Abstract

The main topic of scientific thesis is the creation of a marketing profitable intelligent system for the provision of advertising using elements of neural networks of machine learning and computer vision. The use of machine learning in this area is at an early stage of development and the use of computer vision to deliver effective and profitable advertising is generally unique. Modern solutions allow only using a limited amount of personal information for processing and analysis, while our system allows you to use many external factors to generate it.

The main purpose of the study is to develop software in the field of machine learning and neural networks from which will be used to provide high-quality and profitable advertising. Resulting in a software product that provides high-quality content and meets modern standards for software development and creation of neural networks. For the implementation of this project, the Python language will be used with the libraries tenserflowers pytorch, as well as other libraries used for machine learning, as well as for data processing, video streaming will be used AND a neural network that allows you to determine and record the age and gender of the target audience.
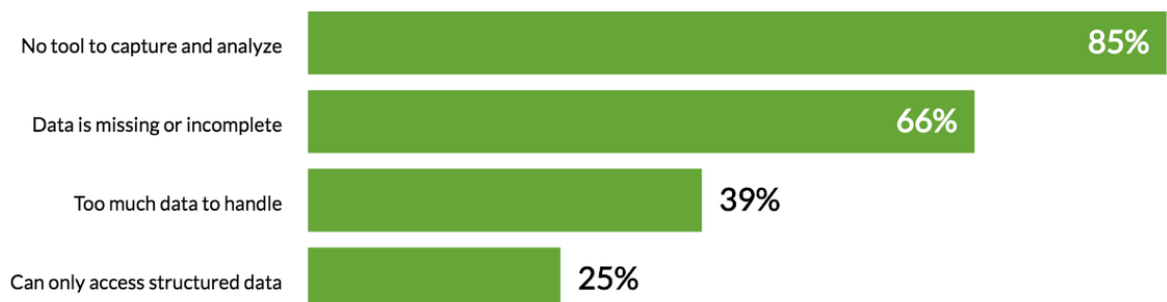
1.1

From digital transactions and retail stock to the temperature of server rooms, there's little the modern business can't track.

And although more data means more opportunities to improve, that's only true if you have what you need to analyze it. Unfortunately, most businesses do not.

Reports show that more than half of today's data goes unused. Known as "dark data," the top reasons it's not being used are as follows:

Why Companies Aren't Using Dark Data
Percentage of respondents that chose each answer

No tool to capture and analyze	85%
Data is missing or incomplete	66%
Too much data to handle	39%
Can only access structured data	25%

Data source: Datumize

splunk>

The solution?

Already, 95% of advertisers have terabytes upon petabytes of demographic data, including personal data, location information, and interests they can use to target prospects they know almost nothing about. Artificial intelligence is a way to tame that data and take it to the next level.

Still, advertisers are only beginning to scratch the surface of AI's potential.

Artificial intelligence is a technology that makes computers capable of performing tasks that would otherwise require a human brain. AI is promoting a revolution in several industries, including our own, allowing the media-buying process to be automated.

To do that, we have to use machine learning, a type of AI that gives computers the ability to learn things by being programmed specifically to the tasks at hand, just the way the human brain does.

Machine learning systems are made of algorithms smart enough to understand data and draw conclusions and correlations from it. As a result, they can diagnose, predict and plan things. They can also teach themselves to become better in a certain area (media buying, for example) and improve their intelligence over time as they get more exposure to data.

In advertising, machine learning allows us to essentially replicate the brain of an experienced buyer as software to make the same optimizations a buyer would. Plus, the system learns over time and generates more accurate results as it works with new campaigns, making correlations that can be tough for the human brain to detect.

There is not a lot of machine learning software available in the market, which means that advertisers who start using it now can gain a competitive advantage.

1. Machine learning can predict and boost ad performance

We know multiple variables can affect your advertising performance on social media. You can optimize your campaign by gender, age, geolocation, device or time of day, among other variables.

With machine learning, you can easily detect what adjustments need to be made in your strategy. The system will generate suggestions based on your goals, the amount of time and money you have and the results of multivariate testing.

