Click prediction using unsupervised learning methods

Vitalija Serapinaite [1], Ignas Suklinskas and Ingrida Lagzdinyte-Budnike[1]

[1] *Kaunas University of Technology, Studentu 50, 51368 Kaunas, Lithuania*

Abstract

Contextual targeting offers a non-privacy-intrusive way to target audiences without the usage of third-party cookies. The idea behind contextual targeting is that when ads are displayed on websites of positively related context, the probability of the user interacting positively with the ad increases. Click-through rate (CTR) has low occurrence between 0.5 and 2 % creating challenges to classify raw advertising data. Machine learning algorithms such as XGBoost are used for CTR prediction but deep learning methods are gaining attention due to better performance. The models reach good classification results, however, they are still based on user historical data. In this paper, unsupervised learning methods such as the isolation forest and the local outlier factor are used as models to predict whether raw contextual data will result in clicks or not. The models learn underlying patterns of the click samples, therefore impression class data seems like an outlier or novelty. The results of the study showed that the best-performing isolation forest algorithm achieved 43% accuracy, which was worse than the baseline of the random classifier. This allows us to conclude that the information described by contextual attributes alone is not sufficient for the solution of such task, but combining it with historical data that is not sensitive in terms of security would probably give a better result. The study also showed that the isolation forest algorithm performs better on lower dimension data than the local outlier factor algorithm. Meanwhile, the effectiveness of the latter one is more related to the quality of the data than its dimensions.

Keywords

Contextual targeting, Click prediction, Machine learning

1. Introduction

Digital marketing domain amounted to 14 billion pounds in the UK in 2019 year showing that the domain became an important part of companies and is lucrative [1]. Digital advertising is an important part of any company that tries to gain better client reach, spread brand awareness, or get more revenue. It includes displaying various kinds of ads on websites, videos, or mobile apps. The goal of digital advertising is to display ads to users that can react positively to them. Positive interaction with the ad can vary from the purchase of an item to an ad impression (view). Targeting relevant groups can maximize the return from advertising, therefore different advertising tactics are used. One of these tactics is based on classification of the Internet users into wider groups that share similar interests, location, education, or age, analyzing users' browsing history or information shared publicly [2]. These segments help advertisers to reach certain audiences and target more potential customers. Another targeting strategy that is successfully used to this day for digital marketing uses browser cookies. Cookies are small pieces of information that are sent to the browser from the websites during users' visit. All user behavior is tracked by cookies such as information about products saved in a cart, time spent viewing each paragraph or an ad and etc. Third-party cookies track user behavior across the Web and allow to use behavioral advertising which is based on historical user data. Therefore, it can have access to sensitive information such as political views, medical history, or sexual orientation [3]. Due to these reasons the users suffer from a lack of privacy, security issues, and data ownership control [3, 4]. In addition, though users can choose whether to allow websites to track their activity, 75 % of

tracking activities can happen before the user was given the choice, therefore users' choices have little impact on actual tracking [5]. More laws are created to regulate how much data can be tracked (EU GDPR) [6]. Third-party cookie removal on Chrome is planned by 2024 and Safari is already blocking the cookies by default. Their removal can cost up to 10 billion in US publisher revenue based on an IAB study [7], therefore alternative methods can be used to improve advertising. Contextual targeting can be an alternative to third-party cookie usage since the method relies on advertisement contexts rather than user history. Based on Connatix and Digiday's survey of more than 100 publishers [8], 23 % of publisher respondents admitted to using machine learning or artificial intelligence methods in advertisement indicating that the algorithms are effective and the percentage of companies using AI will raise. Contextual targeting enables to show ads to the users based on the website content and ad's similarity to it without using any information about the user's browsing history. Due to these reasons, the need of effective context-based targeting methods is increasing notably. The goal of this study is to examine whether unsupervised learning methods can learn to predict clicks using contextual data. Contextual data features are enhanced using GloVe representations. The structure of this work consists of analysis of methods used for contextual advertising (chapter 1), methodology (chapter 2), descriptions of data and preprocessing methods (chapter 3 and 4), result analysis and conclusions (chapter 5 and 6).

2. Analysis of methods that focus on contextual advertising

Different analyses show that not only the content but also the context can affect user behavior. For example, the study [9] focused on investigating different ratios of displayed relevant and irrelevant ads on paginated websites found that when the number of relevant ads is high, the memory of such ads reduces. Another study analyzed contextual cues such as access platform, and gaming device [10]. It was found that the cues affect brand attitude and memory differently: the highest brand memory is achieved when the access platform is a social network, and the gaming device is a PC which is more sensory-rich than other devices. In addition, banner placement positions impact user recall of banners [11]. Static banners placed on the top-right position of the website lead to better recall and are looked at for longer. Another study [12] investigated the effects of online banner content on users' visual attention. The researchers of the study found that banner middle areas are noticed first, higher discount rates in banners are more noticeable. Users that are unfamiliar with the brand, spend more time looking at the image rather than the discount area, opposite to users that are familiar with the brand.

