computer Journal, 37(2), 114-123 (1994).
21. Kapetanakis, S., Petridis, M.,Knight, B.,Ma, J.,Bacon,L. : A Case Based Reasoning Approach for the Monitoring of Business Workflows, 18th International Conference on Case-Based Reasoning, ICCBR 2010, Alessandria, Italy, LNAI (2010)
22. Kapetanakis, S., Petridis, M., Knight, B., Ma, J.,Bacon, L.: Providing Explanations for the Intelligent Monitoring of Business Workflows Using Case-Based Reasoning, in workshop proceedings of ExACT-10 at ECAI 2010, Lisbon, Portugal (2010)
23. Kapetanakis, S., Petridis, M.: Evaluating a Case-Based Reasoning Architecture for the Intelligent Monitoring of Business Workflows, in Successful Case-based Reasoning Applications-2, S. Montani and L.C. Jain, Editors. 2014, Springer Berlin Heidelberg. p. 43-54.