Making Sense:
Extracting Meaning from Text by Matching Entities and Terms to Ontologies and Concepts

DRAFT Chapter 2
How the Brain Solves Tough Problems

Alianna J. Maren
Northwestern University School of Professional Studies

Chapter Draft: 2015-12-31

In this chapter, we will identify three broad architectural strategies that the mammalian brains use to solve tough problems. We'll then see how we can apply them to the challenge of text analytics.

2.1 Three Key Brain Architecture Principles

In the previous chapter, we identified ***three major neurological architectural principles*** that can give us useful insights and guidance:

- **Representation levels** - processing occurs in distinct steps, each one contributing to a new representation level for the resulting data/ knowledge/information,
- **Topographic maps** - useful for identifying data similarities and differences, and are also essential in sensor/data fusion, and
- **Statistical thermodynamics** - modeling coordinated behaviors observed in large-scale collections of simple processing units.

This chapter provides a bit more insight into how each of these guiding principles has been useful for brain processing, how researchers have adapted it for use in a related computational problem (typically vision), and how it applies to text analytics.

Our understanding of how to apply insights from brain-based processing will help us determine how to get greatest value from our effort in building a text analytics system.

2.2 Computer Vision: A Corollary Story

One of the biggest breakthroughs in computer vision came about as researchers discovered that effective computer vision needed to be framed in terms of *representation levels*. The lowest computer vision processing levels were devoted to statistically-based algorithms addressing very small portions of the pixelated computer image. These algorithms found local edges (differences in pixel intensity), determined textures, and found gradients.

At the next processing level, small edge pieces were joined – and missing edges were filled in – so that object boundary edges could be found. Operating in a parallel endeavor, other algorithms characterized regions; size, shape, compactness, texture, intensity and other factors describing small regions. Together, these edge-detection and region-characterization algorithms yielded local image features.

Some three decades ago, computer vision research was temporarily stalled at this level. Researchers were trying to identify objects using algorithms that, more-or-less, could provide local features but not much more. They were attempting to combine the resulting feature lists with object descriptions that resulted from projecting how 3-D object models would appear when viewed in a 2-D image. This turned out to be enormously difficult.

The reason was that the stretch between low-level features and high-level object descriptions was too far for image interpretation algorithms to handle.

The breakthrough came about as various researchers identified the need for an intermediate representation level.

This was a big step, because the researchers – up until this time – had come from one of two major camps. On the one hand, there were specialists in low-level image processing algorithms; everything that could contribute to edge detection and region characterization. The efforts of these researchers were augmented with corollary findings by brain researchers who specialized in the low-level processes of the visual cortex. In fact, several major algorithmic breakthroughs happened when image processing specialists adopted ideas directly from cortical architectures.

On the other hand, the specialists in 3-D object modeling were very akin to cognitive scientists. These researchers worked typically with symbolic logic, and often attempted to emulate the thought processes (as they were understood) as might be done in the mammalian brain. These researchers modeled objects and relations between them, such as "the red cylinder is on top of the blue block."

What was missing was a connection between the low-level signal processing algorithms and the high-level symbolic-based object representations.

Within the computer vision community, the breakthroughs combined two major aspects. First, a group of researchers identified certain "missing" representation levels; ones that were between the levels representing the edges and regions in a 2-D image and the levels representing abstract understandings of 3-D objects and how they would appear as projected into a 2-D plane.

Second, a related group identified a set of perceptual processes that occurred in biological vision; ones that were intermediate between low-level image processing and high-level symbolic manipulation. These processes, which hearkened back to the early days of Gestalt psychology, also drew from insights into pattern recognition as it occurred in biological systems, and as it could be encoded into algorithms.

This was a key step in allowing computer vision breakthroughs to happen.

Of course, many other steps were needed as well. The particular insights about representation levels occurred in the mid-1980's, and we are just now having trustworthy autonomous vehicles with fully-functioning computer vision. Clearly, a lot of additional development has been necessary.

Nevertheless, this key insight has shaped computer vision, and is highly relevant to other complex processing challenges – including text analytics.

2.3 Multiple Processing Stages and Multiple Representation Levels

Just as computer vision requires multiple processing levels, so does text analytics.

