Student questions: Kiri Wagstaff colloquium on "Tell Me Why: Interpretable Machine Learning Discoveries for Earth and Space Exploration"

October 31, 2018

Question 1: Why are so many of the different methods like svd so bad, and how are they consistently worse than random guessing?

In this experiment, a regular SVD was unable to find new classes within the data set more quickly than random selection did. This tells us that the representation of the data that was given to the SVD (in this case, the raw pixel values) did not provide enough information for the image classes (like rover wheel, scoop, brush) to be separated.

Question 2: How long does 1 cycle take to run to suggest the next image?

Great question! This depends on the speed of the machine used to do the analysis. On my laptop, it takes a few seconds per suggestion for a data set with 7000 images. This time scales with the size of the data set and the number of features used to represent the images.

Question 1: How often does DEMUD return false positives for "interesting" images such as the image of blue background with no bug?

By definition of the class discovery problem, each time DEMUD selects an image that contains an object from an already discovered class, it is a false positive. This can be measured as the inverse of the class discovery rate. It varies by problem and often by class – some are more heterogenous, so DEMUD may discover different sub-types of that class which differ in appearance (e.g., color, texture, shape). This is where the definition of "class" gets a bit subjective; one person might divide the insects by their species, while another person would divide them by individual variations or pose.

Question 2: Can you give DEMUD too much information where it become unable to identify novelty on account of having seen too much?

I do not think this is a problem. "Novelty" by definition is something that is different from what has been seen before.

Question 1: Does changing the order of the training images that you show DEMUD change how quickly/efficiently it detects anomalies?

No. It takes in images as a batch and iteratively selects the most interesting image, the next most interesting image, etc. so it is not affected by the order of images within the data set.

Question 2: Do you need to know how many classes there are in the dataset before running it?

No; this is something that you can discover by the process of running DEMUD. You can run DEMUD repeatedly until you are tired of looking at selections ☺ Each time, it will pick the next most interesting item. However, DEMUD cannot on its own determine when it has found all possible classes in the data set – human interpretation is required to make that decision.

Question 1: Since SIFT is also used in structure from motion techniques, could the use of neural networks in SfM be in the future as well?

Structure from motion is not an area I work in, so I don't have a good answer for you there. I encourage you to investigate further, e.g. using Google Scholar, to see if others have tried this out. Perhaps it is something you could develop yourself!

Question 2: Could DEMUD be used to track very slight changes in morphology, potentially as a way to track very slight stages in evolution rather than entirely different species?

Yes, if DEMUD were given a data set that contained images of only one species, it would make distinctions based on finer differences, down to individual variations.

Question 1: How does DEMUD go about creating the image of what it already knows based on the image it's given?

DEMUD operates on feature vectors that represent the content of each item (image). It uses an image "inversion" or "visualization" technique to convert the image feature vector into a 2D image we can see, and it uses the same method to convert the SVD reconstruction ("what it already knows") into a 2D image as well.

Question 2: How quick is the DEMUD algorithm able to process data sets?

On my laptop, it takes a few seconds per selection, given a data set of 7000 images to choose from. This time scales with the size of the data set and the number of features used to represent the images.

Question 1: What is the coolest thing you have found using this algorithm?

I think the discovery of an unexpected molecule (CaF) in ChemCam observations of rocks on Mars is pretty exciting. This occurred simultaneously with human investigation of the same observations, and it provides an example of how DEMUD can quickly identify the most interesting observations that merit human follow-up.

Question 2: When the model is wrong, how does that affect its ability to learn?

If DEMUD selects an item that is not actually novel, that will not adversely affect future selections because that item will immediately be incorporated into the SVD model and therefore ignored (not selected) if it occurs again.

Question 1: Can we really use water and life on Earth today as an analogue for ancient Mars?

That is a great general question and one that we do not have a conclusive answer for yet. We continue to investigate the best analogue settings on Earth that *could* be similar to ancient Mars, but at this point we cannot know for sure.

Question 2: Is machine learning a faster way to interpret data and form a more complete hypothesis?

Faster or more complete than what? I am not sure how to answer this question.

Question 1: Bouncing off of the point that was made today about the network noticing that one of the insect pictures had no insect and it found that to be quite interesting: in your opinion, what is it that makes the "class/explanation" that a neural network outputs useful or interesting to a human, and what informs what we find to be meaningful?

In that case, the neural network's visualization of the novel image content enabled us to immediately understand why an otherwise empty image was selected as "interesting" or "novel." Without that explanation, we might have been scratching our heads for a while or hunting for a bug in the code. In fact, it performed exactly as expected.

Question 2: Where do you see neural networks having the most impact in academia or in society more broadly?

Neural networks are being used in many areas today, from self-driving cars to social media to advertising to finance. Probably just about every discipline is now looking at ways that neural networks could potentially be of use. Some of these efforts will result in useful systems or discoveries, while others may not find a perfect fit. It is valuable to keep an open mind and try multiple different solutions to a given problem.

Question 1: Is there a point where DEMUND can have gone through so much data to where it starts missing something that should have been considered an outlier because it has generalized so much?

DEMUD has one input parameter (k), and this parameter controls how much generalization will happen. The k value indicates how many principal components DEMUD will use to represent the data it has seen. If k is small, DEMUD will only learn very general trends and identify many components as novel. If k is large, DEMUD will more precisely memorize fine details, and it will only identify minor components as novel. The user can control what kind of outliers DEMUD should be sensitive to by changing the value of k. Sometimes it takes multiple runs to decide what a good value of k is for a given problem.

Question 2: Is DEMUND going to be used on Mars 2020 when it is running autonomously?

