Entrepreneur & Innovation Exchange
Published online at EIX.org on January 02 2015
DOI: 10.17919/X9F59N

How Entrepreneurs Can Use Data Analytics to Find Better Business Opportunities

James C Wetherbe (Texas Tech University)
Zhu (Drew) Zhang (Iowa State University)

KEYWORDS: Social Media & Networks, Information Technology, Research Methods.

Abstract

In formulating the opportunity identification process as a search problem, we show the power of marrying big data analytics and entrepreneurial insight to manage the huge hypothesis space in the hunt of winning opportunities.

Introduction

The key question all would-be entrepreneurs face is finding the business opportunity that is right and profitable. An opportunity is an idea for a new product or service. An opportunity is a sensed, rough match between an unmet need and a possible solution. In scientific lenses, with all the uncertainty that shadow the future, an opportunity is a hypothesis about value creation. Some opportunities ultimately become new products or services while others never come to life.

Ulrich and Eppinger (2012) divide the opportunity identification process into six steps as follows.

1. Establish a charter
2. Generate and sense many opportunities
3. Screen opportunities
4. Develop promising opportunities
5. Select exceptional opportunities
6. Reflect on the results and the process

In the eyes of an Artificial Intelligence (AI) researcher, this formulation bears a striking similarity with state space search, a common problem solving strategy. According to Russell and Norvig (2009), a well-formulated problem has the following components, mapped into the corresponding counterparts in opportunity identification:

1. Initial state -- rough product concept
2. Possible actions -- adding/removing feature-value combinations
3. Transition model -- actions leading to change of design
4. Goal test -- satisfying the charter?
5. Path cost -- development/implementation cost

Together, the initial state, actions, and transition model, define the "state space" of the problem. A "search" algorithms traverses the space and finds a sequence of actions that reaches the goal. In the opportunity identification context, the space is the hypothesis space or the design space, and the sequence of actions constitute a design ("solution") that is the winning opportunity. Figure 1 illustrates search happening in a product design space, where the process starts from a very rough initial idea which evolves into more concrete designs through feature additions.

The search formulation of problem solving, though conceptually simple, is practically challenging. The fundamental reason is the size of the state space: suppose there are d features to consider, and each feature takes b possible values, the total number of designs amounts to the order of magnitude b^d. Even if b and d are moderate numbers, this is a daunting space to manage. A classic example of combinatorial explosion, and a practical challenge in entrepreneurial activities. The extremely rich, effectively overwhelming amount of, information available in the world around us gives rise to a strange, yet real, situation for entrepreneurs: we are drowning in possibilities, yet we are starving for opportunities.

The lessons we have learned from decades of AI research is that exhaustive search is not feasible in general, and we need "smart" ways of exploring the hypothesis space. The general strategy, and the key idea, is to "inform" the search algorithm with some estimation of how "promising" a state/hypothesis is.

Copyright © 2015 The Authors. Entrepreneur & Innovation Exchange is published at EIX.org. This is an open access article under the terms of the Creative Commons Attribution-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited and no modifications or adaptations are made. View EIX.org Authorship Terms at https://eix.org/terms

Such estimation is called “heuristics”. In the rest of this article, without claiming completeness or optimality, we argue and demonstrate that the combination of data analytics and entrepreneurial insights serve the exact purpose of a search algorithm in an opportunity space. More specifically, we elaborate the following key components in the search process:

- Sketching the initial idea by understanding industry trends
- Growing the hypothesis space through the addition of features
- Managing the daunting size of the hypothesis space through heuristic pruning.

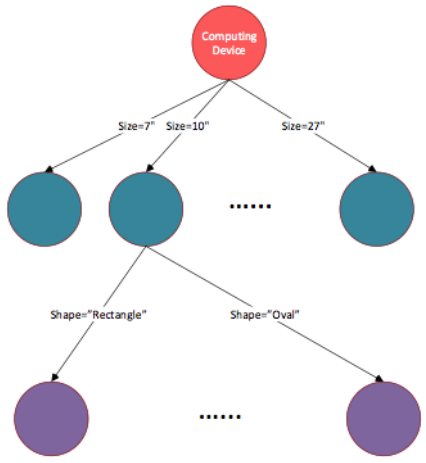


Figure 1 an Illustrative Example of Search in the Design Space

Industry Creation: Fools Rushing in?