This is possible because the system assesses the chances of reaching positive outcomes based on data originated from multiple sources, including your historical ad performance and the performance of similar advertisers.

2. Machine learning can draw correlations

Social media platforms are an incredible source of relevant data. These networks are the place where people talk about their interests, follow their favorite artists and comment on the places they have visited. Machine-learning systems can access this information to generate inputs that will help advertisers reach their primary target audience more precisely.

When you feed a computer a big pile of data, you can obtain interesting correlations that would be harder for the human brain to produce. A system that uses AI might conclude, after analyzing some data, that young men who like a particular type of music and are interested in sports are more likely to download an app, for example. If the goal of your campaign is apps downloads, the system will make sure this audience will be able to see your ads, which will be reached through optimizations a buyer maybe would never think of doing.

3. Machine learning can make media-buying teams smarter

Inexperienced media buyers often run into mistakes. For instance, some might not know that advertising costs are twice as expensive in December as they are in January. That could easily result in overpromising and under-delivering on CPV to a client.

With machine learning, however, you can more easily predict the outcomes of a campaign. The software stores all past campaign data and continuously updates information about customers' behaviors. The story is in the data, and the algorithms can understand and learn from it.

Machine learning also makes it easier to scale a team. Machine learning simplifies the work of media buyers, freeing them to focus on big-picture thinking.

4. Machine learning can reduce costs

Results improve, and costs lower, as you will maximize the results for each ad dollar you invest. It can go as far as dropping the CPA of a campaign from $60 to $30.

This is difficult to achieve if you only have one campaign as a reference or many spreadsheets around you. With machine learning, however, you have a powerful system that can easily comb through a significant amount of data.

Related: The future of ad targeting with computer vision and image understanding

5. Machine learning can create better reports

Media buyers can automatically generate detailed reports that are easy to read and share. The reports get straight to the point, and there is no need to develop advanced calculations to understand the stories the numbers want to tell you.
Boost contextual relevance

Computer vision is an interdisciplinary field that deals with how computers can be made to gain high-level understanding from digital images or videos. From the perspective of engineering, it seeks to automate tasks that the human visual system can do.

Computer vision has been heralded as the future of advertising given its potential for even greater ad targeting, and it's true, it is a real game changer for the industry. Why? Because it has the potential to analyze how consumers react to advertising and to glean consumer information from the millions upon millions of photos posted on social networks.

Advancements in computer vision can help you develop new insights about your customers and more precisely formulate and target your ad campaigns going forward.

People Tracking for Optimization and Non-Digital Ad Attribution

Modern computer vision technology has progressed to tracking human behaviors in real-time through a live video feed, like how autonomous cars can sense pedestrians. Cameras in a retail location, for instance, can draw conclusions about a store layout and shelf arrangement based on customer traffic, how they move throughout a space, and where their gaze falls using facial recognition. For example, the system can accurately track the first area that a majority of customers go to when they enter a store and which products are drawing the most attention.

This technology can also serve as a form of attribution for ad formats such as outdoor signage and potentially TV viewership. Just as how digital ads can track impressions versus clicks, an outdoor sign could track how many people walked past the ad versus how many actually looked at it, how long they looked at it, and even estimate individual demographics like age and gender.

Gathering Data for Emotional Analytics and Tracking Consumer Attention

Similar to how social listening technology can gauge sentiment within written content, computer vision systems can track and measure emotional reactions to ad creative. This data is important because self-reported emotions can be inaccurate, especially if the subject's face is telling a different story.

Using Visual Data for Customer Personalization
Using computer vision, companies can gather real-time visual data on customers to personalize experiences and inform marketing strategy. Select McDonald's locations have implemented camera-equipped kiosks that suggest menu items based on the customer's perceived age and gender.

In another example, some analysts theorize that the addition of a camera-equipped smart speaker to the Amazon Echo lineup could give Amazon the ability to gather customer data for more effective cross-sells. By observing what people wear and what they bring into their homes, the company can learn which products to restock or suggest for purchase.

