BBVA

DATA & ANALYTICS

Show & Tell

Shaping TRUST

with machine learning

Summary of presentations

October 2017

BBVA

DATA & ANALYTICS

Authors of this document

Pablo Fleurquin
pablo.fleurquin@bbvadata.com
Fabien Girardin
fabien.girardin@bbvadata.com
Israel Viadest
jacobisrael.viadest@bbvadata.com

Last edit: November 13, 2017.

Index

Introduction

What is Show & Tell EDGE

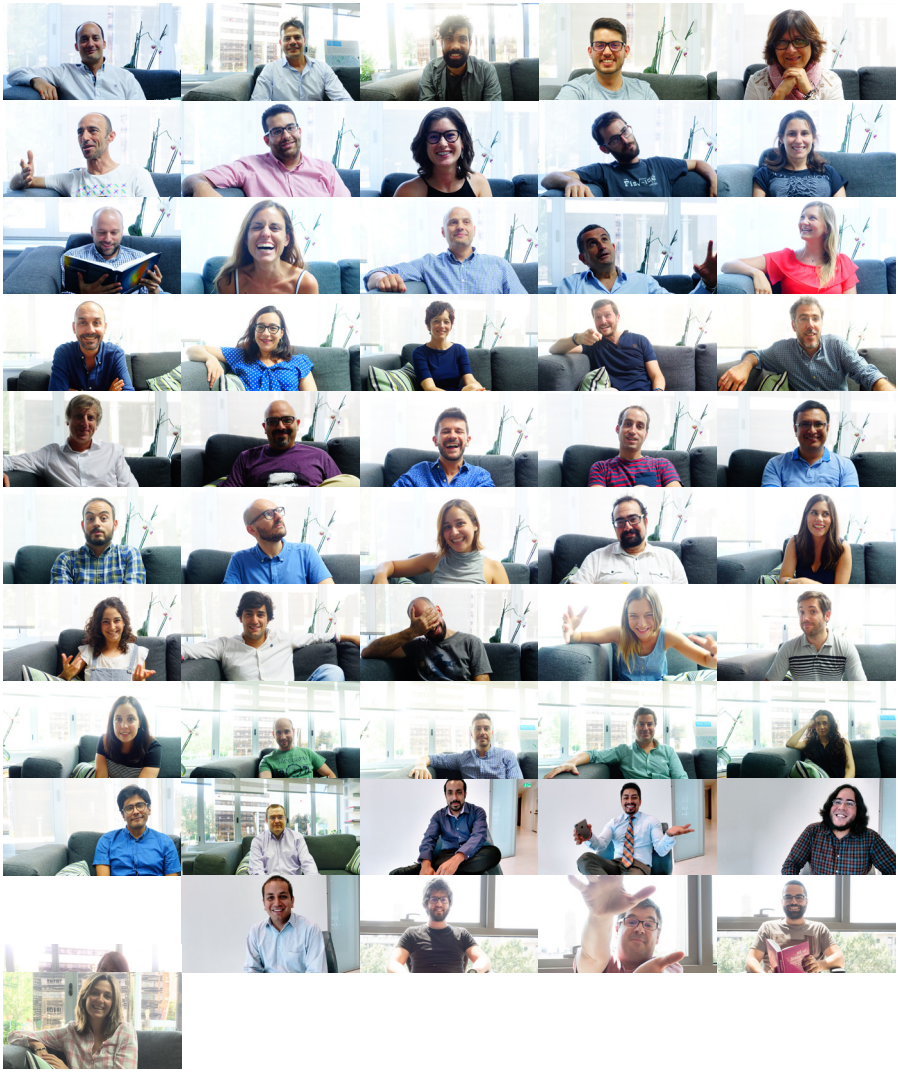
We live in a world where 50% of Google's code is rewritten daily; Where 90% of the world's data is two years old or less. At **BBVA Data & Analytics** we believe in creating the future before it happens to us. Show & Tell EDGE is our regular public gathering to share the recently possible that emerged from our research and innovation initiatives in advanced analytics and data science. It is a hub of inspiration and exploration for passionate BBVA designers, engineers, data scientists and product managers who need to get ahead in this complex world with ingredients of the futures of the data-driven bank.

Show & Tell EDGE is about provoking dialogues and push research and innovation beyond their natural environment within BBVA. Along with the presence of 30+ people from Data and Open Innovation, Service Design, Global Products, AI Program, Innovation Labs, BBVA Chairman's Office and more... we hope to discuss new practices, new data products, new models, new analytical capabilities and try to link them to your domains of activity.

What is BBVA Data & Analytics

BBVA Data & Analytics is a center of excellence in financial data analysis. We are 40+ data scientists, technologists, domain experts, strategy practitioners, and visual design thinkers. Along with Banco Bilbao Vizcaya Argentaria (BBVA) we believe the intelligence derived from algorithms can transform the banking industry, its relation with customers and its role in the world.