The decision to commit to a particular market sector, existent or not, is a strategic entrepreneurial move that affects the long-term well-being of a business, startup or established. From the search perspective outlined above, it is analogous to picking a landing point for your starship in a vast universe of potential opportunities, which has a huge effect on the prospect of treasure hunt in the neighborhood.

While traditional ways of market study (e.g., interviews and surveys) are certainly still useful (yet expensive), we can increasingly rely on the power of big data to shed light on the strategic positioning of a business. The key driver of industry creation or involvement is sensing consumer needs. Incidentally, in formation sciences, it has been a long-held belief that what a person searches reveals her intent, which very often translates into a purchase need.

Google Trends is a powerful tool that can provide a quick birds-eye view of how some information need (potential consumption need) has been evolving and how it will evolve. Figure 2 illustrates the trends of four different product concepts, as measured by search volume over time. Clearly the public is more interested in commodities such as smartphones and energy drinks than luxury vehicles or science fiction, as indicated by the average volume of search. Google has different projections (represented by the dotted portions of the lines) for the two commodities, and a flying car seems more attractive than a real thing with a lot of drawbacks? Which trajectory do you want to ride as you look for intrapreneurial and entrepreneurial opportunities? Let both the history and forecast guide you. While intuitions can be very valuable when we have direct experience, they are very likely prone to errors in an unchartered territory. The trends are not necessarily scientific, but when quantified, provide good empirical grounds for something that we call “a hunch.”

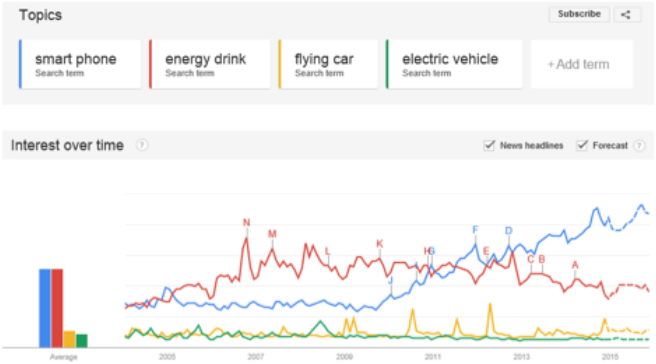


Figure 2 Search Volume as Proxy of Aggregate Consumer Interest

The potential opportunity may call for geography-related consideration, Google Trends also provides location-based break down (Figure 3). If the entrepreneur had a hunch, for example, about smart phone markets being less saturated in the developing countries, here is some data support for further expansion of the idea.



Figure 3 Location-based Breakdown of Search Volume for “smart phone”

Copyright © 2015 James C Wetherbe, Zhu (Drew) Zhang, Published by Entrepreneur & Innovation Exchange
EIX.org (2015)
DOI: 10.17919/X9F59N

For further brainstorming, Google Trends also provides volume statistics for related queries, and characterize their trends. See Figure 4.

Queries	Top	Rising
smart phone samsung	100	
smartphone	90	
best smart phone	80	
on smart phone	75	
smart phones	70	
android smart phone	65	
smart phone price	60	

Figure 4 Trending Statistics for Related Queries

Search engines do not represent the only type of platform that embodies public interest. Social media platforms such as Twitter lead us to a more viral view of the picture. Twitter is not only a place for marketing or political campaigns, but also a gold mine for business opportunites. While search behavior may be motivated by a broad spectrum of information needs, people's tweeting behavior is sentimental by nature, typically associated with feelings such as initiation, endorsement, or disagreement.

Twitter does an amazing job keeping its users stay on top of the "pulses" in the current time window, yet it does not lend itself to an easily usable tool for opportunity identification, the essense of which requires keeping track of statistics and seeing patterns through clouds. Until very recently, seeking intelligence similar to that provided by Google Trends has had to involve programming the Twitter API ("Application Programming Interface"), a non-trivial task that requires some non-trival computer science training.

Luckily enough, powerful services have been emerging that bridge this gap. For example, Topsy, a new social analytics engine, provides something very similar to Google Trends, based on Twitter historical data. See Figure 5.