To accomplish effective text analytics, we need to do the low-level statistical processing algorithms. One of the features that differentiates how we develop statistics-level text analytics algorithms vis-a-vis similar low-level processing for computer vision is that we must take into account how different terms (typically nouns and noun phrases) occur across entire sets (corpora) of documents.

Thus, a major distinction between low-level text analytics algorithms and low-level computer vision (and also brain-based vision) is that text analytics looks at entire corpora early in the game, whereas vision processing can focus on just one image at a time.

That said, at the lowest levels, we do preprocessing, feature (term) extraction, and statistics.

At the highest level, we create a world-view that expresses our understanding of what physical and abstract things comprise our universe, and identify the relations between them.

Our ultimate goal is to match texts with a set of known objects in our world-view, identifying those objects and relationships that are expressed in each text, as well as grouping together texts that have similar objects and relationships.

Thus, for both computer (and biological) vision, as well as for text analytics, our highest representation levels are very abstract.

In between the low-level statistical processing and our high-level object representations, we need transitional mechanisms. These are so important that we devote the entire next chapter to describing text analytics representation levels and transitions in greater detail.

This subject also comprises the majority of the chapters in this book.

With that said, we briefly turn out attention to two other aspects of brain architecture that are useful two us: topographic maps and statistical thermodynamics.

2.4 Topographic Maps

One of the biggest clues that a brain architecture principle might be important is when that architecture has components that go to great and extraordinary lengths to make the principle work in practice.

A primary example of this is how the brain uses topographic maps to represent data.

2.4.1 Strange Things That Happen in the Brain

As a quick aside, the notion of topographic mapping is a brain-architecture principle that we will use on our data after it has gone through several processing levels. The primary way in which we'll use this principle is to create order from large text corpora.

We can use one of two different algorithm types here; one clusters related text items, and the other distributes these items over some pre-defined "topographic space" (which is usually 2-D) so that similar items are next to similar ones. Both of these algorithm classes are useful. They have slightly different purposes, and it will be up to us to decide which we would prefer to implement.

With this in mind, we turn to the brain, and notice that it does *topographic mapping* of certain input data.

The brain creates topographic maps of visual data. That is, something that is perceived by one small area of the retina activates brain cells that are neighboring to those that are activated by a neighboring retinal region.

This may be so obvious that it seems almost odd to make a point of this organizational principle, but bear with me for a moment.

The brain's tactile representation system similarly is represented with topographic mapping. Neurons from one finger activate brain cells that are near to neurons on the next finger, and so on. Again, this is entirely what we would expect.

What is perhaps a bit more subtle and interesting is that this tactile representation has the same kind of organization as that used for vision; neurons carrying stimulus from the "front and center" of our bodies (mapping roughly to the center of our visual field) are central in the topographic representation and get a much higher mapping precision (just as with vision). Those neurons that represent our tactile sides and back get a much lower degree of precision, just as peripheral vision gets less neural processing (and precision) than the processing given to foveal-centric stimulus.

Turning our attention to auditory stimulus, we find (once again) that the brain's representation of auditory stimulus forms a topographic map.

What is particularly interesting here is that when the brain creates topographic maps for visual and tactile processes, the architecture seems very logical and straightforward. In contrast, the brain architecture in mapping auditory stimulus to a comparable 2-D space is extraordinarily complex.

The auditory neurons process sound frequency information, and to create a 2-D spatial representation map, the timing (and also the frequency differentials) between the signals arriving in the two auditory-input processing areas (the cochlea) must be interpreted in a very specific way.

Creating a spatial representation of auditory stimulus is complex. Yet the brain invests in this effort. Clearly, there is survival value.

One value - an obvious one - is being able to spatially-locate the source of sound production. This alone would be a strong argument in favor of developing an auditory 2-D topographic map.

However, there is another reason for this architecture; one that has only been recognized for the past two decades – and which is gaining more and more attention from biologists.

2.4.2 Sensor Fusion - An Essential Survival Skill

"What I tell you three times is true."

The Hunting of the Snark, by Lewis Carroll

"Two out of three ain't bad."

Bat Out of Hell, performed by American musician Meat Loaf, lyrics written by Jim Steinman.

Sensor fusion is a survival skill. Sensor fusion is what lets us correlate information coming in from multiple sources (visual, tactile, auditory) about the same thing. This is a powerful clue that helps us put the information together to create a meaningful mental construct.