Contextual advertising is the practice of optimizing audience targeting effectiveness by placing ads in media with favorable or similar contexts. Common goals are to improve brand awareness, increase the click-through rate of an ad, or create an item purchase intent [13]. Artificial intelligence or machine learning can help optimize contextual targeting practices by predicting the probability of a click, making better recommendations in which web context ad should be placed based on its contextual data. A study [14] focused on predicting the probability of click-through-rate trained a model that used basic features (advertiser ID, name, space ID, space name, category ID) and image features (RGB information), characteristics of an image extracted via ResNet. Basic features were transformed into low-dimensional real number vectors and combined with image features as input to a deep neural network. The model reached 0.0206 root-mean-squared error (RMSE) which was better than the results achieved by 6 other models. Root-mean-squared error measures the performance of the model. Lower values of the metric indicate better model performance. Another study [15] authors created a framework for a better mobile in-app advertising using large Asian countries' in-app advertising data of 30-day period. The observed clickthrough rate (CTR) in data was around 0.90 % indicating that most of the interactions are impressions but not the clicks. App and ad categories, an hour of the day, province, smartphone brand, connectivity type, user mobile provider, and user's internet provider were used as categorical features. Feature functions of the data contain information about impressions, clicks, CTR, distinct ads shown over prespecified history, entropy, app count, time variability of users' CTR, and app variability. The study used the XGBoost algorithm and reached better results on the test data compared to the baseline. However, the increase in CTR is due to collected behavioral information and contextual information adds little value. In another paper [16] researchers proposed the DualMLP model that consists of two-stream multi-layer perceptron (MLP) models as a base algorithm for CTR prediction. Such architecture

increases models' ability to learn various features from each stream that supplement each other. In this study, outputs of MLP models are fused using bilinear interaction aggregation layer. The experiments were performed using Frappe, MovieLens, Criteo, Avazu datasets and measured using AUC metric. FuxiCTR was used as a baseline model. Results were compared with other single stream and two stream models such as HOFM, CrossNet. The DualMLP model reaches competitive results compared to other two-stream models (98.33 vs 97.42 and 95.94 AUC). The results show, that two less complicated network architectures combined can reach good performance.

Another method created for advertisement recommendation [17] used natural language processing (NLP) techniques on microblogs by extracting the top 5 keywords with positive sentiment and matching them to ads using their heading, industry field, bid phrase, title, and URL data. Wikipedia Topics were used as reference points to get better recommendations. Machine learning can also be used to find the best ad insertion point in video material. Multimodal model [18] based on convolutional neural network (CNN) extracting fusion-representation from semantics, scene, sentiment, object, audio, and color information of a video was used to select ads that are shown in a video during predicted timestamp. Tested on 3000 ads with over 6000 movie clips from movieQA, the model reached 2.5 % higher accuracy compared to different methods such as C3D, and TSN I3D models. Contextual targeting can also be improved by using the contextual information-based relation extraction method. Extended Long short-term memory model [19] with gate mechanism, convolutional graph network, entity-centric logical adjacency matrix combined with GLoVe has reached competitive results on TACRED and SEMEVAL-2010 tasks. Website content categories can also be enhanced by using the wikitocat model [20] to predict categories in the content. An algorithm called LiveSense [21] that advertises ads in live streaming videos based on context, uses three principles to determine contextually relevant ads: textual relevance, local and global visual relevance. The textual relevance is calculated by measuring the vector space of the title of the steam of an ad, and the description of the host webpage. Visual relevance is calculated between the current live stream frame and ad frames using 64-dimensional histograms in HSV color space. HSV color space indicates image hue, saturation and value. This data was extracted from the LIVID dataset and was used to train multi-layer perceptron. The trained model reached a lower root mean squared error than other state-of-the-art methods.