Email: hello@bbvadata.com
Twitter: @bbvadata
Web: bbvadata.com



EDGE #1:
Introduction to Trust and Machine Learning: Fairness & Interpretability

Contributors

Juan Murillo
Partnerships & New Data Products / Data Project Management

Juan holds a MSc in Civil Engineering and an executive MBA. He's currently in charge of territorial analysis and knowledge sharing at BBVA D&A where, thanks to his background as an urban planner, he contributes to develop applications of data reuse and to create new informational services that provide macroeconomic insights built upon microtransactional data. He has been co-author of several scientific papers in cooperation with MIT Senseable City Lab and other academic institutions.

Contact:
juan.murillo.arias@bbvadata.com

Pablo Fleurquin
Risk & Fraud Analytics / Data Science

Pablo holds a Ph.D. in Physics by the CSIC center Institute for Cross-Disciplinary Physics and Complex Systems in Palma de Mallorca. The theoretical framework of his Thesis rest upon what is known as Complex Network Theory or Graph Analytics. He has been co-author of several scientific papers and a spanish patent related to his PhD work . Outside academia he worked in Machine Learning models for online credit, card fraud analytics & pricing strategies. He currently works developing and enhancing models, tools and techniques in the D&A Risk & Fraud team.

Contact:
pablo.fleurquin@bbvadata.com

Fig. 1. The Trust virtuous cycle of customer experience

Abstract

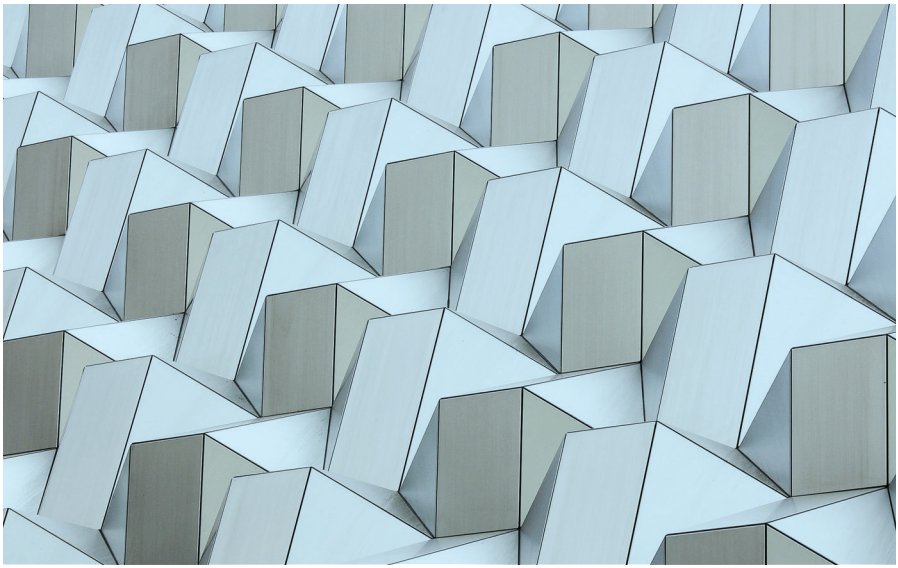
Trust is at the bedrock of our human social system. Historically, the financial businesses have been based on how it could trust customers, and not the other way around, hTrust is at the bedrock of our human social system. Historically, the financial businesses have been based on how it could trust customers, and not the other way around, hence quality, transparency and altruism was demanded on the side of the customer. Therefore, a bank is often perceived as a partner people need to live with, but that are prone to mislead, provoke unfair situations and take advantage of opaque processes. Nowadays the huge digital footprint of our customers and our analytic capabilities provide us with a much deeper vision about our customers, opening new means to build trust. For instance, to personalize the experience of products and services, to advise our customers with accurate forecasts and alerts, to generate opportunities as more credit access with less risk, as well as to automate user interaction.

Those technological opportunities also open risks that may drive mistrust. There is an increasing amount of dysfunctional solutions based on partial data or in bad quality data that erode trust and disengage customers, besides posing a risk proportional to the kind of service provided (e.g. bad recommendations about leisure choices or financial decisions imply very different levels of risk). In addition, privacy violation is a major concern nowadays. Similarly, discrimination like unfair access to societal goods is becoming pervasive and has reinforced the threat. Examples are everywhere such as 2013 Ally Financial 98M

US$ suit on auto-loan discrimination. However in this particular case, the Consumer Financial Protection Bureau's (CFPB) used an algorithm to infer a borrower's race based on other information on Ally Financial applications. Other border-line use of technology is in recidivism models such as the LSI-R in the United States. These help the judicial system to assess the danger posed by each convict. But the question still remains; do they eliminate human biases in historical data or just camouflaged it into a black-box model?