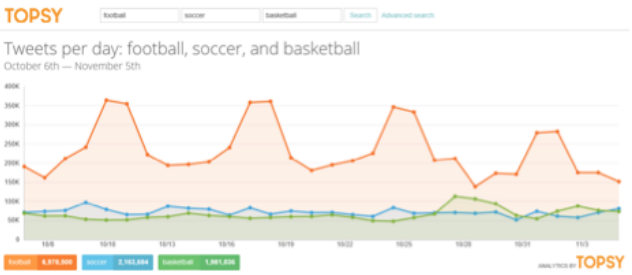


Figure 5 Tweets Trends (Source: topsy.com)

Where do we want to land our treasure hunting ship in the universe of possible opportunities? Some coordinates featuring a "hot" topic surrounded by a lot of positive sentiments, or maybe surrounded by a lot of controversy! While we all dream of our product being unanimously loved, Zhang et. al. (2012) show that sentiment/opinion divergence actually leads to good market performance for consumer electronics. Whichever you believe, Topsy, powered by Twitter data, again offers a nice lens (Figure 6).

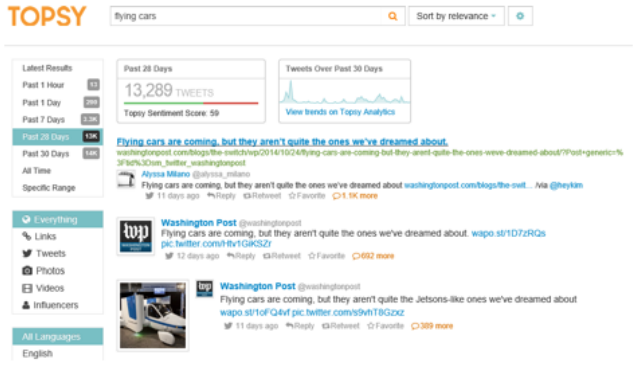


Figure 6 Tweets and Sentiments Surrounding a Topic (Source: topsy.com)

Idea Refinement: Growing the Opportunity Space

Running an effective opportunity tournaments (Ulrich and Eppinger, 2012) demands generating a large number of high-variance, high-quality candidate opportunities. Adopting better methods for generating opportunities and mining better sources of opportunities can increase the average quality of the opportunities under consideration, which will also increase the quality of the best ideas resulting from the tournament.

In the treasure hunt metaphor, now that the ship has landed in a promising land -- i.e., we have an initial concept -- next on the agenda is to scout the neighborhood, i.e., to fine-tune the idea and generate concrete designs to be examined. In light of Figure 1, we have roughly pinned down the root of the tree, and now we are ready to grow the tree by adding new

Copyright © 2015 James C Wetherbe, Zhu (Drew) Zhang, Published by Entrepreneur & Innovation Exchange

DOI: 10.17919/X9F59N

features to the initial concept. The question is: where do the features come from? Clearly a lot of domain knowledge is indispensable in this process. We demonstrate how such brainstorming can be empowered by social media analytics.

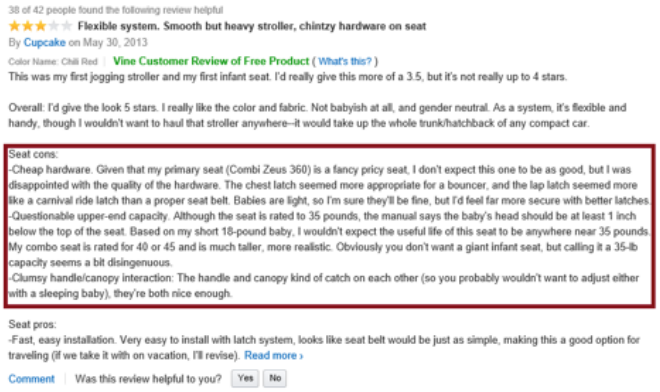
New product or service innovations almost always arise from unmet needs. In the old days, to find new business ideas, one would do this:

- Ask the group of people around himself/herself (friends, neighbors, relatives, social circles, etc.): What bothers you? What's inconvenient? What wastes your time or money? What makes no sense? What frustrates or angers you?
- After hearing the responses, work with people to come up with a better method or product, one that would solve the problem.

And yes, the entrepreneur can ask himself/herself the aforementioned questions! The new perspective is, enabled by Web 2.0 technologies, such questions can be answered at a much larger scale --- unmet needs arise from negative evaluation of existing products or services, and social media provide an unparalleled platform for consumers to share their product experiences and opinions, i.e., through word-of-mouth (WOM) or consumer reviews. There has been increasing amount of work in the marketing community to understand how WOM content and metrics thereof influence product sales and firm performance. We believe social media embody as much, if not more, value for intrapreneurial and entrepreneurial opportunities.