Another study [22] researchers created a model called MiNet for CTR prediction that uses auxiliary data of users' interests. Long-term interests are expressed as profile features such as age group, gender, short-term interest like the type of ads the user clicked on before, information of recently clicked news in the website where an ad is displayed. Offline experiments using Amazon Books-Movies dataset and the News-Ads dataset showed that the created network outperformed some of the state-of-the-art methods such as CoNet, and MV-DNN for CTR prediction. Differentially private stochastic gradient descent algorithm [23] can be used on ads data that contain class imbalance and sparse gradient updates to predict CTR, conversion rates, the number of conversion events and evaluate privacy-utility trade-offs on datasets that were taken from the real world. The results show that the method is both beneficial and privacy-inducing for ad-related tasks. Another method called the deep multi-representation model (DeepMR) [24] was proposed for CTR prediction that incorporates deep neural networks and a multi-head self-attention mechanism. In addition, it contains the ReZero method that uses novel residuals with zero initialization connections to a deep neural network for learning better representations. The model outperformed state-of-the-art models for CTR prediction such as xDeepFM, DeepFM on three real-world datasets Frappe, MovieLens, and LastFM. Another effective way to use contextual advertising to increase CTR rate is to use contextual competitive targeting [25] which increases CTR more effectively than contextual targeting but does not result in increased conversion rates. An experiment using Google's AdSense contextual targeting platform showed that advertising banners in competitor websites of similar contexts lead to a CTR increase. However, a follow-up survey revealed that it is caused by users' curiosity and customer loyalty to the brand reduces conversion rates.

Contextual targeting can also be seen as a contextual bandit problem that focuses on finding the best action based on loss, cost, and reward functions given a specific context. The algorithm is best applied to situations with a dynamic environment that has rapidly changing settings making it suitable for contextual targeting. In addition, it uses external side information such as context for making decisions. An enhanced version of the contextual bandit problem, that does not assume a simple reward function or states not affected by previous actions, called Policy Gradients for Contextual Recommendations (PGCR) [26] was used to personalize music recommendations. However, it can also be applied to

personalized advertising. The model has fast convergence and outperforms classic contextual bandits and basic policy gradient methods. Another contextual bandit-based method that uses CNN networks for displayed images learns the reward function and upper confidence bound for exploration [27]. The model outperforms other strong baselines such as LinUCB, KernelUCB on four datasets. Contextual bandit algorithm can also be used to find matches between advertising creatives and target audiences by solving overlap issues using partitioning of target audiences [28]. The algorithm learns to assess which creative has the best fit in the space of potentially overlapping target audiences while simultaneously learning an ideal creative display policy in the disjoint space. The algorithm performs more efficiently than non-adaptive A/B testing or naïve split-testing.

Based on analyzed works it can be concluded that most of the models are based on classification of clicks or prediction of CTR. However, these models require large amounts of data, computation resources due to prolonged training times. All analyzed models require user historical data to make predictions whether user will click an ad or not. In addition, classification methods are based on class separation, therefore big data imbalance that exists in click and impression data impacts classifier's ability to find any patterns in minority class. Another approach to learning minority class patterns is to use unsupervised learning methods on a single class. Unsupervised learning methods that learn to recognize anomalies or find novelties are faster algorithms that require less data to learn data patterns. Therefore, isolation forest and local outlier factor algorithms are chosen in this study and will be trained on context-only data to see if models can capture underlying patterns of click data and distinguish it from impressions. Prediction of a click is important task since it allows to maximize returns from digital marketing by changing advertising strategy in a way that results in more clicks.

3. Methodology

As discussed in the previous section, multiple types of models can be used for CTR prediction ranging from machine learning methods to deep learning. Classification models learn such model weights, that the classes have the best separation based on seen samples. CTR prediction on raw data is difficult since from all impressions, clicks vary between 0.5 and 2 %. This causes big imbalances in raw data and models trained on such data are more prone to predict negative class. It is possible to counter this issue by applying random sampling methods that reduce the majority-class quantity or increase the minority-class size sample. Methods that create artificial samples based on K-NN algorithms such as ADASYN or SMOTE can also be used to solve this problem. Another approach to dealing with highly imbalanced data is using models to learn how normal representations of the selected class behave. In such cases, samples that belong to the same class should be considered normal but samples from other classes should be considered as outliers. For such tasks, outlier or novelty detection algorithms can be used. Outlier detection algorithms expect that outliers exist in the training data, however, novelty-based methods can be trained on a single class. Such methods predict novel samples on test data. Both types of algorithms are unsupervised, meaning they do not require samples to be labeled.

In this study, models from the scikit-learn library such as isolation forest and local outlier factors are used for outlier and novelty detection. The isolation forest algorithm is a decision tree-based algorithm that tries to isolate outliers from data. The algorithm randomly selects subsets of features and split values. Random partition creates shorter paths in trees with outliers. The local outlier factor is similar to the K-nearest neighbor algorithm, it computes the local density variation of input samples. Outliers have lower densities than their neighbors. Local outlier factor algorithms can also be used solely on a single class. In such cases, the model learns to recognize novel samples, therefore novelty parameter is set to true.

