




WHITE PAPER

Contextual Machine Learning: It's Classified

By: Seth Grimes, Alta Plana Corporation





Modern, cognitive computing – the application of adaptive machine learning to diverse real-world challenges – enables an advance from descriptive to predictive to prescriptive analytics. We expect systems to anticipate needs, round up and crunch relevant data, evaluate alternatives, and make targeted recommendations that help us reach our goals, whether they involve choosing a travel route, bringing a product to market, or enabling personalized delivery of precision medicine. We no longer need to be satisfied with static, after-the-fact pictures of what was. Today’s leading-edge solutions predict and suggest actions likely to lead to the outcomes we seek. But while the newer technologies are stunning, they are built on, and will continue to rely on, the power of established, proven classification and analysis methods such as taxonomy. This paper explains why and how.

Machine learning picking up speed

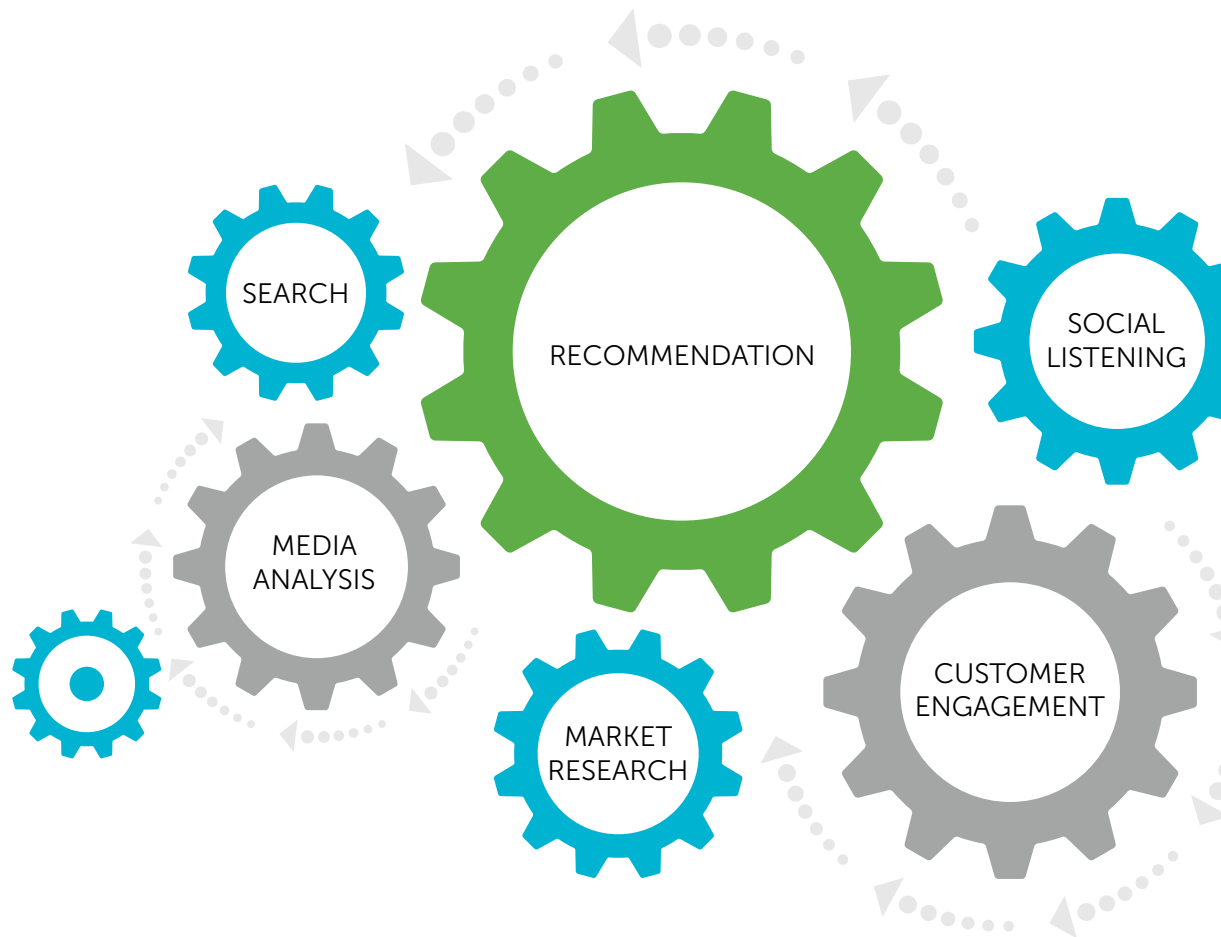
Technology advances lead to higher expectations, which motivate further innovation: A virtuous cycle.