Up to this point, we must realize that technological innovations are not intrinsically "good" or bad. It is how humans use their "superpowers" that pose ethical questions. As a data based organization we must be transparent and responsible through our decision-making process, being it algorithmically driven or not. Hence from a data-driven perspective we must address potential problems as the ones described turning the tide to generate trust from the customer perspective. The solution is not straightforward and several organizational aspects must be involved. First, dysfunctional solutions must be tackled by fostering the organization culture towards customer-centric design, with a transparent and effective Data Governance (QA) and promoting critical thinking. Second, privacy violation must be addressed with functional and safe infraestructures and channels, and working with anonymous data whenever the use case allows it. Finally, discrimination should be restrained by advocating data literacy and responsibility at all levels of the organization. Furthermore, from a perspective of advanced data analytics things can be done in order to tackle this burden. The main foundations of Trust from a Machine Learning perspective are the concepts of transparency and fairness.

"Nowadays the huge digital footprint of our customers and our analytic capabilities provide us with a much deeper vision about our customers, opening new means to build trust."



"Fairness is always the result of a comparative process. This can be twofold; as a comparative process with a past situation with my own self or a comparative process with another person independently of time."

Fairness is always the result of a comparative process. This can be twofold; as a comparative process with a past situation with my own self or a comparative process with another person independently of time. For example, in the former case, we can consider a price increase in a certain product, given incomplete market information, as unfair. In this, anticipating buyers discrepancies and the transparency of the vendor explaining why price has increase can ameliorate the unfairness sensation. In the latter case, we base our fairness assumptions by comparing to others. Things are more intricate, because one must address, subjectively, how alike one is to the comparative others. If there is a price reduction in a certain product for people considered as peers, odds are that the comparison will provoke an unfair situation. A good example of it was the uproar that took place with Amazon dynamic pricing model when people realized that the model had charged some people more than others. Unfairness of the

second type can be explicitly solved by including fairness metrics as another component of the algorithm development (see EDGE #2 & EDGE #3).

In addition, transparency also known as Machine Learning interpretability is a key part of the toolset to tackle mistrust in our algorithmic decision-making processes. It can be used to promote fairness of the first and second type, and moreover pervade the organizational culture with ethical responsibility. As the great 20th-century physicist Richard Feynman put it: "if you cannot explain something in simple terms, you don't understand it". This maxima that is so accepted in the hard sciences, it is not that extended in Data Science. It implies a bidirectional association between explainability and understandability, which ultimately oppose transparency against blackbox-ness. It should be noted though, that black-box algorithms are not exclusively those of a non-linear nature; high dimensional and heavily tuned Generalized Linear Models can be also vastly opaque. Fortunately, interpretability frameworks such as LIME (EDGE #4) clear the way to take-apart the machine and explain its pieces.

A more in-depth and hands on explanation of these applied concepts is obtained from the following Show & Tell proposals. Enjoy the ride!

"transparency also known as Machine Learning interpretability is a key part of the toolset to tackle mistrust in our algorithmic decision-making processes."

```
        in range(FLAGS.num_episode
        self.logger.debug("       => Environm
        self.environment.__init__(actions_

        # Init environment
        state = self.environment.state.get_
            int(np.round(np.random.uniform(
            np.random.choice(self.environme
            np.random.choice(self.environme
        # Init DS to save epoch statistics
        segments_rejects_round = {}  segmen
        learning_rate = {}  learning_rate:
        rewards_round = []  rewards_round:
        bids_round = []  bids_round: []
        predicts_round = {}  predicts_round
        fairness = []  fairness: []

        i = 0
        while i < FLAGS.n_iterations:
            self.logger.disabled = False if
            self.logger.debug("
```

EDGE #2:
Interpretability and Trust in credit risk scoring

Contributor

Manuel Ventero
Risk & Fraud Analytics / Data Science

Manu holds a BsC in Telecommunications Engineering. Before joining BBVA he combined his work for the wholesale multinational services within a telecommunications enterprise, with a startup that he co-owned. Right after he decided to make a switch to the wonderful and dusty world of data. At the moment he works for the D&A Risk & Fraud team focused on disrupting SME online lending by creating innovative, fast and accurate credit scoring algorithms.

Contact:
manuel.ventero@bbvadata.com

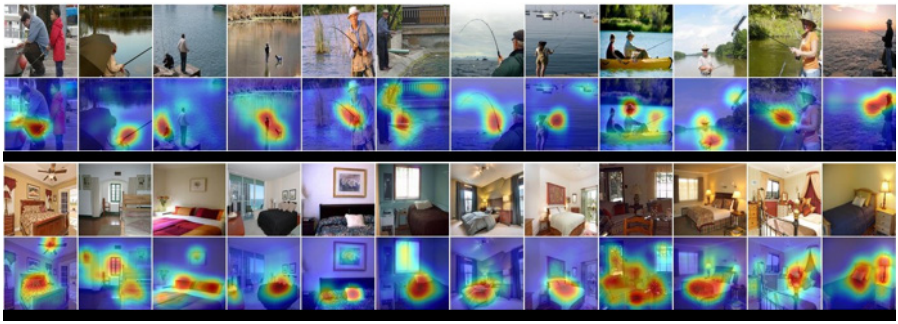


Fig. 1. Learning Deep Features for Discriminative Localization

Background

Next year due to **GDPR EU regulation**, users will have the right to request an explanation on an the algorithm's decision. Regardless of the intense debate on the matter, the form and vehicle of the explanations, highly regulated industries such as banking and insurance already have requirements on explanations.