To be more specific, social media mostly contain consumer-generated textual data that describe and comment on the purchase and usage experience. Figure 7, for example, captures a consumer review of a baby travel system on Amazon.com. Text mining tools are available, and are still being developed, that are capable of finding the following:

- The locale of negative opinions, highlighted by the box
- The factors in the negative comments, e.g., "hardware", "upper-end capacity", and "handle-canopy interaction", which will naturally become candidates of features/considerations (in other words, branches in the search tree) for someone who's looking at the stroller market.


38 of 42 people found the following review helpful

★★★☆☆ **Flexible system. Smooth but heavy stroller, chintzy hardware on seat**

By Cupcake on May 30, 2013

Color Name: Chili Red | Vine Customer Review of Free Product (What's this?)

This was my first jogging stroller and my first infant seat. I'd really give this more of a 3.5, but it's not really up to 4 stars.

Overall: I'd give the look 5 stars. I really like the color and fabric. Not babyish at all, and gender neutral. As a system, it's flexible and handy, though I wouldn't want to haul that stroller anywhere--it would take up the whole trunk/hatchback of any compact car.

Seat cons:
-Cheap hardware. Given that my primary seat (Combi Zeus 360) is a fancy pricy seat, I don't expect this one to be as good, but I was disappointed with the quality of the hardware. The chest latch seemed more appropriate for a bouncer, and the lap latch seemed more like a carnival ride latch than a proper seat belt. Babies are light, so I'm sure they'll be fine, but I'd feel far more secure with better latches.
-Questionable upper-end capacity. Although the seat is rated to 35 pounds, the manual says the baby's head should be at least 1 inch below the top of the seat. Based on my short 18-pound baby, I wouldn't expect the useful life of this seat to be anywhere near 35 pounds. My combo seat is rated for 40 or 45 and is much taller, more realistic. Obviously you don't want a giant infant seat, but calling it a 35-lb capacity seems a bit disingenuous.
-Clumsy handle/canopy interaction: The handle and canopy kind of catch on each other (so you probably wouldn't want to adjust either with a sleeping baby), they're both nice enough.

Seat pros:
-Fast, easy installation. Very easy to install with latch system, looks like seat belt would be just as simple, making this a good option for traveling (if we take it with on vacation, I'll revise). Read more ›

Comment | Was this review helpful to you? Yes No

Figure 7 a Review of a Baby Travel System on Amazon.com

The rationale behind the approach above is the so-called "imitate, but better" strategy that has been discussed in the product development literature. When another player innovates successfully, it in effect publishes the location of a gold mine. A keen innovator can exploit this information by either considering alternative solutions that could address the same need or alternative needs that could be addressed with the same solution.

The "mining for negative comments" approach can be generalized to discover unmet needs in a competitive picture, instead of focusing on one target. Zhang et. al. (2013) discover from social media an implicit product comparison network underlying a market segment, where each product is situated in a competitive landscape (Figure 8). Notice that now the "imitate, but better" strategy can now be implemented in a much more holistic space like this, and every feature addition onto an idea-in-development is going to be based on much more informative understanding of consumer sentiments. More specifically, for example, one can deploy "crossover" and "mutation" operations in light of genetic metaphors.

Copyright © 2015 James C Wetherbe, Zhu (Drew) Zhang, Published by Entrepreneur & Innovation Exchange