Machine learning is this decade’s great technical advance. The algorithms discern patterns in source data and generate predictive models. The technology has existed for decades, but new sophistication, in the form of hierarchical deep learning, powered by low-cost, on-demand computing resources and fueled by a robust data economy, now make machine learning practical for everyday problems. Yet results remain highly reliant on the choice of inputs and algorithms.

Traditional approaches continue to out-perform machine learning for many of the most common tasks, including especially classification, that are at the heart of so many business processes and decisions. Leading analytics providers continue to apply traditional, high-precision, taxonomy-based-classification, for instance, for text and social analysis needs.

Taxonomy + ML for better predictions

Pattern detection and classification are at the heart of search, social listening, and customer engagement, as well as recommendation, media analysis, and market research. In each domain, application of human knowledge, captured for instance via taxonomy, helps deliver the most accurate and relevant insights. Outcomes are more favorable if human expertise trains models, tunes them via active learning, and evaluates and interprets the insights produced. Apply the co-joined technologies to model consumer behaviors and interests, to messaging, video, and voice data, to enhance interactions with virtual assistants. The classification advantage is unbounded.





There are many approaches that harness data and analytics to meet common business challenges. We define analytics as the systematic application of numerical and statistical methods to derive and deliver quantitative information. The power and complexity of approaches has grown, and will continue to grow, hand-in-hand with business (and personal) needs and expectations.

Needs and expectations have evolved beyond descriptive analytics: a first-generation analytics that is essentially a picture of the What of a situation. The questions, however, are still relevant:



Which of your company's Web pages were visited most frequently, and which sources drove the most traffic and revenue?



How have sales performance and profitability evolved year-on-year, measured monthly for principal product lines, in each region?



How many social mentions has your brand generated recently, and posting at what days and times of day generated the greatest social engagement?

These are important questions; the insights gained in answering them can help you optimize your business. Yet they merely describe. They don't explain, and they don't suggest best courses of action.

Enter predictive analytics

Enter predictive analytics, a discipline with two basic forms:

- 1 From Descriptive to Predictive to Prescriptive Numerical projections:
Past performance predicts future outcomes.
- 2 Classification: Category Y is a best fit for case/person/object X based on shared or similar qualities or characteristics.

Neither variety of predictive analytics is new, but the state of the science is constantly improving, driven by new methods such as deep learning, by the on-demand availability of inexpensive computing resources, by data culture, and by API-enabled application flexibility.

It's classification that's our central interest.

Consider common questions such as:

QUESTION	ANSWER
What are the key points in the Fitbit review posted on Amazon.com, and how did the writer feel about the various product features she mentioned?	Topic, feature, and sentiment extraction, and resolution and disambiguation of identified entities (people, products, laces, etc.), are significant classification challenges.
Given the items an individual views, can we recommend additional interesting content?	Classification can create "semantic signatures" of content, of single items and of collections.
Given the words and phrasing of a customer interaction, can we infer inclination to cancel service or just to seek a discount, or perhaps openness to an extension, upgrade, or add-on?	Classification can assign individuals to persona categories and pattern-match particular interactions to understand intent and to identify deception and fraud.

But here's where analytics gets really interesting, in the jump from predictive to prescriptive...

From predictive to prescriptive

Prescriptive analytics is about the path to a goal:

We know where we'd like to be.

Which actions – which decisions – will take us there?

Think of descriptive and predictive analytics as contributing steps. Take your best-fit predictive model and evaluate what-if scenarios to find the set of controllable conditions that promises to land you closest to your goal.

Prescriptive analytics isn't easy. Ability to execute fast, exhaustively, and accurately is key. Your modeling choices include machine learning and also traditional methods. The first excels at discerning emergent patterns in big data. Traditional methods, especially for the central classification challenge – where taxonomy, especially when constructed the variety of data domains, excels – provide reliability and high precision. Imagine the advantage that can be gained via a combined approach.



Given: Machine learning is this decade's great technical advance.