In this talk we explored one way of making a black-box algorithm transparent using an interpretability framework called LIME

According to **Bank of Spain's 6th notice from June the 30th 2016**, an SME could request their bank to explain its rating as well as why has it been rejected. For this reason, we must explore algorithm transparency techniques that will allow BBVA to build, more complex and better performing models while ensuring transparency and trustworthiness.

Algorithmic transparency is openness about the purpose, structure and underlying actions of the algorithms used to search for, process and decision making. In this talk we explored one way of making a black-box algorithm transparent using an interpretability framework called **LIME** (University of Washington, August 2016),

and other approaches to model interpretability and understanding depending on their nature.

This set of solutions have been applied to Trust·u's non-linear risk model. **Trust·u** is a co-lending platform for newborn SMEs based on transactional data, and built around the concept of trust through social support. This model has been developed using an ensemble method called **Gradient Boosted Trees**. The model creates complex decision boundaries, where interpretability plays a key role in order to enhance trust for customers, regulators and analysts.

With some of the proposed approaches we aim to make risk models such as Trust·u's, trustworthy, as well as answering the what? why? and how? about Machine Learning interpretability and deep dive into LIME, pros and cons, and future work.

Analytical Framework

LIME methodology provides a framework for both local and global explanations.

Local explanations are implemented in their Open Source package, using this algorithm to find explanations. With these explanations, we could answer questions from Trust·u customers, such us, why have I been rejected? or why have I been approved? to explain an specific instance. We are responsible for choosing, the sampling size, N, the number of desired explanations K, as well as the distance metric (defaults to cosine distance for text, L2 for images or euclidean for tabular data).

In the other hand, LIME presents SP-LIME, their second algorithm. It allows us to provide a set of globally representative instances with explanations to address the "trusting the model" problem, via submodular optimization.

To solve the non-redundant examples, coverage is modeled as a function of W, an explanation matrix, I, denoting the global importance and V, is the set of global explanations.

$$c(V, W, I) = \sum_{j=1}^{d'} 1_{[\exists i \epsilon V : W_{ij} > 0]} I_j$$

That is, finding the set, that maximizes the weighted coverage function.

$$Pick(W, I) = \underset{V, |V| <= B}{\operatorname{argmax}} c(V, W, I)$$

“The model creates complex decision boundaries, where interpretability plays a key role in order to enhance trust for customers, regulators and analysts.”

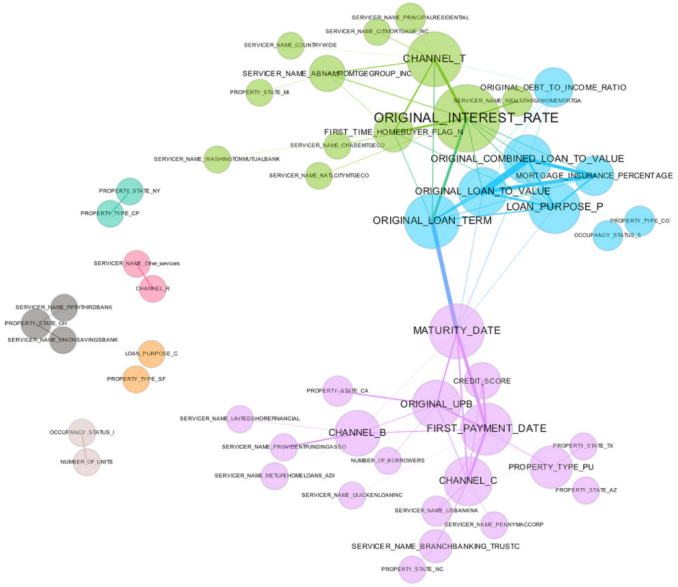


Fig. 2. Correlation graph. 2D representation of one-to-one relationships in data.

This second algorithm, is introduced in the paper, but hasn't been implemented in the OS package. However, Jordi Aranda (Data Scientist at BBVA Data & Analytics), has already implemented it.

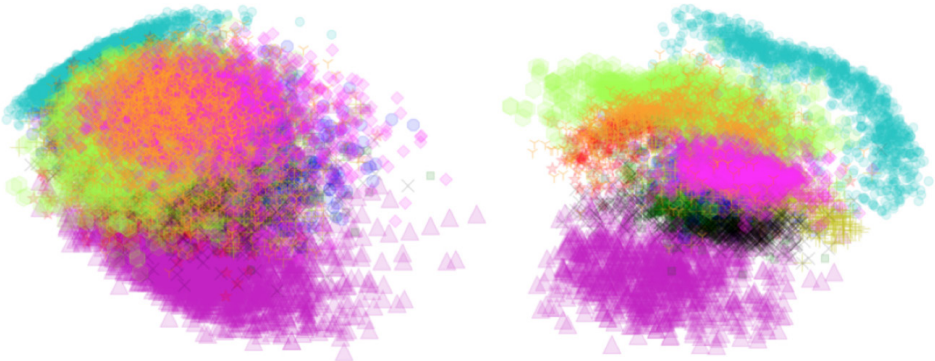
Results

We found out that LIME can help not only to fulfill with regulator demands, but also to give feedback to users on the decisions that affect them as well as helping scientists and analysts on the process of training and assessing models.

