XONA PARTNERS

Data Sciences Focus

Mobile Eco-System Contributions

Dr. Riad Hartani, Dr. Alex Popescul,
Dr. James Shanahan

November 2013

Data Science Evolution – A Brief Revisit

As a team, we have first contributed to the field of artificial intelligence and overall data sciences since the early 90s, with early work in the area of machine learning, multivalued logic and neural networks. We then followed the evolution of these data analysis and knowledge discovery techniques over the years, as algorithms became more elaborate, computing models more efficient, and live data generated and collected at increasingly higher rates, often for completely novel applications.

Very recently, our focus has been on analyzing applications of recent techniques such as deep learning and the various additions to random forest and gradient boosted decision trees to practical industry problems. Over this 20 years lapse of time, one thing became clear to us: Data Science has made the big leap of being a research area for a select few applications, to a set of tools, accessible in various shapes and forms to various industry verticals, and optimized to resolve some of their more challenging problems. In this paper, we synthesize our experience on the experimental front through a recent case study, applying and customizing select advanced Data Science algorithms to a new set of Internet services applications.

Specifically, we focus our interest on real world case studies in the realm of mobile and cloud networks optimization, and corresponding business intelligence models running on top of these networks. In fact, up to today, commercially available data analytics products on the market have had the following shortcomings: (1) a limited scalability of for the data collection models as measured in terms of data generation, collection and storage (2) a lack of efficient machine learning and predictive modeling algorithms to process collected data in real time, (3) an open loop data analysis feedback, that is not dynamically correlated with the operator's business logic and (4) computing and pricing models that are based on centralized localized processing in operators' IT infrastructure, that is not taking advantage of the pay- as- you- go cloud- based computing.

Our stated goal over the past couple of years was to address the above shortcomings, taking advantage of technology innovations recently introduced, and just now getting to a sufficient level of maturity to be commercially applicable to mobile networks data analytics, taking into account operators' business logic goals in mind.

Data Science in a Mobile World – Why now?

To achieve the stated goals above, we opportunistically leveraged the fact that three fast converging concurrent trends. A brief snapshot is presented here.

First is the maturity of Data Management models.

We are witnessing the fast adoption of novel architecture to store and access large data sets (Hadoop, MapReduce, HDFS, Yarn, etc. – commonly known as Big Data models), as well the commercial availability of various cloud deployment architectures (OpenStack, vCloud, Cloudstack, AWS, etc.). This is removing significant logistical obstacles to embracing management of large data structures. The move is likely to be even more significant moving forward, given the immense number of contributions of the open source community in this area (as an example, we were part of a 3000 contingent at the last Openstack summit in Hong Kong – November 2013 -, the largest ever, which shows the strong interest of the computing community). Key here is convergence onto universally adopted platforms versus what was before seen as a proliferation of diverse platforms.

Second is the evolution of Data Sciences.

This applies to the large set of data analysis models in a broad sense, and specifically machine learning and mining algorithms that are more accurate and computationally tractable, leveraging distributed cloud-based computing models. Current developments in Deep Learning, for example, illustrate well how an older field of neural networks achieved breakthroughs in accuracy when its algorithm improvements were fueled by much increased computational power. Taking advantage of the introduction of new computing models, such as algorithms parallelization, GPUs and alike, then porting that to distributed cloud compute models, not only the existing algorithms have been optimized to run better and faster, but a number of additions and optimization have been developed and run in a computationally tractable way.

Third is Data Availability.

Leveraging software and hardware architectures that are increasingly scalable to selectively and dynamically process large volumes of data, relying on various models of data capture, via sensors, devices, and management modules. Larger data sets influence algorithm choices by easing the risks of over-fitting, which leads to better generalizable insights. The sheer size of data available is likely to increase, either as front-end data in real time or backend data stored as historical patterns. In the networking world specifically, hardware-based data-path architectures have evolved in a way that allows for data to be captured fast enough for deeper analysis, and software-based management architectures in a way that data can be queried, received and presented to relevant data processing models.

It is in fact, the first time ever, that such trends are coming together, which opens up the opportunity to leverage the vast amount of real time users and services data available for processing, through a correlation of to its underlying business processes, to optimize bottom line business logic, and dynamically derive new revenue streams and optimize existing modes of operations.

One specific real-life case study we have recently worked on with Tier 1 global mobile operators is briefly explored. It sits in the context of optimizing and monetizing their mobile data along select dimensions. Various similar case studies, in the areas of mobile data fraud and revenue assurance, public cloud migration enablement with underlying performance measurement and enforcement, as well mobile payment models optimization have been or are being worked on. This builds on very similar set of tools developed by the team over the last few years, in the world of digital and online advertising, web search optimization and related topics.

A Mobile Network Optimization Case Study

The Data Science solution we have worked on is inherently modular, and part of a more elaborate solution umbrella, composed of:

(1) A hybrid local/cloud based data gathering and storage, leveraging novel techniques optimized for the variety of data models. Adaptations of Hadoop-like models and their underlying MapReduce computing paradigm for large scale distributed file systems, are leveraged to present the various

data sets, that normally gathered in silos into a common data representation accessible to data processing models and

(2) A set of machine learning and data mining algorithms, specifically focused on clustering and predictive modeling in high dimensional spaces based on imprecise, uncertain and incomplete information, efficient statistical data summarization and features extraction algorithms as well as large scale real time data streams management. These tools will be at the core of the processing engine, and will aim at deriving optimization to the existing business logic and augment it with new revenue generating business logic, which would be mapped to a set of new revenue generating services.