The technology aims to identify, detect, classify, and predict interesting features in source data, both text and structured datasets. The underlying process involves modeling, evaluation, and feedback/reinforcement, the latter making the method "cognitive," mimicking human learning. The hope is to improve on established methods, to achieve greater accuracy, robustness (model coverage and maintainability), and speed-to-production without sacrificing performance or ability to support the effect. (Availability and cost of data science talent is a significant concern.)

Some of the terminology is esoteric – words such as cognitive and reinforcement – but the concepts are relatively straightforward. In supervised machine learning, the software infers general decision rules – a predictive model – from training data. A human analyst annotates features of interest in a training set, choosing labels from a predefined set of type or categories. (Some organizations use crowd-sourcing for this labor-intensive task.) In unsupervised learning, by contrast, the machine makes a best guess as to the categories, grouping cases with similar characteristics. Feedback or other forms of reinforcement confirm or correct the machine's choices.



Despite advances, results remain highly reliant on the choice of inputs and algorithms. Model validation to ensure accurate results and reliable performance is an essential step.

Concerns aside, the case for machine learning is clear. The prime motivator is ability to flexibly generate purpose-suited models from data. The ingredients for adoption – low-cost, on-demand computing resources and lots of data – are in place. Steps to put machine learning in production, however, can get quite complicated.



Cognitive is complex; don't miss context

Consider IBM Watson, an example of a cognitive system. Watson feeds a knowledgebase by combining text-sourced data, extracted via text mining, with information from structured data sources. There's a curation processing involved: Humans assess, select, and correct acquired knowledge. The system interprets natural-language queries and generates candidate responses. The machine weighs possibilities and offers the answer most likely to respond to the question/ query.

What we have is, in essence, contextualized machine intelligence: a system generated by machine learning and context-focused via classification. The results speak for themselves: In 2011, a Watson computing system that could beat human Jeopardy champions. In 2014, Watson was made available on-demand, via IBM's Bluemix cloud, and more recently, specialized healthcare, for smarter cities, and for the spectrum of business challenges involving natural language.

Starting very recently, commodity machine learning from a variety of sources, often open source – from Google TensorFlow and Microsoft Azure Machine Learning to startups such as MonkeyLearn and MetaMind – has brought machine learning to the masses. Powerful tools in under-trained hands, however, will not produce best results. The contextualization we've discussed can be applied to improve outcomes, systematically, contributing at several stages to the accuracy, relevance, and usability of models and results, as we now discuss.

You seek to model diverse features, the features that matter for your business:



Not just terms, but also higher-level categories and detail-level components and attributes.

Fine-grained classification is high-precision classification, essential for highrelevancy search and recommendations.



Temporal – sequential – associations of terms, categories, elements, and attributes.

The ability to detect that, for example, people who search for maternity wear (in all the variations of that term) later search for a crib is the basis of predictive intent modeling.



Interests broken out by segments.

Can your technology profile young Latina women, aged 16-19, versus those aged 20-24, and differentiate the goods and services each segment purchases? Or easily interpret different seasonal buying habits of teenage boys living in San Francisco versus San Diego? One classification approach is to take a semantic fingerprint of visited and shared content, for purposes such as recommendation and ad matching.



Significance.

What points and patterns stand out, judging from a holistic understanding of consumer conversations across categories, and do those anomalies matter?

How can you improve machine learning accuracy with classified context?
Consider five ways:



Apply classification to create a high-relevancy training set.

If you train your model on data that isn't representative of the sources you'll use in production, your models will fail to deliver. Consider: Don't train a sentiment model on a set of movie reviews if you'll be analyzing Twitter reactions to automakers' announcements. You'll mix up Harrison Ford and a Ford Focus. Instead, draw only from sources that provide on-topic inputs, and apply contextual classification to ensure that each input is relevant. Looking for keyword hits, on "Ford," say, won't do the job. You need fine-grained classification to ensure accuracy.



Apply classification for automated, context-sensitive training-set preparation.

Annotation – labeling features of interest – for training-set preparation can be a labor-intensive process. In many cases, you'll need to hire subject-matter expert annotators. In other cases, you can crowd-source annotation although, due to quality concerns, crowd-sourcing requires careful management. Instead, consider applying linguistic resources to automate annotation.

Start with lexicons and gazetteers, which are lists of terms, names, places, and other entities. A thesaurus lists synonyms: a step more sophisticated but not enough to disambiguate a polysemous term, a term with multiple meanings. (Is Ford a carmaker, an actor, or a president?) You can apply lexical networks, which capture the words that frequently precede and follow a term of interest, and look at co-occurrence of other terms with a given term. Also consider contextual frequency of use whatever the domain. (If you're working with recent movie reviews, odds are that Ford will be Harrison rather than Henry).