Smart advertising offers

- *Personalized Smart Billboard System (PSBS)*, which enables the personalization of promotional messages. The system detects faces from waist-up photos and segments the facial region into subregions, such as forehead, eye, mouth and neck. Within these sub-regions, PSBS attempts to find predetermined accessories, including hats, glasses, sunglasses, necklaces and facial hair such as a beard or moustache. Once any of these clues are found, related advertisements are shown to the customer. The contribution of this study is threefold. First, a new system is proposed in which a smart billboard shows the advertisements concurrent with customers' accessories and some of their facial features. In this respect, the system can be regarded as novel and complementary to common smart advertisement systems, which are typically based on the identification of gender and age profiles. The second contribution is the low cost of the proposed system. In PSBS, feature extraction can be performed using a single image, instead of employing a video sequence or a motion sensor, which reasonably reduces the information system resources needed to implement the model in practice. Final contribution is the proposed computational pipeline that integrates relevant computer vision techniques to detect domain-specific objects.

- *Dynamic Yield's AI-powered personalization* engine works to create the optimal customer experience by automatically tailoring an interaction with a user based on their past activity or behavior, as well as by leveraging real-time signals. Powered by machine learning, a product recommender system is the technology used to suggest which products are shown to individuals interacting with a brand's digital properties. Fueled by a number of algorithmic decisions, recommendation algorithms mine user, product, and contextual data – both onsite and offsite – to present every user with a personalized experience. And when it comes to product recommendations, there is no archetypal strategy marketers should use for every widget. Different strategies must be applied for different users, depending on the amount of information available about the customer, their behavior, and the context

of products on a site. This includes site behavior, status, geo-location, time of day, past purchases, and more.

Instagram algorithm. Instagram identifies accounts that are similar to one another by adapting a common machine learning method known as "word embedding." Word embedding systems study the order in which words appear in text to measure how related they are. So, for example, a word embedding system would note that the word "fire" often appears next to the words "alarm" and "truck," but less frequently next to the words "pelican" or "sandwich." Instagram uses a similar process to determine how related any two accounts are to one another.

Content on Instagram is varied, so the company looks more at what accounts you might like rather than posts. Photo by Amelia Holowaty Krales / The Verge
To make its recommendations, the Explore system begins by looking at "seed accounts," which are accounts that users have interacted with in the past by liking or saving their content. It identifies accounts similar to these, and from them, it selects 500 pieces of content. These candidates are filtered to remove spam, misinformation, and "likely policy-violating content," and the remaining posts are ranked based on how likely a user is to interact with each one. Finally, the top 25 posts are sent to the first page of the user's Explore tab.

1.2

Age and Gender Classification using Convolutional Neural Networks

Age and gender play fundamental roles in social interactions. Languages reserve different salutations and grammar rules for men or women, and very often different vocabularies are used when addressing elders compared to young people. Despite the basic roles these attributes play in our day-to-day lives, the ability to automatically estimate them accurately and reliably from face images is still far from meeting the needs of commercial applications. This is particularly perplexing when considering recent claims to super-human capabilities in the related task of face recognition (e.g.).
Past approaches to estimating or classifying these attributes from face images have relied on differences in facial feature dimensions or "tailored" face descriptors. Most have employed classification schemes designed particularly for age or gender estimation tasks, including and others. Few of these past methods were designed to handle the many challenges of unconstrained imaging conditions [10]. Moreover, the machine learning methods employed by these systems did not fully challenges of age and gender estimation from real-world, unconstrained images. Most notably, extreme blur (low-resolution), occlusions, out-of-plane pose variations, expressions and more.
In this paper we attempt to close the gap between automatic face recognition capabilities and those of age and gender estimation methods. To this end, we follow the successful example laid down by recent face recognition systems:

Face recognition techniques described in the last few years have shown that tremendous progress can be made by the use of deep convolutional neural networks (CNN). We demonstrate similar gains with a simple network architecture, designed by considering the rather limited availability of accurate age and gender labels in existing face data sets.

Age and Gender Classification

Age classification. The problem of automatically extracting age related attributes from facial images has received increasing attention in recent years and many methods have been put fourth. A detailed survey of such methods can be found in and, more recently. We note that despite our focus here on age group classification rather than precise age estimation (i.e., age regression), the survey below includes methods designed for either task.