Sensor fusion does more than just help us correlate information, though.

Sensor fusion helps us cut through the chaff of too much stimulus, and isolate and extract stimuli that may actually be important.

Think about this for a moment. Most of the brain's neurons are inhibitory; they are calm-down neurons. Substantially fewer are excitatory; the pay-attention-to-me types of neurons.

We – and all living creatures – are constantly bombarded with stimulus. The important task is not to react to incoming stimulus, it is to determine that which is *worth* reaction.

Consider how this works for a cat who is out on her morning hunt.

When a cat is hunting, she may ignore a little flicker of movement in the grasses. She may ignore a little rustling sound. However, when the flicker of movement happens at about the same time and in about the same location from when and where the sound is detected, she is immediately alerted.

Her cognitive mind is not involved in this decision, at least not at first. She doesn't have to be thinking, "Im hungry, is there a mouse nearby?"

Instead, the sensor fusion area called the *superior colliculus* triggers her response automatically. Her head turns and her eyes foveate; they center on the source of the movement. Her ears pivot towards the sound source. The combined inputs from two sensor modalities (visual and auditory), in about the same timeframe and coming from approximately the same location, are enough to fully grab and direct her attention.

Only after she has immediately responded to the combined sight and sound, and assessed what that combination might be, can she then decide whether or not to pursue the potential target.

This is an important survival skill, since the natural world is full of both sights and sounds, and there are a huge number of largely-random triggers. If the cat or any organism (including ourselves) responded equally intensely to all stimuli, we would all forever be distracted. Instead, we are able to preferentially respond only when stimuli combinations override our innate "calm-down-and-ignore" mechanisms. (Neurologically, we have far more inhibitory connections that cause our brains to calm down than we have excitatory connections, which get our attention focused on a new stimulus.)

There are many biological factors that make sensor fusion possible, and recent research has shown that sensor fusion is even more widespread far more commonly down throughout the brain than was thought to be the case some decades ago.

One architectural component that makes sensor fusion work, though, is the notion of topographic maps.

Topographic maps are a key aspect of low-level visual processing. This makes enormously good sense from an architectural point of view; it lets us literally "connect-the-dots." Simply put, things that are near each other in what would be a 2-D image of our visual observation space will trigger neighboring cones and rods in the retina, and these then trigger neighboring neurons in the input layer of the visual cortex. This forms a topographic map representing the input stimulus.

This topographic map projects upward through the visual cortex layers, so that very simple connections at the lowest levels (similarities or dissimilarities of brightness and texture) will generate larger-scale features at the upper visual processing levels. All throughout the processing levels, though, the notion of a topographic map persists.

This is easy enough to understand in terms of the visual cortex. What is very interesting, and also remarkably compelling, is that our auditory system and our tactile sensation system each also project to analogous topographic maps. In fact, the auditory and tactile systems project to maps that have the same configuration, more-or-less, as does the visual system.

This is a *tour de force* of neurological architecture.

While we can envision that tactile sensations, from all over our 3-D bodies, might easily be projected to a 2-D map that is aligned with the visual map, it takes far more neurological complexity to project auditory sensations to the same topographic configuration. This is because auditory processes begin with frequencies of sounds heard, and the projection requires a delicate and precise combination of frequency and time differentials in order to project the source into a 2-D map that can represents the space around the body.

The immense amount of neural complexity involved in creating this projection ability attests to the great value in having the inputs from three very different sensory systems visual, tactile, and auditory mapped into the same processing arena, with close overlaps of the stimulus source.

Earlier, biologically-based sensor fusion inspired various algorithms that helped computational sensor fusion. The earliest applications were to various military interests. Now, we can readily envision their relevance to various commercial applications, such as automata-driven cars and other vehicles.

Combining, for example, visual and IR sensors can help guide such vehicles under various conditions of weather and illumination. It is because the biological system finds such great value in topographic maps that we are encouraged to see how this kind of data representation can help us with text analytics. There is a class of algorithms that has already been widely used in creating topographic maps; we will consider how they can form robust and valuable tools for interpreting texts in several different ways.

2.5 Statistical Thermodynamics in Text Analytics and in the Brain

In addition to drawing inspiration from brain research, we will also call upon one other area; statistical thermodynamics. This is the branch of physics that deals with how very large collections of very simple objects behave.