Use the abstraction and detail – concepts and attributes – capture in taxonomy to boost annotation breadth and precision.

“Ford” belongs to the conceptual class (category) of vehicle manufacturers, along with Toyota, Fiat, GM, and others. Here, we’re climbing up a level of abstraction in our classification taxonomy. And Ford vehicle models include Focus, Mustang, and F-150... descending a level. In effect, use of taxonomy allows you to provide implied annotations, for instance, to label a car-model instance with a tag for manufacturer, even when the manufacturer’s name isn’t explicitly present in the training data.



Use classification resources to test machine-learning outputs, for model validation.

Model validation involves checking outputs against gold-standard results, which are typically produced by human evaluators. But just as automated methods provide for high-precision training-set annotation, they can provide for checks on outputs of models produced via machine learning.



Use classification resources in a reinforcement learning approach that will enhance your models and keep them current.

Currency is ensured by using output corrections to adaptively retrain the ML-produced model. And model enhancement: An example is use of taxonomy to associate entities and topics to annotate and make point-searchable a video based on words spoken in the video’s sound track.

One other potential accuracy booster we’ll mention: Use of an ensemble approach. Combine outputs of multiple methods – perhaps machine-learned and traditional – to arrive at a best consensus result.

These are some of the many ways to improve machine learning accuracy via classified context. Creative minds can surely come up with others. Where can these methods be applied?

Finally, we consider the question:

Who can make best use of contextual machine learning approaches?

We choose a few representative examples for purposes of illustration.


Brands, agencies & marketers study social status updates, consumer-generated reviews and forum postings, survey responses, e-mail and other customer contacts, and other, diverse insight

sources with requirements that include:

- Audience and market profiling and segmentation.
- Identification of behavioral signals that predict commercial activity (e.g., 'path to purchase').
- Social listening, customer engagement, sentiment analysis, and customer experience management.
- Competitive intelligence.

Social media platforms & online publishers serve constituents who include visitors and subscribers – who both consume and produce content – advertisers, syndicators and aggregators, and their own editorial and business needs. Their analytics-dependent tasks include:

- Data monetization, applying classification for topic tagging and filtering.
- Whom-to-follow recommendations.
- Ad matching, and content recommendation, based on the semantic signatures of the ad/content and of the visitor, based on content consumption.




Retail, manufacturing, and logistics deal with often-huge counts of product and service items and their categories, components, attributes, and specifications, as well as associated information describing usage scenarios, events, and sentiment. Analytics powers functions such as:

- Search keyword expansion to capture product categories and attributes.
- Self-service customer support, via semantic search that understands categories, components, and attributes.
- Product recommendation, matching product and visitor profiles.

But these are only examples. Really, the answer to our opening question, “Who can make best use of contextual machine learning approaches?” is:

Anyone with a lot of data – hence the applicability of machine learning – and with a real world problem where common-sense knowledge comes into play.