4. Data

The data used for this research was taken from one of the European advertising companies. Mobile in-app advertising from the United States, Australia, United Kingdom was used. In addition, click data was filtered out to contain mobile app ids that occur at least 1000 times in data. The number of samples

is 869031. Click data was taken for a full month from 2022 November 1st 00:00:00 to November 29th 23:59:59. The data contains attributes such as:

- Categorical features: agency id, client id, campaign id, placement id, banner id, tag id, inventory source id, region id, city id, zip code id, device type id, browser id, and mobile app category ids.
- Textual features: client industry vertical name, agency name, inventory source name, mobile app name, mobile app categories;
- DateTime: click and impression date-times;
- URL: landing URL.

Additional data that contained 5000 impressions and 5000 clicks data from December 28th was used for testing the best model and performing supplementary analysis.

5. Data processing

The data processing pipeline can be seen in Figure 1 left side. Mobile app ids that have lower occurrence than 1000 were filtered out with the samples that contained many NaN values (row contains more than 10 % missing values). Unknown values were replaced by mode values. Some features were engineered using existing attributes. Categorical features were converted to integers where each number represents a category. Textual data was also processed using natural language processing methods. The data was later joined and used with principal component analysis (PCA) compression algorithm.
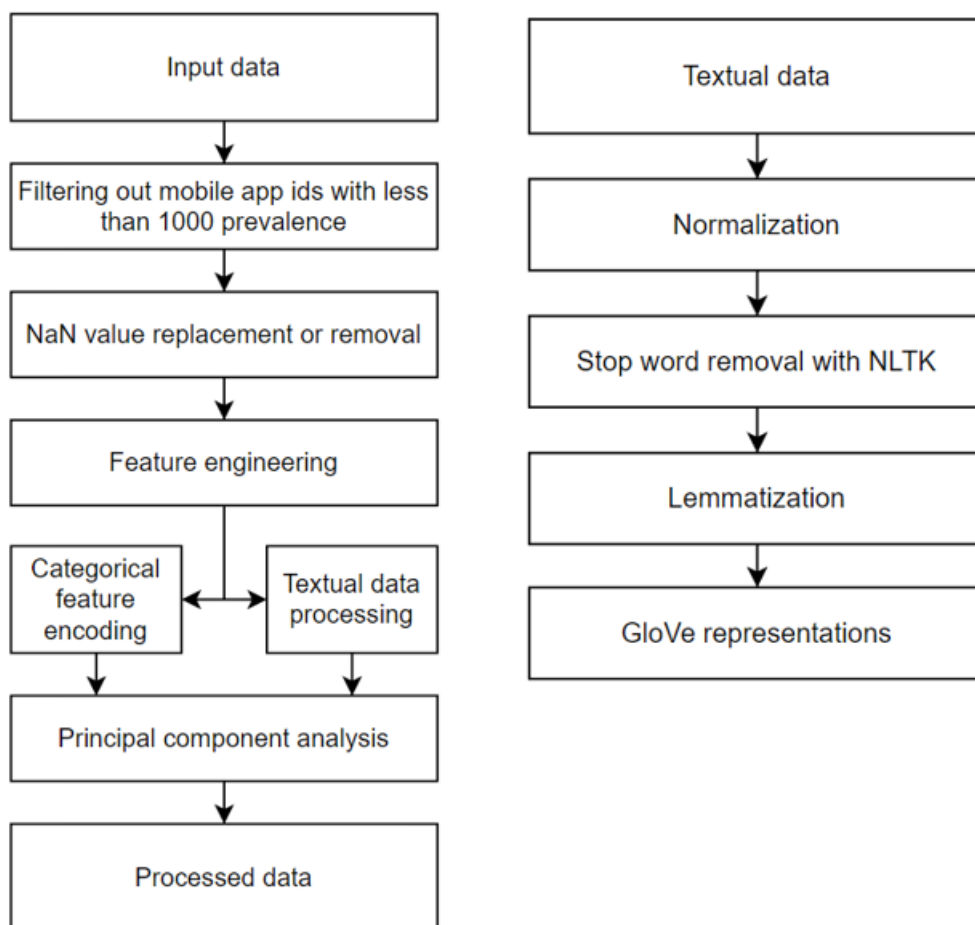

Figure 1: Data processing pipeline (left) and textual data processing pipeline (right).