3G and 4G networks are built over flat IP packet based networks. With the flexibility and scalability of IP based networks and services, comes the requirements for more stringent traffic and resources management mechanisms, and underlying challenges, unseen in previous circuit based switching technologies (for both user data where TDM circuits are replaced by IP / MPLS sessions) and control data (where SS7 is progressively replaced by SIP and Diameter IP based signaling).

The new architectures introduce various network elements in order to tackle such challenges. This would include data path processing models such as Deep Packet Inspection devices, used for marking and rate limiting traffic, to data compression/rating devices used for video optimization to topology and state aware control plane devices such as PCRF engines and SON resources load balancing engines among others.

In order to optimize customer user experiences (Quality of Experience (QoE): defined along various KPI metrics as perceived by the user), 3G/4G networks require the introduction of more sophisticated predictive, preventive and/or corrective resources management models in the networks. This is specifically where we have introduced novel data processing models, leveraging machine-learning algorithms, and demonstrated their value. As such, a real world traffic control scenario is developed, addressing a very specific problem that is causing major challenges in mobile networks today. The problem is formulated as follows: How to maximize the aggregated users QoE utility function over time, based on observation of real time and batch historical network level data measurements, and enacting semi real traffic control mechanisms in specific network enforcement points (either directly through dynamic provisioning or via a policy proxy function, such as a PCRF spell out for non-specialists).

This problem is instantiated via the following specific case study: QoE of the users is, in this case, modeled as the proportion of users traffic facing admission control rejection on setup (during the signaling phase between the mobile user device and the Radio Network Controller in 3G networks or equivalent in 4G networks between the user and the mobile network packet core), the network measurements are observed off the radio base stations either directly or through some level of aggregation, and the enforcement policies are based on pushing rate limiting decisions on the data-path, as well as other mechanisms focused on the video angle for transcode/transrate, etc..

The following assumptions have been made, for illustration purposes, but without impact when a more elaborate network model is taken into account (example: various dimensions are observed to determine congestion levels in different parts of the networks and different multi layer mechanisms are enacted to push policy enforcement decisions at various interfaces): data off the radio base stations are captured in various network conditions, and in various network locations, where no resources management model is enforced (besides the ones intrinsic to the radio access and core access intrinsic resources management models as used in a standard configuration). Data is modeled along an input / output dimension space, where the input shows a multidimensional aggregate packet setups entering the network over time, and the output showing the blockage levels over time. Machine learning algorithms (time series, neural networks, deep learning models explored) are trained on such data to model this function and provide an approximation of such function over time. The learning model would optimize the time horizons in the past over which the data is read (as input to the approximation function) and the time horizon in the future over which the function is being approximated.

Thresholds are defined where the network entry acceptance reaches some configured level would be identified as a threshold over which resources management mechanisms would need to be pushed down the packet core to reduce traffic volumes (either per user or an aggregate across users). The higher the projection of blockage levels into the future, the more aggressive the rate limiting would have to be.

Rate limiting functions would force Internet and private traffic (assumed to be a mix of TCP and UDP, with different rules applied to each) levels to drop by a well-defined function over some defined time interval (delayed in time versus the time where the policy action is pushed down). Assumptions are that this decrease will result in a step function reduction for a set of recommended policy actions. It is also assumed that such aggregate reductions in traffic would cause a slowdown in non-interactive traffic without affecting the interactive traffic, and hence marginally affecting the QoE utility function.

Based on the assumptions above, the machine learning models, coupled with the closed loop control feedback model would demonstrate the following:

The traffic projection is modeled with a sufficient accuracy, over time, leading to an appropriate approximation of entry into the network rejection levels. The model would run based on input data, and as soon as the projection shows a high level of rejection in the future time horizon, a control policy action is pushed onto the network. This control policy action would then force traffic levels down, and as such feed updated data into the prediction model. The overall system would run with this closed loop feedback and overall maximization of the utility function is proven, while the overall network stays in stable conditions.

Conclusions and Key Take-Aways

A brief description of some of the Data Science applications to mobile networks have been highlighted, as a way to demonstrate applicability and value of such techniques in the real world. Specifically, one demonstrates that existing vertical industries (mobile telecom world in this specific case), that have historically been fairly slow moving in terms of pushing new data analysis techniques, are starting to get disrupted. Disruption in this case is beneficial, as it will likely converge on making operations way more efficient, build a platform for new revenue generating services and push towards a new generation of players, taking full advantage of the potential of Data Science models.

Xona Partners team, with its diverse technology expertise in the Data Science space as well as select industry verticals, along with its global insight into new business models developing across the globe, has been working with select players, in a win-win model, to solve some of the leading multinational pain points – or allow them to develop an edge in what is, and will increasingly become, a highly competitive play, where winners take it all.

Xona Partners (Xona) is a boutique advisory services firm specialized in technology, media and telecommunications. Xona was founded in 2012 by a team of seasoned technologists and startup founders, managing directors in global ventures, and investment advisors. Drawing on its founders' cross functional expertise, Xona offers a unique multi-disciplinary integrative technology and investment advisory service to private equity and venture funds, technology corporations, as well as regulators and public sector organizations. We help our clients in pre-investment due diligence, post investment life-cycle management, and strategic technology management to develop new sources of revenue. The firm operates out of four regional hubs which include San Francisco, Paris, Dubai, and Singapore.

Xona Partners

www.xonapartners.com

advisors@xonapartners.com