

**DARPA Tech, DARPA's 25<sup>th</sup> Systems and Technology Symposium  
August 9, 2007  
Anaheim, California  
Teleprompter Script for Dr. Kendra Moore, Program Manager,  
Information Exploitation Office**

Patterns, Patterns, Everywhere...

» **KENDRA MOORE:**

Patterns.

I am passionate about patterns.

I am passionate about learning patterns from complex data and using those patterns to improve and maintain our information advantage.

As we develop new technology to automatically process large amounts of heterogeneous data,

I want to exploit the resulting information to learn patterns of behavior.

I want to learn our enemies' normal operating modes and detect when they change modes.

I want to determine when a specific target, be it a traditional military threat, an insurgent, or a suspected terrorist, is deployed in a new way.

I want to use the information we collect from sensors on our own forces – on both our vehicles and our soldiers – to learn our forces' normal patterns of movement and use the knowledge of those patterns to detect when they've run into a problem or to determine when our own patterns have made us vulnerable.

Now, you might ask: Are there patterns out there?

Well, of course there are.

Every individual has his or her own patterns of behavior.

Every organization has its patterns of behavior.

And, yes, even insurgents and terrorists have patterns of behavior.

Their actions are *not* random.

They train, they plan,  
and they develop their own tactics, techniques, and procedures.

These patterns may change over time,  
but they do exist.

Our challenge is to learn our adversaries' patterns and exploit them –  
so that we can interrupt their actions and quickly detect when they are  
changing their patterns of behavior, so that we can stay ahead of them.

So, where do we find these patterns?

The simple answer is that we can find these patterns in the data.

Ah, but there's the rub... there are a LOT of data *and* a lot of different  
data sources to process, analyze, and exploit.

My IXO colleagues have told you about the new data they want to  
collect and analyze.

In addition to imagery and video, we have all of our other traditional

military data: radar, signals intelligence (SIGINT), human intelligence (HUMINT), and so on.

And as my colleagues have pointed out, our ability to collect and store these data exceeds our current capability to thoroughly process and exploit it.

But, that's just the tip of the data iceberg.

In the future, we can expect our adversaries to operate amidst the civilian population and communicate via broadcast media and the Internet.

As a result, we need to exploit many other data sources to learn their patterns of behavior.

These sources include: documents, ranging from pocket litter to terrorist training manuals; radio and television news and other broadcasts; Internet sites and blogs; and general economic and political information.

All of these data sources can be used to inform us of our adversaries' intentions, activities, and their patterns... if we can only exploit them.

Our challenge is to exploit all of this data... all of this rich, dynamic, multi-dimensional, ever-changing, and uncertain data ... even as it is generated at increasing speeds, in multiple formats, in multiple languages, and at multiple locations... or at no location at all, in cyberspace.

So, where are we today on the challenges of learning patterns in this sea of data?

In recent years, we've made great strides in learning and using patterns in single types of data.

We can extract entities, topics, meaning, and semantics from text.

We can convert the spoken word to written text and recognize voices using aural patterns.

We can automatically detect and recognize objects such as buildings, equipment, and military units in images, and perform change detection on those images.

We can automatically detect and track moving objects in video.

And, we can track blog topics, and conduct network and trend analyses.

While work still remains to fully exploit these individual data sources, especially to handle the volume of data, we've made significant progress and overall, we're doing a pretty good job.

That said, today's exploitation techniques have some significant shortfalls when it comes to exploiting our adversaries' patterns of behavior: they have been developed and applied to single types of data, they assume fixed patterns, and they assume complete or nearly complete data.

The general approach has been to define or learn a vector of information that serves as an example pattern and then compare new data to that vector.

If the new data falls close enough to the exemplar, then we declare that it is an instance of the pattern.

But, if the instance is only partially visible or if the pattern has evolved, today's approaches fail... even though a human analyst could likely succeed.

For example, today's image recognition algorithms can find a tank in a picture;

however, if the tank is 75% obscured or there's a new type of tank that hasn't been seen before, the algorithms won't detect it.

On the other hand,

a human could detect the tank with partial information or use knowledge of the pattern to infer that a new type of tank exists.

IXO is beginning to address these shortfalls.

In the Predictive Analysis for Naval Deployment Activities program, we are exploiting multi-source data including vessel tracks, weather data, and shipping news, to automatically learn models of normal behavior and detect deviations.

One of the goals of this program is to continuously learn patterns, detect when a vessel deviates from its normal behavior, determine if that change represents a threat or other problem, and if not, update the model for that vessel.

This program represents a leap forward in pattern learning and change

detection in complex, multi-source data, for a single application domain, the maritime domain.

This single application domain provides us with a specific grounding or context in which to develop the technology.

But, this is just a beginning.

We need to go much farther.

We know that our domains of interest will evolve.

We know that the data sources we need to exploit will evolve.