Up until the past decade, this connection might have seem far-fetched. Now, however, the collections that we consider are so vast that it is not that unreasonable to step very far back and treat them as collections of very simple objects.

In statistical thermodynamics, we typically model collections of objects that can be in one of two states; think of this as "on" and "off," or "black" and "white," or simply as states "A" and "B." The point is that only two states are allowed for each particle or unit in this statistical thermodynamics world.

Sometimes, we apply some sort of external force to the system, and something happens. For example, in the statistical mechanics (related to thermodynamics, not quite the same) of magnetic materials, we can have substances that behave a certain way only when an external magnetic force is applied.

Specifically, we can have a substance that under normal conditions contains an equal mixture of units having spins of either "up" or "down." When the external magnetic field is applied, though, the overwhelming majority of these units line up with the field, creating a system whose units are hugely in just one of the two allowable states.

By corollary, we can imagine that a query acts as a sort of "external magnetic field." It isn't, of course, but we can envision it having some sort of effect on the units in a text data corpora.

Prior to the query, there is no real differentiation between units. Once the query is applied, though, the units are either in state “A” (having relevance to the query) or in state “B” (not being relevant).

Now, there is something very interesting about statistical thermodynamics systems. We can model them with equations, of course; that is the entire purpose of having these models — we get certain kinds of predictable and describable behaviors.

More than that, though, depending on just a couple of parameters (one of them being the strength of the applied field or stimulus), we can control how the system responds. We can choose parameters so that the system goes into an “all-or-nothing” response; when the external force (or query) is applied, they are either almost all go into state “A” or almost all go into state “B.” There are also a lot of behaviors in between.

This means that we can get a range of behaviors from the corpus as a whole in response to queries, if we set it up right — analogous to a statistical thermodynamics system of magnetizable particles.

The more we can create and work with this analogy, the more we can generate certain predictable and even controllable behaviors from our data corpus.

This, by itself, has potential value.

There are two more reasons, though, to think about using statistical thermodynamics applications to modeling text corpora. The first of these reasons deals with dynamic properties, especially phase transitions.

Phase transitions are what happens when a system changes state. When liquid water turns to ice, that is a phase transition. When a previously unorganized set of magnetizable units suddenly take on the same spin directions (and becomes magnetized, or having its own magnetic field), that is also a phase transition.

There are all sorts of interesting and subtle things about phase transitions that can be of value to us, especially when we think of them as being the outcome of a dynamic process.

This is an interesting – and largely new – thought for us. Most of the time, we have been thinking about static data corpora, or data corpora that are updated from time to time with new batches of data.

However, with better text analytics processes becoming available, we are more and more interested in detecting change. We are interested in novelty detection (new things), sentiment change (human responses to certain things), and even changes in volumes of text on one subject or another.

Methods that can give us a handle on the dynamics of large-scale systems

– potentially even large text corpora – are thus of great value, because they can help us to model change.

The final reason that we have an emerging interest in statistical thermodynamics of data corpora actually relates to both brain behaviors and the behaviors of large, complex systems.

This discussion is brief – the goal here is simply to justify introducing a realm of physics to text analytics and mining – but bears just a little elaboration.

In the brain, there are occasions in which a few neurons start firing, and they induce other neurons to fire as well. This in itself is not at all unusual; neurons in the brain do this all the time. What is unusual, in this case, is that the neurons which are induced to fire are very far away from the initial firing neurons. Not only does this specialized set of far-from-each-other neurons fire together, they keep it up. They have coherent firing patterns.

Similarly, physicists identify something that they call "long-range correlations" in substances just before they undergo a phase change. That means that certain of these simple units start acting coherently, before the big drama of a phase transition takes place.

We can think of these behaviors as "early warning indicators." Moreover, these behaviors (as they show up in physical and biological systems) are exactly the kinds of behavior that we may seek to discover in data corpora. This will especially be the case if we are looking for advance detection of emerging new events.

For this reason, one of the last chapters in this book provides (for the adventurous and intrepid reader) a brief overview of statistical thermodynamics as it can potentially relate to analyzing text data corpora. It is just enough to establish some vocabulary and ideas, and to provide a jumping-off point for those who will later follow up on this line of thinking.

Bibliography