The textual data processing pipeline can be seen in Figure 1 right side. Textual data was firstly normalized by removing symbols and numbers, changing words to lowercase. After it, stopwords that do not give any useful contextual information were removed. The list of stop-words was taken from the NLTK library. The remaining words were lemmatized to their basic form using the NLTK library. The word representations were extracted using GloVe model developed by Standford and trained on 2 billion tweets, using 50-dimension vectors. Textual data word embeddings were averaged, however, if the model did not contain data words, they were replaced with the same length zero vector.

Some additional features were generated from existing characteristics using feature engineering. The number of a weekday was used as an additional column and its' textual representation was extracted from GloVe. The logic behind this decision was that the model has learned the semantic meaning of each weekday name and their associations. For example, Saturday and Sunday are weekend days, indicating that the majority of people associate these words with rest, spending time with friends and families, investing time in hobbies. Therefore, these representations can help machine learning models to find patterns between the context of ads, mobile applications, and the clicks received by banners. The resulting set of characteristics contained 363 features. Principal component analysis (PCA) was used to reduce dimensions to 50 and 25.

In addition, hour and minute integer attributes were later added to the data creating 2nd version of data. Hour integer was converted to a daytime string:

- Morning between 6 and 12;
- Noon between 12 and 17;
- Evening between 17 and 23;
- Night as the remaining time.

The daytime string was processed using textual data pipeline, GloVe representations were extracted from strings. After additional columns were added, the total number of features extended to 415, creating 2nd version of the dataset.

6. Results

Both models (isolation forest and local outlier factor) were trained on data that contained over 800000 samples of click data. The models were trained to learn what is the standard behavior of users who clicked on mobile app ads using contextual data such as banner, app, date, and other information. The best model was used for additional testing on data that contains both click and impression data. If the model learns what cues determine whether the user will click on an ad, it should predict that click data is normal but should categorize impression data as outliers.

The results seen in Table I shows that the isolation forest performs better than the local outlier factor in all cases. Isolation forest with 300 trees performed the best when PCA with 50 components was applied to 1st version of the dataset, however, the difference between using PCA with the same parameters on 2nd version dataset was only slight – less than 1 % accuracy difference. Using PCA with 25 components yielded an even lower difference, while the same algorithm was used – less than 0.5 % difference. The results indicate, that additional generated features such as hour and minute did not provide any useful information for the isolation forest algorithm. Added features did not help the algorithm to distinguish user behavior that resulted in a click, only making outlier separation slightly harder. In addition, using lower dimension data (363 features versus 50 features after PCA) with isolation forest increased model performance by almost 20 %. This can indicate that the isolation forest algorithm works better with lower-dimension data.

The local outlier factor algorithm performed better when the 2nd version dataset with 50 PCA components was used and reached the best results, compared to the algorithm trained on original dimension data without the added features as well as its PCA version. In addition, the 2nd version dataset results with 50 and 25 PCA components differed only by 4 %, while the difference between the 2nd version dataset with 50 components and the 1st version with 50 components was 15 %. This can indicate that hour, minute, and daytime GloVe representation features give additional information to the local outlier factor algorithm that help to determine whether the sample is novelty (a click) or not. The model reached higher accuracy by almost 4 % using original dimension data compared to the same

type of data with PCA compression indicating that the model can handle higher dimension data well. Using compression can worsen results since some information is lost using the PCA algorithm. However, the model reached slightly higher (by over 1 %) accuracy when data was compressed to 25 components versus 50 components.

The isolation forest algorithm performed better than the local outlier factor algorithm on all datasets except the original data which was not compressed – the accuracy difference was around 9 %. The only cases, when the local outlier factor algorithm was close to the performance of the isolation forest was when the 2nd version dataset was used. The accuracy difference between the models ranged from 1 % to 2 %. The isolation forest algorithm performed best when lower dimension data was used, while local outlier factor algorithm performance was not associated with data dimensions but rather the data quality.

Table 1

Model results on different types of datasets. Datasets include samples only from a single class.

Description	Isolation forest (300 trees)	Local outlier factor
Dataset with 363 features	73.57 %	82.12 %
PCA 50 components on the dataset 1st version	93.99 %	77.18 %
PCA 25 components on the dataset 1st version	90.52 %	78.58 %
PCA 50 components on the dataset 2nd version	93.13 %	92.33 %
PCA 25 components on the dataset 2nd version	90.21 %	88.3 %