However due to its youth, the library lacks some details on its implementation, and sometimes because how it simplifies complex decision boundaries, it can lead to counterintuitive explanations. On this matter, we are developing our own interpretability framework, taking LIME as a starting point and benchmark.

Additional Resources

- Slide Deck



EDGE #3:
AlphaBid: Learning to bid in real time using a fair strategy

Contributor

Roberto Maestre
Payments / Data Science Coordination

Roberto holds a Ph.D. in Artificial Intelligence (UPM-UCM) in the field of "non-linear equation systems solving" applied to control aircraft and railway operations. MsC in Artificial intelligence and BsC in Computer Science. He has large experience at both academy and industry, in domains like Machine Learning, Logic and Engineering. He currently works developing and enhancing new models, tools and techniques related to Technical Pricing, Risk and Automatic Decision Making.

Contact:
roberto.maestre@bbvadata.com

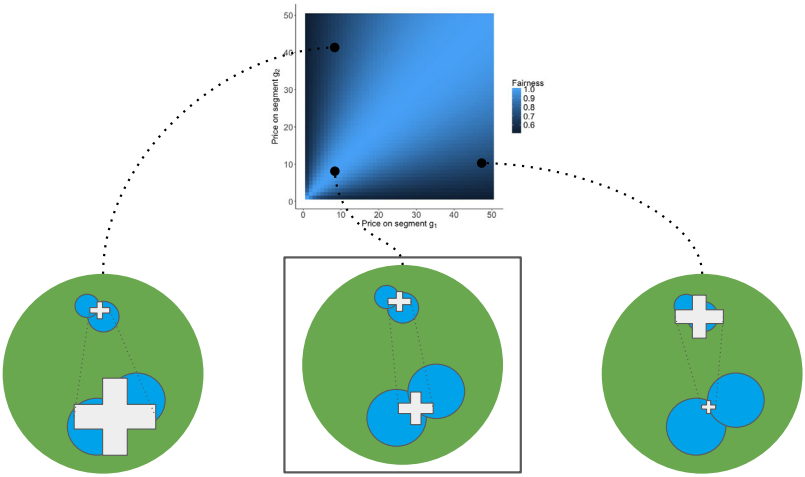


Fig. 1. Jain index as a fairness metric on two customer segments for three cases, respectively, overfitted match, perfect match and partial match (from left to right).

Background

Dynamic pricing models help companies maximize their revenues. However, traditional models may produce unfair price differences between groups of customers, and consequently negative perceptions about the pricing strategy. This negative perception can result in losses for a company in the long-term. This talk shows how to include a fair strategy into the pricing model, keeping the balance between revenue and fairness, using Reinforcement Learning (RL). We demonstrate that RL provides two main features supporting fairness in dynamic pricing: on the one hand RL is able to learn from the recent experience adapting the prices policy to complex market environments; on the other hand RL provides a trade off between short and long-term objectives, integrating fairness into the model's core. Considering

This talk shows how to include a fair strategy into the pricing model, keeping the balance between revenue and fairness, using Reinforcement Learning (RL)

these two features, we propose the application of RL in order to maximize the revenue integrating fairness and common sense in the long-term objective. Results on a simulated environment demonstrate a significant improvement in price fairness together with revenue maximization.

Analytical Framework and Data Sources

Q-learning is a RL algorithm introduced by Watkins. It provides a simple way for agents to learn, in dynamic environments, by trial and error. The heuristics in this model is directly related to the rewards provided by the environment in each iteration. Q-Learning is a well adapted method for learning in sequential decision problems. Thus it can be applied to dynamic pricing as each action performed by agent will modify the state of the environment (related to fairness) providing a reward (the bid itself). Next equation represents the generic way in which Q-learning is expressed:

$$Q(s,a) = Q(s,a) + \alpha\Big(r' + \gamma \arg\max_{a'} Q(s',a')\Big|s,a\Big)$$

where α is the learning rate; γ the discount factor and r' is the reward obtained after perform action a in state s. Thus, the goal of RL is to find a policy π that maximizes the expected discounted utility.