DOI: 10.17919/X9F59N

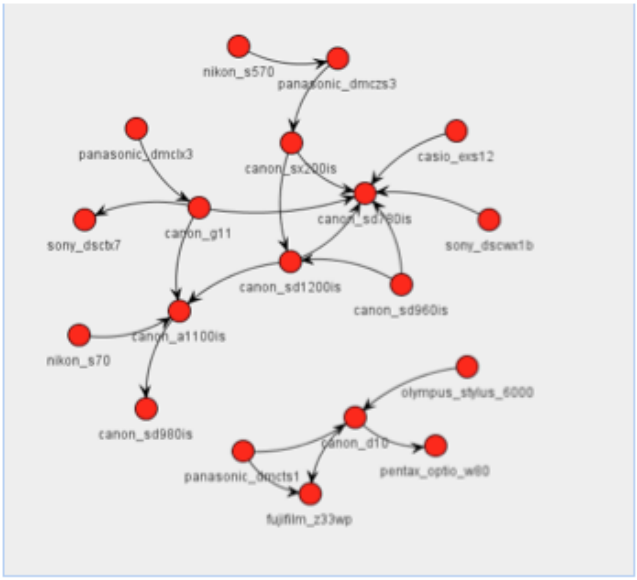


Figure 8 a Competitive Landscape in the Social Media

Each node in the network is a product; the directed links indicate consumer opinions on relative superiority. A link A->B represents a favorable opinion of B over A. Source: Zhang et. al. (2013).

Opportunity Screening: Pruning the Search Space

The search process illustrated in Figure 1 does not continue infinitely, even though there are still candidate features available for consideration. A path/branch may not be pursued due to:

- High barrier of entry
- Strong competition
- Prohibitive implementation cost
- Repetition of ideas or existence of substitute products (notice that adding new features onto an existing product based on consumer comments does not guarantee novelty.)

Computationally, it is also important to "prune" some branches in the tree instead of further growing them, in order to manage the daunting size of the search space that arises from combinatorial explosion of candidate features. While again domain knowledge plays an important role in the pruning process (e.g., estimation implementation cost), let's demonstrate how data analytics can come to aid, and more specifically, how Google Adwords keyword tool can help us understand several (though not all) of the pruning considerations above.

Google AdWords is an online advertising service that places advertising copy at the top, bottom, or beside, the organic search results Google displays for a particular search query. The choice and placement of the ads is based in part on Google's proprietary determination of the relevance of the search query to the advertising copy. As we all know, AdWords has evolved into Google's main source of revenue. In this context, we argue that it can be used as a tool to quickly approximate entry cost of product ideas. See Figure 9 for an example. Several components are worth noticing:

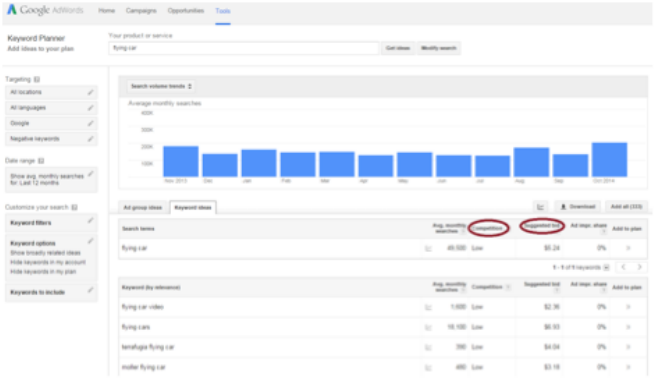


Figure 9 Google Adwords Keyword Planner

- The search volume for "flying cars" and related keywords are quantified over time, and as we discussed before, this measure can be used as a very rough proxy of consumer interest.
- Competition: based on the number of competing advertisers relative to all keywords across Google, the competition level for a keyword is labeled as "low", "medium", or "high." A measure of advertising competition in the cyber space, it is a good proxy for actual competition in the market space.
- Suggested bid: calculated by taking into account the costs-per-click (CPCs) that advertisers are paying for this keyword. The amount is only an estimate, and your actual cost-per-click may vary. Another way to look at it is that it is Google's proprietary estimation of other online advertisers' willingness-to-pay. Though a high dollar amount may be perceived as a cost, it can also indicate the potential lucrativeness of the potential market. For comparison purposes, "apple computer" is a high-competition keyword with a suggested bid of only $0.64, yet "flying cars" in our example is "worth" $5.24.
- The data view can be customized for specified location and language, a good tool to create a more "localized" understanding.

Copyright © 2015 James C Wetherbe, Zhu (Drew) Zhang, Published by Entrepreneur & Innovation Exchange

DOI: 10.17919/X9F59N

It is also helpful to gain further insight by examining the SERP (Search Engine Result Page) on Google. In Figure 10, for example, the search engine returns over 3.6 million pages for "flying cars" as organic search results, yet there is no paid ad! Here are the possible responses:

- Pessimists say: the lack of paid advertisers suggests that this query word is not lucrative and that the flying cars is a tough market to make money in. At 3.6 million competing pages, i.e., possible product/service providers, it will be extremely hard to launch a new product/service/website and rank above the fold on Google SERP for the term "flying cars." Hence it is not worth the time to further develop the idea.
- Optimists say: looks like most of the 3.6 million pages are information sources. The fact that there is no paid ad indicates that there is no mature business available yet in this market. Let me occupy the territory by developing an exciting product and become a leader in the industry!