Data with 10000 samples were used to measure the ability of the best model to distinguish clicks from impressions. The isolation forest algorithm with 300 trees trained on the 2nd version dataset with 50 PCA components was used for testing. The model reached 42.7 % accuracy indicating that it could not differentiate clicks from impressions. The testing result confusion matrix can be seen in Figure 2. It is visible, that the model struggles to distinguish an impression from a click since it predicted over 86 % of impressions as clicks, while almost 72 % of clicks were predicted as clicks. The model correctly predicted only 13.6 % of impressions as outliers, showing that it mostly predicts both impressions and clicks as clicks. Random classifier reached 50 % of accuracy by predicting a single class for all samples. This indicates, that the isolation forest algorithm performed worse than the random classifier, therefore the trained model is not usable in real life.
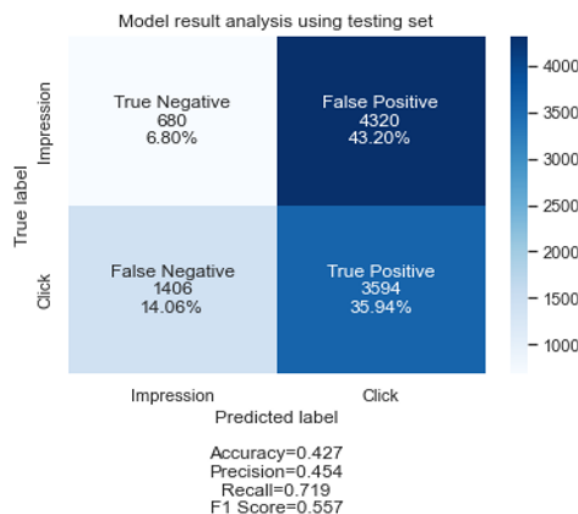

Figure 2: Confusion matrix results on additional testing data with 5000 click and 5000 impression data trained on the best-performing isolation forest algorithm.

Model results were visualized in Figure 3 using 2 component PCA. The left graph indicates the distribution of compressed impression and click data. Three clusters of data are visible: the middle cluster contained click data, while others had a mixture of both classes. The right side of the graph indicates the model predictions of each sample. The samples of a middle cluster that belonged to the click class were mostly predicted as false impressions. The model predicted these samples as outliers, although they were samples from click data. Other clusters had a mixture of predictions with no clear patterns visible.
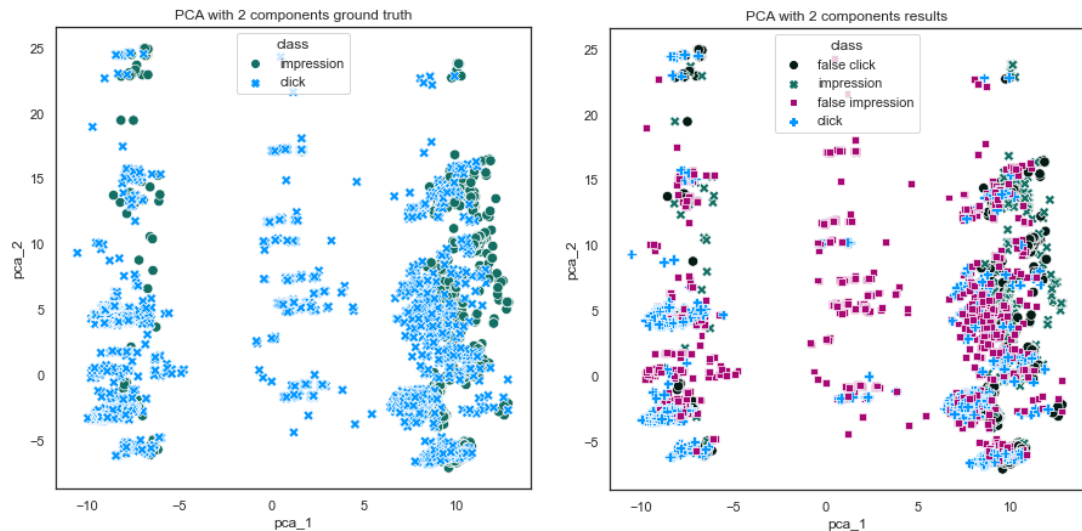

Figure 3: The left side shows data with 5000 click and 5000 impression data visualized using 2 component PCA. On the right side, best-performing isolation forest algorithm predictions were visualized.