Advances in neural networks provide a variant of the Q-learning algorithm in which a neural network is trained by stochastic gradient descend. By minimizing the next loss function we can approximate the values of i.e. $Q(s,i;\theta) \approx Q(s,a)$:

$$L_i(\theta_i) = \left[\Big(y - Q(s,a;\theta_i)\Big)^2\right]$$

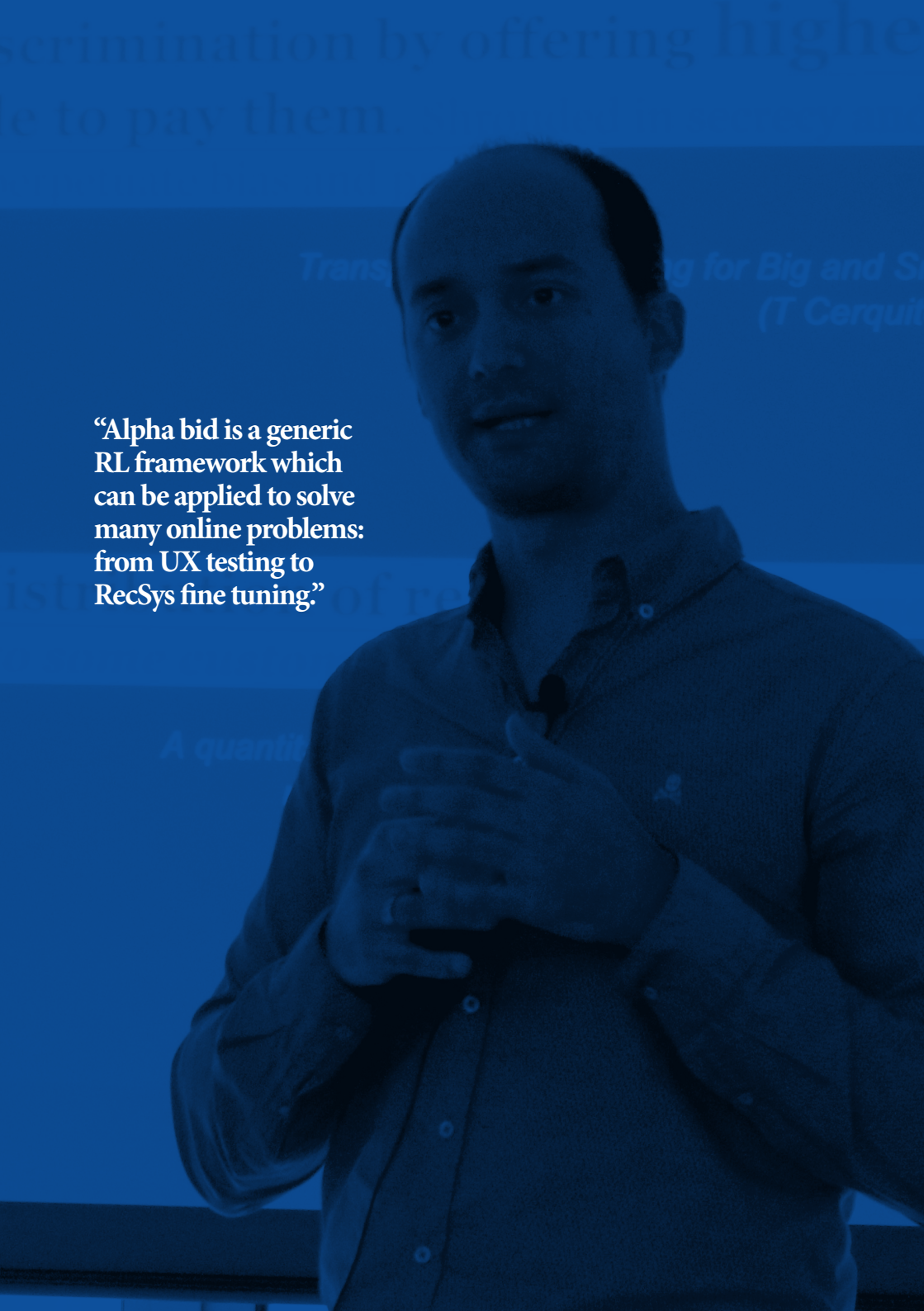
This framework integrates fairness metrics supporting fair policies.

Results

Main results point out the algorithm's convergence on a given fairness and bid target. Thus, we empirical prove that RL is a model capable of learning revenue maximization while the fairness metric provides a more egalitarian dynamic pricing strategy between groups of customers. Alpha bid is a generic RL framework which can be applied to solve many online problems: from UX testing to RecSys fine tuning.

Additional Resources

- Slide Deck


“Alpha bid is a generic RL framework which can be applied to solve many online problems: from UX testing to RecSys fine tuning.”

EDGE #4:
From niches to riches: A fair approach on Recommender Systems

Contributor

Juan Duque
Payments / Data Science

Juan is currently studying a Ph.D. in the topic of olfactory search at Universidad Politécnica de Madrid. In the past, he worked in a Project of vascular network hemodynamics in collaboration with the Hospital Universitario Puerta de Hierro and the Instituto de Matemática Interdisciplinar (UCM). Nowadays, he works in the area of recommender systems as a data scientist in BBVA D&A.