Early methods for age estimation are based on calculating ratios between different measurements of facial features. Once facial features (e.g. eyes, nose, mouth, chin, etc.) are localized and their sizes and distances measured, ratios between them are calculated and used for classifying the face into different age categories according to hand-crafted rules. More recently, uses a similar approach to model age progression in subjects under 18 years old. As those methods require accurate localization of facial features, a challenging problem by itself, they are unsuitable for in-the-wild images which one may expect to find on social platforms. On a different line of work are methods that represent the aging process as a subspace or a manifold. A drawback of those methods is that they require input images to be near-frontal and well-aligned. These methods therefore present experimental results only on constrained datasets of near-frontal images. Again, consequently, such methods are ill-suited for unconstrained images. Different from those described above are methods that use local features for representing face images.

Gaussian Mixture Models (GMM) were used to represent the distribution of facial patches. In [54] GMM were used again for representing the distribution of local facial measurements, but robust descriptors were used instead of pixel patches. Finally, instead of GMM, Hidden-Marko Model, super-vectors were used in for representing face patch distributions.

An alternative to the local image intensity patches is robust image descriptors: Gabor image descriptors were used in along with a Fuzzy-LDA classifier which considers a face image as belonging to more than one age class. In a combination of Biologically Inspired Features (BIF) and various manifold-learning methods were used for age estimation. Gabor and local binary patterns (LBP) features were used in along with a hierarchical age classifier composed of Support Vector Machines (SVM) to classify the input image to an age-class followed by a support vector regression to estimate a precise age. Finally, proposed improved versions of relevant component analysis and locally preserving projections. Those methods are used for distance learning and dimensionality reduction, respectively, with Active Appearance Models as an image feature. All these methods have proven effective on small and/or constrained benchmarks for age estimation. To our knowledge, the

best performing methods were demonstrated on the Group Photos benchmark. In state-of-the-art performance on this benchmark was presented by employing LBP descriptor variations and a dropout-SVM classifier. We show our proposed method to outperform the results they report on the more challenging Adience benchmark, designed for the same task.

Gender classification.

Here we quickly survey relevant methods. One of the early methods for gender classification used a neural network trained on a small set of near-frontal face images. In the combined 3D structure of the head (obtained using a laser scanner) and image intensities were used for classifying gender. SVM classifiers were used by, applied directly to image intensities.

Rather than using SVM, used AdaBoost for the same purpose, here again, applied to image intensities. More recently, used the Weber's Local texture Descriptor for gender recognition, demonstrating near perfect performance on the FERET benchmark. In, intensity, shape and texture features were used with mutual information, again obtaining near-perfect results on the FERET benchmark.

Deep convolutional neural networks

One of the first applications of convolutional neural networks (CNN) is perhaps the LeNet-5 network described by for optical character recognition. Compared to modern deep CNN, their network was relatively modest due to the limited computational resources of the time and the algorithmic challenges of training bigger networks. Though much potential laid in deeper CNN architectures (networks with more neuron layers), only recently have they became prevalent, following the dramatic increase in
both the computational power (due to Graphical Processing Units), the amount of training data readily available on the Internet, and the development of more effective methods for training such complex models. One recent and notable examples is the use of deep CNN for image classification
on the challenging Imagenet benchmark. Deep CNN have additionally been successfully applied to applications including human pose estimation, face parsing, facial keypoint detection, speech recognition and action classification. To our knowledge, this is the first report of their application to the tasks of age and gender classification from unconstrained photos.

A CNN for age and gender estimation

Gathering a large, labeled image training set for age and gender estimation from social image repositories requires either access to personal information on the subjects appearing in the images (their birth date and gender), which is often private, or is tedious and time-consuming to manually label. Data-sets for age and gender estimation from real-world social images are therefore relatively limited in size and

presently no match in size with the much larger image classification data-sets. Overfitting is common problem when machine learning based methods are used on such small image collections.

This problem is exacerbated when considering deep convolutional neural networks due to their huge numbers of model parameters. Care must therefore be taken in order to avoid overfitting under such circumstances.