7. Conclusions

Contextual targeting is an alternative to third-party cookie-based digital advertising methods. Contextual targeting focuses on context in which the banner is displayed, rather than the history of user who visits the website. Such method creates targeting strategy that protects users' privacy, identity and helps to regulate personal data ownership. The need of such algorithms increases as most popular browsers are planning to remove third-party cookies in following years and their removal can cost up to 10 billion in US publisher revenue [7]. CTR prediction methods can be used in solving these problems. However these methods rely on deep learning and machine learning classifiers that require abundance of data for training process in order to learn to distinguish clicks. Huge amount of data creates multiple other problems – data imbalance and long training time. Data imbalance is natural phenomenon in this domain, since most of the user interactions with an ad lead to an impression rather than a click. In addition, using large amounts of data for training increase the number of resources required to complete the model training. Unsupervised learning methods such as isolation forest and local outlier factor algorithms are used as the alternatives. Both models are fast, can be trained on a single class data. In such case, models learn that normal samples are clicks and outliers or novelties are impressions.

In this paper, contextual data enhanced with GloVe extracted embeddings was used for model training. The performed analysis showed that isolation forest algorithm with 300 trees was better at recognizing clicks than local outlier factor algorithm (1 – 2 % difference on 2nd version dataset and 12 – 13 % on 1st version after PCA compression). The local outlier factor algorithm reached similar performance to isolation forest when dataset was enriched with additional daytime features and embeddings. However, when isolation forest algorithm trained on 50-dimensional PCA data was tested on additional data, it reached 43 % accuracy which was lower than a baseline of a random classifier that reaches 50 % accuracy. The results show that contextual attributes such as ad, mobile app

information, click metadata are not enough for the creation of a model that can predict whether the user will click on an ad or not. Additional information about the user retrieved from browsing history is required to make better recommendations and thus make better CTR predictions. This is required to achieve better performance, since users have different preferences and interests, therefore relying solely on context will not yield good results.

The results of this study are consistent with research [15] which found that context does not add much value. However, the latter research was focused on CTR prediction, whereas the purpose of our study was to investigate whether anomaly or novelty detection methods can learn to recognize patterns in click data from context. For this reason, the contextual data used in the study has been augmented with additional engineered features such as hour and day of the week and their extracted GloVe embeddings, in hope that these might provide important information about the semantic meanings of the attributes. However, despite the data showing that the day and time of the week affect the number of clicks (since most clicks occur when users are actively using the Internet but not resting), the results of this study showed that anomaly and outlier detection methods using GloVe embeddings do not capture these contextual data relationships.

8. References

[1] Competition & Markets Authority (CMA): Online platforms and digital advertising. Market study final report, 2020. URL: https://assets.publishing.service.gov.uk/media/5fa557668fa8f5788db46efc/Final_report_Digital_ ALT_TEXT.pdf

[2] CookiePro, what is cookie profiling? 2021. URL: https://www.cookiepro.com/knowledge/what-is-cookie-profiling/

[3] Bleier, Alexander. On the Viability of Contextual Advertising as a Privacy-Preserving Alternative to Behavioral Advertising on the Web, 2021. SSRN Electronic Journal. 10.2139/ssrn.3980001.

[4] Deloitte.Digital: Goodbye third-party cookies. Hello human experience, 2020. URL: https://www.deloittedigital.com/content/dam/deloittedigital/us/documents/offerings/offering-20200206-third-party-cookies_new.pdf

[5] Papadogiannakis E., P. Papadopoulos, N. Kourtellis, and E. P. Markatos. User Tracking in the Post-cookie Era: How Websites Bypass GDPR Consent to Track Users. In Proceedings of the Web Conference 2021 (WWW '21). Association for Computing Machinery, New York, NY, USA, 2130–2141. DOI: https://doi.org/10.1145/3442381.3450056

[6] Franken G., T. V. Geothem, W. Joosen: Who Left Open the Cookie Jar? A Comprehensive Evaluation of Third-Party Cookie Policies, 2018. URL: https://www.researchgate.net/publication/333133018_Who_Left_Open_the_Cookie_Jar_A_Com prehensive_Evaluation_of_Third-Party_Cookie_Policies

[7] IAB: The demise of third-party cookies and identifiers: What it means for digital advertising in the US, 2021. URL: https://www.iab.com/wp-content/uploads/2021/03/IAB_McKinsey_State_of_Data_2021-03.pdf

[8] Digiday: The State of Contextual Targeting, 2021. URL: https://digiday.com/wp-content/uploads/2021/07/CONNATIX_State-of-Contextual_071521.pdf

[9] Kononova A., W. Kim, E. Joo, K. Lynch: Click, click, ad: the proportion of relevant (vs. irrelevant) ads matters when advertising within paginated online content, International Journal of Advertising, 2020. DOI: 10.1080/02650487.2020.1732114

[10] Sreejesh S., T. Gosh, Y. K. Dwivedi: Moving beyond the content: The role of contextual cues in the effectiveness of gamification of advertising, 2021. Journal of Business Research, Elsevier, vol. 132(C), pages 88-101. DOI: 10.1016/j.jbusres.2021.04.007

[11] Resnick, Marc L. and William S. Albert. "The Impact of Advertising Location and User Task on The Emergence of Banner Ad Blindness." Proceedings of the Human Factors and Ergonomics Society Annual Meeting 57, 2013: 1037 - 1041.