Contact:
juan.duque@bbvadata.com

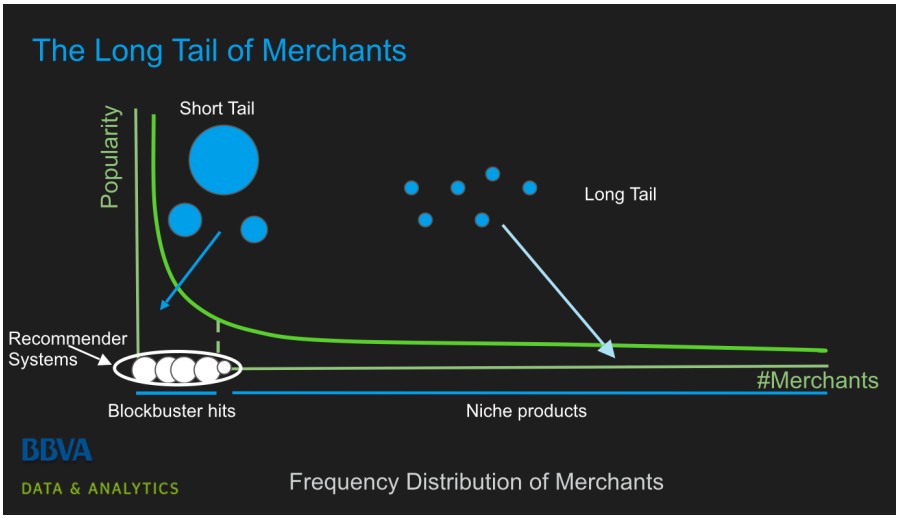


Fig. 1. Long tail frequency distribution of merchants signaling the difference between popular and niche items.

Background

This talk presents two techniques that allow keeping a balance between popular and niche products in the recommendation.

For the last two decades Recommender Systems (RS) have attracted the attention of industry, given her influence in the way people consume products. While RS aim to provide an appealing list of items to users, most algorithms suffer from a bias in the recommendation towards popular items. As a consequence, the recommended list often goes away from users' true interest. However, less popular, long-tail items are desirable for recommendations because of its novel and diverse character. Likewise, Long Tail products represent new chances for industry to increase its sales and profits by potentially selling in a vast and unexplored market.

Therefore, how can we overcome the "bigger was better" approach of RS and get a broader recommendation across the pop-

ularity distribution? In this talk I introduce the concept of fairness in recommender systems, so that all items have the same chance to be presented to users. To this end, this talk presents two techniques that allow keeping a balance between popular and niche products in the recommendation. The first one makes every product available in the recommendation, whilst the second distributes better the recommendation itself. As I will demonstrate, these contributions are critical to get more personalize, diverse and novel recommendations, while keeping relevance in the recommendation.

Analytical Framework

We propose a new objective function specifically conceived to deal with missing information in RS. As main characteristic, this loss function has a term that explicitly forbids a predicted zero preference when missing values occur.

Besides that, we introduce a popularity dependent scaling factor in the loss function which prevents the model from over-recommending popular items. Following[*], we take this factor as

$$s_i \propto \frac{1}{N_i^{\beta}}$$

where N_i is the popularity of item i.

Results

We show that fairness in RS can be achieved by using the two methods described previously improving several aspects of the recommendation itself. First, our new loss function is able to make recommendations for all items with a non-zero mean preference, solving the problem of "making everything available to the recommender" presented in the abstract. Second, our model outperforms previous ones in terms of relevance–oriented metrics, such as recall or mean averaged precision. Third, more frequently recommendations in the long tail of the popularity distribution are guaranteed by the popularity dependent scaling factor. In particular, we report improvements of **+17%** in the recommendation of Mid-Long Tail products.

Additional Resources

- Slide Deck

[*] *Item popularity and recommendation accuracy. Harald Steck. Proceeding RecSys '11. Pages 125-132.*

“While RS aim to provide an appealing list of items to users, most algorithms suffer from a bias in the recommendation towards popular items.”

Going Forward

Trust is a complex term with multiple dimensions investigated in psychology, sociology, economics, information systems and even philosophy. From a Machine Learning perspective, we realize to only grasp the tip of the iceberg. Technology can provide attributes to build trust like competence, quality, simplicity, and convenience. We have seen that Machine Learning techniques can help contribute to further experiences of trust like transparency and fairness.

What we have tried to communicate with this Show & Tell, is that models are not sanitized abstractions of reality; on the contrary, explicitly or not, they are being created with our biases and unfair judgments. These must not be seen solely as profit seeking machines, because the choices they made in the end are fundamentally moral.