Which characterizes you? A question to be answered by reality and further research, of course. The key in this section, though, is that data analytic engines such as Google Adwords keyword planner play the role of "heuristic functions" (Russell and Norvig, 2009) in the search problem formalized at the beginning of this article. Based on the demonstration above, one can conceivably develop a cost estimation heuristic for a potential opportunity based on some function of search volume, CPC pricing, and keyword competitiveness, as well as other domain-knowledge-based variables. Heuristics are estimations, so they do not have to be perfect. If we remember the lessons learned in AI, a good heuristic function that leads to completeness and optimality should be "admissible." What does that mean? It should never over-estimate the cost. What does that really mean in English? Well, it says, we should be optimistic in life.

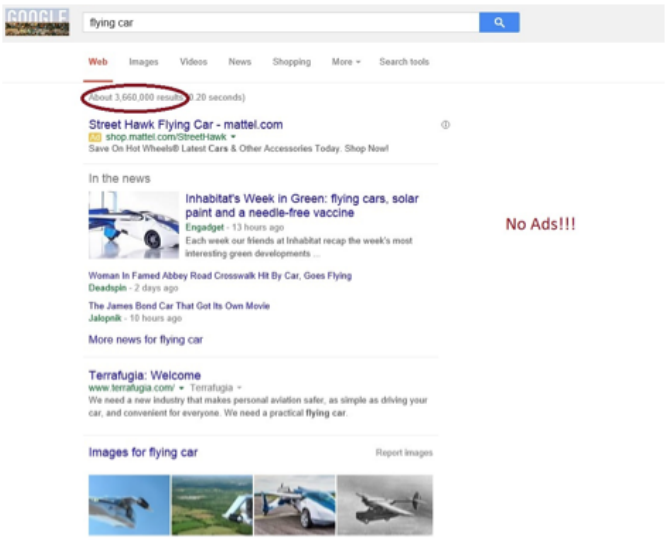


Figure 10 SERP for "Flying Cars"

Entrepreneurial Insights? More Indispensable than ever!

The techniques discussed above are specific, but not mechanic enough to be fully automatable (i.e., programmable). As tempting as it seems, we may never get there. Why? In our discussion of industry creation, the "seed" ideas have to come from human insights or intuitions; when growing and pruning the hypothesis space, human judgments are critical in the assessment of fuzzy situations. Think of all the tools as an embodiment of HAL; it is still David Bowman's call to land his ship.

There exist different types of opportunities (Terwiesch and Ulrich, 2009). In Figure 11, the techniques described in this article exhibit strength in identifying horizon 1 and 2 opportunities, but clearly fall short on horizon 3 ones. The data analytic algorithms are as good as robots scouting a neighborhood for gold nuggets, yet the "teleport" capability in the idea space is inevitably human.

Copyright © 2015 James C Wetherbe, Zhu (Drew) Zhang, Published by Entrepreneur & Innovation Exchange

DOI: 10.17919/X9F59N

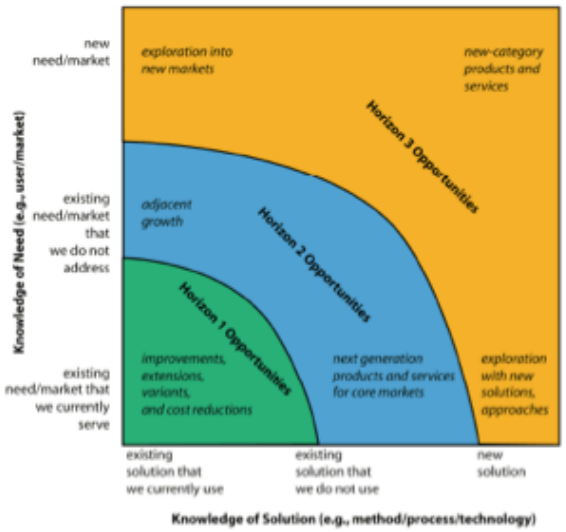


Figure 11 Types of opportunities

Horizons 1, 2, and 3 represent increasing levels of risk, reflecting different types of uncertainty. Source: Terwiesch and Ulrich (2009).