[12] Peker, S.; Menekse Dalveren, G.G.; ˙Inal, Y. The Effects of the Content Elements of Online Banner Ads on Visual Attention: Evidence from An-Eye-Tracking Study. Future Internet 2021, 13, 18. https://doi.org/10.3390/fi130100181

[13] Furong Wang and Li Zhao. A Hybrid Model for Commercial Brand Marketing Prediction Based on Multiple Features with Image Processing Security and Communication Networks Volume 2022, Article ID 5455745, 10 pages DOI: https://doi.org/10.1155/2022/5455745

[14] Omid Rafieian, Hema Yoganarasimhan. Targeting and Privacy in Mobile Advertising MARKETING SCIENCE Articles in Advance, pp. 1–26 ISSN 0732-2399 (print), ISSN 1526-548X (online). URL: http://pubsonline.informs.org/journal/mksc

[15] Ouyang, Wentao Robin, Xiuwu Zhang, Li Li, Heng Zou, Xin Xing, Zhaojie Liu and Yanlong Du. "Deep Spatio-Temporal Neural Networks for Click-Through Rate Prediction." Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (2019): n. pag.

[16] Mao, K., Zhu, J., Su, L., Cai, G., Li, Y., Dong, Z. 2023. FinalMLP: An Enhanced Two-Stream MLP Model for CTR Prediction. arXiv e-prints. doi:10.48550/arXiv.2304.00902

[17] Song, Xue, Baohan Xu and Yu-Gang Jiang. "Predicting Content Similarity via Multimodal Modeling for Video-In-Video Advertising." *IEEE Transactions on Circuits and Systems for Video Technology* 31 (2021): 569-581.

[18] Jun Long, Lei Liu, Hongxiao Fei, Yiping Xiang, Haoran Li, Wenti Huang, and Liu Yang. Contextual Semantic-Guided Entity-Centric GCN for Relation Extraction. Mathematics 2022, 10(8), 1344; DOI: https://doi.org/10.3390/math10081344

[19] Jin, Yiping, et al. "Bootstrapping Large-Scale Fine-Grained Contextual Advertising Classifier from Wikipedia." Proceedings of the Fifteenth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-15), 2021. Crossref, https://doi.org/10.18653/v1/11.textgraphs-1.1.

[20] Chen, Xiang & Nguyen, Tam & Shen, Zhiqi & Kankanhalli, Mohan. LiveSense: Contextual Advertising in Live Streaming Videos, 2019. 392-400. 10.1145/3343031.3350888.

[21] Ouyang, Wentao, et al. "MiNet." Proceedings of the 29th ACM International Conference on Information & Knowledge Management, Oct. 2020. Crossref, DOI: https://doi.org/10.1145/3340531.3412728.

[22] Denison, Carson E., Badih Ghazi, Pritish Kamath, Ravi Kumar, Pasin Manurangsi, Krishnagiri Narra, Amer Sinha, Avinash V. Varadarajan and Chiyuan Zhang. "Private Ad Modeling with DP-SGD." *ArXiv* abs/2211.11896 (2022): n. pag.

[23] Elsayed, Shereen and Lars Schmidt-Thieme. "Deep Multi-Representation Model for Click-Through Rate Prediction." *ArXiv* abs/2210.10664 (2022): n. pag.

[24] Yiping (Amy) Song, Chee Wei (David) Phang, Shuai Yang & Xueming Luo (2018) The Effectiveness of Contextual Competitive Targeting in Conjunction with Promotional Incentives, International Journal of Electronic Commerce, 22:3, 349-385, DOI: 10.1080/10864415.2018.1462952.

[25] Pan, Feiyang, Qingpeng Cai, Pingzhong Tang, Fuzhen Zhuang and Qing He. "Policy Gradients for Contextual Recommendations." *The World Wide Web Conference* (2018): n. pag.

[26] Ban, Yikun and Jingrui He. "Convolutional Neural Bandit: Provable Algorithm for Visual-aware Advertising." *ArXiv* abs/2107.07438 (2021): n. pag.

[27] Geng, Tong, Xiliang Lin and Harikesh S. Nair. "Online Evaluation of Audiences for Targeted Advertising via Bandit Experiments." *AAAI Conference on Artificial Intelligence* (2019).