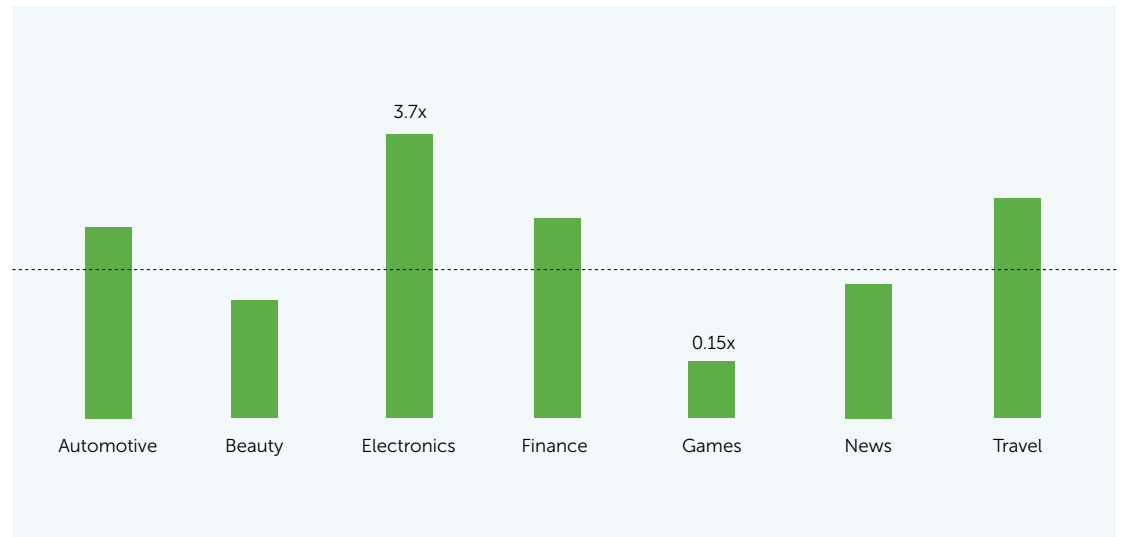
This paper was written for data scientists, software developers, marketing analysts, product managers, and the executives who work with them, crafting organizational data strategy. The assumption is that you and/or your colleagues have an aptitude for data wrangling and a degree of coding experience, whether for data analysis or product creation. That is, you have the facility necessary to work with the tools and techniques discussed in the paper. We assume that you're currently applying machine learning to pressing analytical tasks or have an initiative in the works. You're looking to maximize model performance – precision and results relevance in particular.

The choice of machine-learning methods is out of scope for this paper – although we'll offer the hints that a) recurrent neural networks offer best results for text and other sequence-dependent data and b) supervised methods, with models built from annotated training data remain quite popular, for good reason – so we'll focus on implementation of the ML-classification hybrid we've been describing.

Proof via prototyping

There is a degree of systems integration to be done, whether at the software level – via application programming interfaces – or by devising a processing pipeline where the output from one step is fed as input to the next. The focus should be on creating a repeatable process that will generate reproducible results with consistent performance. Given the plethora of cloud-deployed services and installable components available, and the possibility of scripting your own workflow for experimentation or for production deployment, there are few barriers to prototyping and development via an agile, iterative approach. Go for it!

Do prototype use of taxonomy. Focus on a detailed, holistic understanding of consumer conversations across categories to allow not only for right-level classification – by category, topic, brand, product, component, or attribute – but also providing indexation multiples – measures of in-category frequency expectations – that help you assess contextual significance.



Test on your own data, judging the correctness of results for yourself and evaluating the boost that contextualization, via classification tools, provides in test cases. You'll wish to assess classified context at multiple process points, as described in Section IV of this paper. Use of an on-demand processing service with a subscription model will allow you to make efficient use of resources and manage costs. Do ensure that the system not only meets accuracy and performance needs, providing analytical lift, but also that it has the capacity to scale to meet production needs.



This paper has described *classified* context, a technical approach that boosts the accuracy of models built via machine learning. Classified context improves training-data relevancy. The approach provides for rich, expanded training-data annotation and supports model validation and reinforcement learning.

The hybrid is contextual machine learning, analytical modeling for text-rich business applications drawing on social and other online media and a spectrum of enterprise information sources. Contextual machine learning makes the most of analytical advances, the data economy, and human expertise, as captured in traditional classification methods, notably taxonomy.

Prototype with your own data, using best-fit machine-learning algorithms, and experience the advantage for yourself.

About the Author

Seth Grimes consults on business applications of natural language processing, text analytics, sentiment analysis, and data visualization. He founded Alta Plana Corporation in 1997 and the Sentiment Analysis Symposium conference series in 2010. He co-organizes the LT-Accelerate conference in Brussels, a European forum for language-technology adoption. Follow him on Twitter at @SethGrimes.

About the Sponsor

This paper was sponsored by **eContext**, the world's largest text classification engine. Powered by a dynamically growing knowledge base of over 55 million curated rules and billions of documents and online consumer interactions, eContext can classify any text — including social posts, search, site content, and user messages — to any of its 450,000 topic categories. eContext's service accepts inputs from 35 languages, and can be used to predict user behavior, inject natural language intelligence into voice-activated assistants, classify video content, validate image recognition and structure training corpuses. eContext is owned by metasearch company Info.com.

USA office
180 N. Michigan Ave.
Suite 1400
Chicago, IL 60601
+1 (312) 477-7310

UK office
9 Belgrave Road
Westminster, London SW1V 1QB
United Kingdom
+44 (0)20 7834 5000



hello@econtext.ai
www.econtext.ai