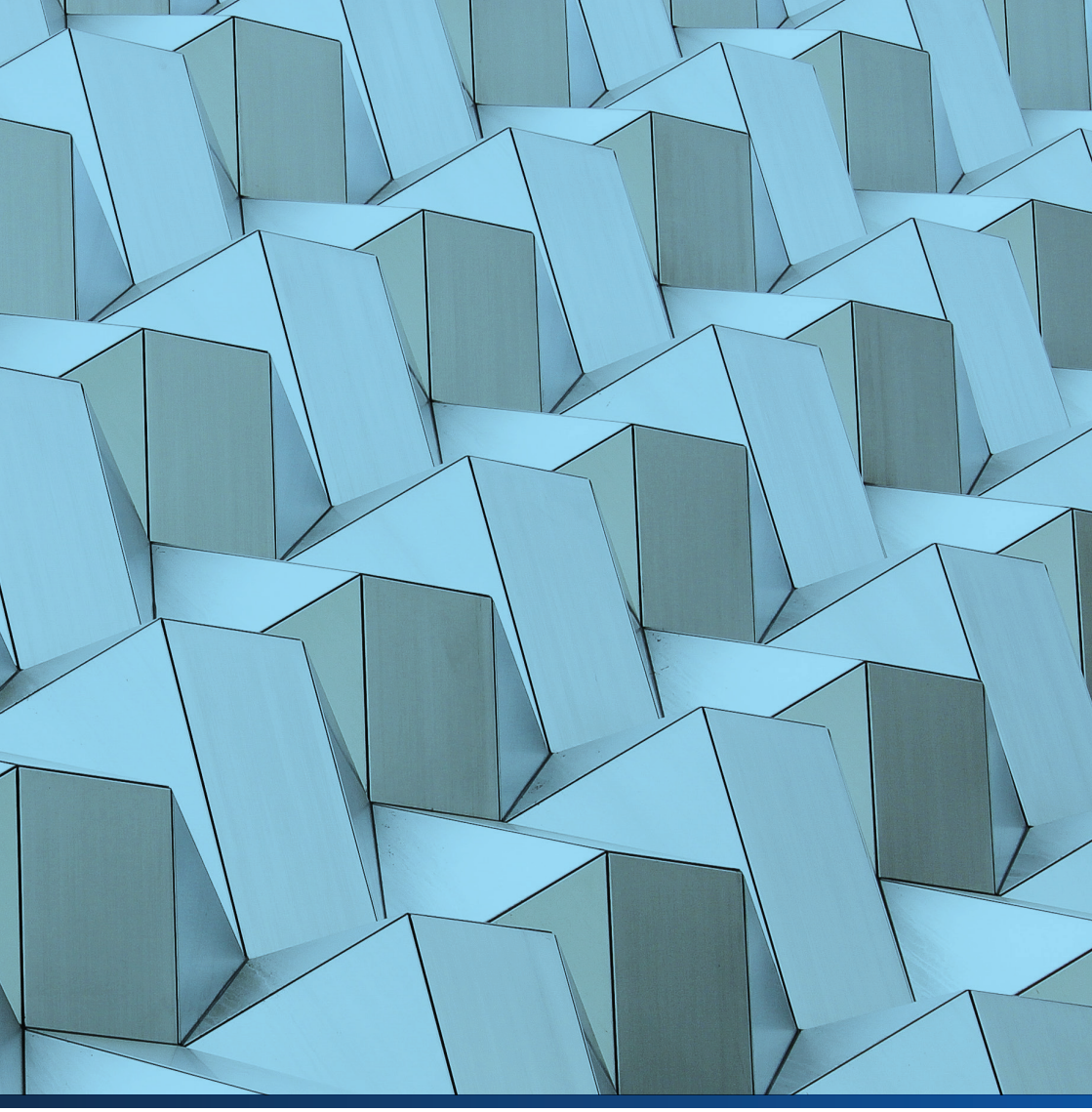
In addition, we have shown ways to include fairness and transparency as central elements of model development, that eventually will foster a trustworthy relation with our customers. From the different talks we have learned one way of making opaque algorithms transparent positioning us one step ahead of the new regulatory demand which comes into force next year under the EU General Data Protection Regulation (GDPR). Using model interpretability we can fulfill the regulatory "right to explanation" and give feedback to customers on the decisions that affect them, as well as help scientists and analysts on the process of training and assessing models. Regarding fairness, we proved with empirical evidence that Reinforcement Learning is a model capable of learning revenue maximization while providing a more egalitarian dynamic pricing strategy between groups of customers. Concerning recommender systems, we presented a way of effectively dealing with the extended bias in recommendation towards popular items. Avoiding this bias means new ways for industry to increase its sales and profits by potentially selling in a vast and unexplored market. In particular, we report improvements of +17% in the recommendation of Mid-Long Tail products.

Show & Tell

Shaping TRUST with machine learning

BBVA
DATA & ANALYTICS

October 2017



Shaping TRUST with machine learning

EDGE #1: Introduction to Trust and Machine Learning: Fairness & Interpretability
EDGE #2: Interpretability and Trust in credit risk scoring
EDGE #3: AlphaBid: Learning to bid in real time using a fair strategy
EDGE #4: From niches to riches: A fair approach on Recommender Systems