More importantly, when big data and analytics tools are more and more commoditized, the playground is more and more leveled. Expert insights and intuitions are more critical for success than ever. Big data analytics makes it easier and cheaper to test concepts, yet big money has to arise from big ideas. The bad news is, few of us are like Steve Jobs, who was daring enough to believe that customers did not know what they wanted until he showed it to them. Though a much needed virtue, gut is more often unreliable than not. In his seminal work, Tetlock (2005) describes a twenty-year study in which 284 experts in many fields, including government officials, professors, journalists, and others, were asked to make 28,000 predictions about the future, finding that they were only slightly more accurate than chance, and worse than basic computer algorithms.

Looks like we are in a very unfortunate paradox here: data are great at quantification, but terrible at imagination; analytics are less transcending, human intuition is less reliable. Yet another instantiation of the classical Maravec's Paradox (Maravec, 1988). What do we do?

After decades of battle between human and machine intelligence in the game of chess, here is a story told by Gary Kasparov, the human chess master:

In 2005, the online chess-playing site Playchess.com hosted what it called a "freestyle" chess tournament in which anyone could compete in teams with other players or computers. Normally, "anti-cheating" algorithms are employed by online sites to prevent, or at least discourage, players from cheating with computer assistance. (I wonder if these detection algorithms, which employ diagnostic analysis of moves and calculate probabilities, are any less "intelligent" than the playing programs they detect.)

Lured by the substantial prize money, several groups of strong grandmasters working with several computers at the same time entered the competition. At first, the results seemed predictable. The teams of human plus machine dominated even the strongest computers. The chess machine Hydra, which is a chess-specific supercomputer like Deep Blue, was no match for a strong human player using a relatively weak laptop. Human strategic guidance combined with the tactical acuity of a computer was overwhelming.

The surprise came at the conclusion of the event. The winner was revealed to be not a grandmaster with a state-of-the-art PC but a pair of amateur American chess players using three computers at the same time. Their skill at manipulating and "coaching" their computers to look very deeply into positions effectively counteracted the superior chess understanding of their grandmaster opponents and the greater computational power of other participants. Weak human + machine + better process was superior to a strong computer alone and, more remarkably, superior to a strong human + machine + inferior process.

So, do we have the verdict?

Conclusions

The techniques illustrated in this article, as well as the co-evolution history of human and machine intelligence, suggests a secret formula for successful opportunity identification in today's information-rich business world:

Entrepreneurial insights + big data analytics = winning opportunity

What makes our story different from chess playing is: the search process for opportunities has an infinite time horizon by nature, while a chess game has a clearly defined end. In a fast-changing business world, what is today's winning solution may be worth less than a dime tomorrow. Following, or even in parallel to, the exploitation of a champion opportunity, is a new episode

Copyright © 2015 James C Wetherbe, Zhu (Drew) Zhang, Published by Entrepreneur & Innovation Exchange

DOI: 10.17919/X9F59N

of exploration.

References

Moravec, H. (1988) Mind Children. Boston: Harvard University Press.

Russell, S. J. and Norvig, P. (2009) Artificial Intelligence: A Modern Approach (3rd ed.), Upper Saddle River, New Jersey: Prentice Hall.

Terwiesch, C. and Ulrich, K.T. (2009) Innovation Tournaments – Creating and Selecting Exceptional Opportunities. Boston: Harvard Business Press.

Tetlock, P.E. (2005) Expert Political Judgment: How Good is it? How Can we Know? Princeton: Princeton University Press.

Ulrich, K.T. and Eppinger, S.D. (2012) Product Design and Development. Fifth Edition. New York: McGraw-Hill.

Zhang, Z., Li, X., and Chen, Y. (2012) Deciphering word-of-mouth in social media: Text-based metrics of consumer reviews. ACM Trans. Management Inf. Syst. 3(1): 5

Zhang, Z., Guo, C., and Góes, P. (2013) Product Comparison Networks for Competitive Analysis of Online Word-of-Mouth. ACM Trans. Management Inf. Syst. 3(4): 20

Copyright © 2015 James C Wetherbe, Zhu (Drew) Zhang, Published by Entrepreneur & Innovation Exchange

DOI: 10.17919/X9F59N