Our focus will shift from a single geographic area, a single set of activities, a group of individuals, or a political/economic group, to a global focus, and back.

We know that the threats will evolve, merge, and fragment.

We need technology that can rapidly adapt to these changes.

In the past, we relied on analysts to learn patterns, detect instances of them, and detect when those patterns evolved...  
to identify when something unexpected or different happened.

But, in the past, we had a smaller set of threats and a smaller set of data sources to exploit.

In today's world, we must leverage technology to better monitor all possible threats,

learn their patterns,  
and detect when those patterns change.

We cannot continue to rely on people to manually review all of the information, looking for unique indicators or things that just don't make sense.

There's just too much information for humans to manually exploit.

We need to develop vastly improved information exploitation capabilities that free our analysts and warfighters from processing data so that they can focus on evaluating potential threats.

Imagine the following scenario:

US troops are deployed to a war-torn country supported by the United States.

There is a well-known network of roads between the airport and a US base.

Over time, we learn the patterns that vehicles follow and tactical intel analysts are regularly alerted on anomalies.

Across town,  
soldiers automatically collect data at multiple checkpoints and upload it to sound, image, and  
text databases.

That data is currently analyzed using a network analysis package.

At another base in a neighboring country, analysts look for patterns in foreign travel by known bad actors.

Today, anomalies are investigated manually or dismissed by analysts who are de-sensitized by a high level of false alarms.

The analysts typically use subjective mental models that they have developed over time and through extensive experience; however, they are likely to operate in isolation from one another.

They often lack the opportunity to jointly analyze and compare their data and mental models.

I think we can do better.

I think we can develop technology that will allow our warfighters to better assess the situation, and detect and interdict threats.

I envision technology that will allow an agile, dispersed team to handle all of the data on vehicle movement, networks, imagery, HUMINT, and more.

The system would work alongside the analysts, developing dynamic, comprehensive situational awareness models that understand normalcy across many domains.

Analysts could use the technology to delve deeply into a problem space as necessary, and otherwise let it alert them to anomalies as they arise.

Let's imagine that these models identify a new attribute among the vehicles traveling on the road network.

The models show that some trucks over a certain size slow considerably just out of sight of a checkpoint before continuing their journey.

Soldiers at the checkpoint notice that some of the passengers are more nervous than usual.

An analyst in the neighboring country reports a higher-than-normal travel rate among members of a certain group, whose leader is known to have connections to suspected insurgent groups.

I would like to see technology that can recognize that a pattern is changing and can look for other patterns that are changing.

I would like to see technology that can recognize that a new pattern is emerging and look for related changes or impacts and alert analysts to investigate.

For the scenario I just described, this technology would examine these seemingly unrelated events and alert an analyst that there may be a link between them.

The analyst examines the alert and requests imagery and a physical inspection.

A patrol then discovers that the trucks are dropping off bad actors into a trap door that leads to an extensive underground network.

With this information, the warfighters are able to thwart a previously unknown attack on our forces or allies before it occurs.

So, what do we need to do to support this vision?

We need to develop several new capabilities.

I want to see improvements in dynamic representation discovery.

We need the ability to automatically identify new features within and across multiple data streams.

This means we need lightweight, flexible, and rapidly adaptive representation schemes and mechanisms for managing them.

I want to develop a capability to automatically learn patterns of behavior in all-source data *and* recognize when those patterns have changed.

I envision algorithms that can incrementally learn patterns of behavior across multiple data streams.

These algorithms need to learn online in unsupervised and semi-supervised modes, with feedback loops to detect and manage model drift, and automatically learn what additional data sources are needed to resolve anomalies.

In addition, I want to develop techniques that recognize when the *domain* itself has changed and when *new* domains emerge that should be examined.

Effective models of domains and information should be lightweight and adaptable in order to respond to a wide variety of situations and questions.

I see a system that compares many dynamic models simultaneously, to ensure that the data and surrounding context is fully exploited.

In this way, we ensure that the system won't fruitlessly try to fit new data into an old model when the situation has changed too dramatically to be represented as was once thought.

I see a new generation of human-machine interactions to ensure that users are getting everything possible out of these advanced systems.

Users need to see enough of the underlying data to understand why a model formed the way it has and why a particular model was chosen for a particular problem.

Users need to understand the models' alerts.

Users also need the ability to investigate real world explanations of the attributes discovered by the system.

We need new information exploitation visualizations to support the wide array of information flows, modeling approaches, and domain analyses that are needed.

I'd like to hear your ideas for how we can achieve any or all of this vision.

Data is the past.

Information is the present.

Information exploitation is the future.

We need *your* help to turn that information into the knowledge our analysts and decision-makers can use.

Come join us in the future of information exploitation.

And now I'd like to introduce you to the bear who will bring us to a roaring conclusion,

Mark Davis,

our Deputy Director.