No, DEMUD is currently only used to analyze data after it is transmitted to the Earth. We envision future operational scenarios in which it could be used onboard spacecraft to help prioritize the data for transmission (e.g., most novel first) or even to help decide which rocks or other targets to investigate next.

Question 1: Why was it expected that ChemCam would find only elements and not molecules?

Because the laser it employs is powerful enough that it breaks molecules into individual atoms (elements).

Question 2: Do you think machine learning could be used to identify previously undiscovered exoplanets in direct imaging data?

If the image quality is sufficiently good and the planets sufficiently large to be distinguished from the background, then yes, in theory this would be possible.

Question 1: Is there a possibility that DEMUD could miss a feature that a human would find interesting?

Absolutely. DEMUD does not know what humans will find interesting. It relies on the representation (feature vector) used to represent the data content and is only sensitive to what is captured in those features. It can be useful to run DEMUD with several different representations to see which ones are most beneficial. In my talk, I showed results when analyzing images and using either the raw pixels, SIFT features, or CNN-based features to represent their content. Generally, the CNN features provided the best match for how humans interpret the images.

Question 2: Why is there so much pixel loss through the algorithm?

I am not sure what is meant by "pixel loss." If you mean "loss of spatial resolution" in the image visualization, this is because the original 227x227 image has a dimensionality of 227x227x3 (for red/green/blue channels) = 154,587 values, while the CNN representation is a compressed version with only 4,096 values. Some spatial detail is therefore lost and cannot be reconstructed (think of it like a lossy compression). We give up some detail to be able to abstract away from pixel values to content like "insect leg" and "Mars rover part".

Question 1: What kind of changes do you think they will make to the camera on the next large mission on Mars?

New cameras tend to try for higher spatial (and spectral) resolution, subject to onboard storage available and mission downlink restrictions.

Question 2: What would advancements in neural networks provide for your research, and what do outliers represent in your data sets?

Further advances in visualizing image content (converting CNN feature vectors back into 2D images) would further improve the explanations that DEMUD is able to provide. The outliers we find tend to represent new classes or object types, which can aid in further scientific investigation when exploring new environments.

Question 1: You mentioned that reconstructing the images from 4096 data points leads to lower resolution images, is this limitation a computational one, and would the process be improved by having higher resolution?

We give up spatial details to be able to abstract away from pixel values to content like "insect leg" and "Mars rover part". If we use more features, the representation becomes more tied to the specific pixel values and therefore would not recognize the same insect if moved or rotated (for example).

Question 2: How are you able to perform such complicated computational tasks with the very limited computational power of something like a rover?

Excellent question! To operate onboard a spacecraft, we take a machine learning model and simplify and compress it to fit within available memory and computational resources. Often this means giving up some accuracy (using a simplified model), so we also do a careful trade study to decide where on the simplicity vs. performance tradeoff we want to be. The code is ported to C and optimized for use in the onboard environment.

Question 1: Has demud ever identified something incorrectly as novel and if so did it then teach itself to correct that to non-novel?

The basic DEMUD algorithm never gets feedback from humans about whether it was right or not. However, it assumes that anything it previously selected is (now) no longer novel. Therefore, everything gets "corrected" to being "non-novel" after it is selected. We later developed a variant to allow the opposite kind of feedback: you can tell DEMUD when an item it selects should *stay* novel/interesting and, rather than being ignored, similar objects should be prioritized for selection because you would like to see more of them.

Question 2: Can you use demud to identify unique planetary bodies or planets?

If you have a data set that contains observations of planets as a whole, DEMUD could analyze that data set and tell you which ones were unusual and why.

Question 1: Several data sets were mentioned that you have applied your algorithm to. What data set haven't you applied it to that you are most interested in applying it to and why?

We have explored applying DEMUD to aerosol observations from Earth orbit to automatically detect and characterize forest fires or other atmospheric activity. I think there is a lot of potential for this kind of analysis to detect unusual events and teach us new things about what is happening on our own planet.

Question 2: What was the most unexpected or interesting scientific result so far from using DEMUD and why?

I think the discovery of an unexpected molecule (CaF) in ChemCam observations of rocks on Mars is pretty exciting. This occurred simultaneously with human investigation of the same observations, and it provides an example of how DEMUD can quickly identify the most interesting observations that merit human follow-up.

Question 1: What kinds of things about rocks can DEMUD reveal?

That depends on what information you have collected about the rocks to provide to DEMUD. If the observations are images, then DEMUD can identify unusual colors, shapes, textures, etc. If the observations are from a spectrometer, then DEMUD can identify unusual compositions, mineralogy, etc. Other kinds of data would reveal other kinds of properties.

Question 2: Is there a way to teach the program to better distinguish between what is novelty and what us just natural variation?

It is possible to initialize DEMUD with a starting data set that contains many existing observations of already known phenomena. That way, DEMUD doesn't have to start from zero and discover things we already know; it can instead start with a good idea of what is normal (including natural variation) and focus on truly novel discoveries.

Question 1: With so little known about the planets found with the kepler telescope, how do you determine what counts as a 'novelty' to detect somewhere where you can't take spectral analysis or other tests?

"Novelty" is always with respect to the features available to describe the objects. This could be unusual elements within images, unusual spectral features, unusual sensor readings, etc.

Question 2: Do you see programs like this moving into other tech areas such as facial recognition or product quality control?

Yes, you could potentially use DEMUD to identify new people within a large data set of human faces, or to identify manufacturing defects in images of products on an assembly line. These would be interesting applications of the technology.

Question 1: How can you be sure nothing crucial is overlooked in the model when human inspection is removed?

In our envisioned use of DEMUD, human inspection is not removed. Instead, DEMUD prioritizes the data and determines the order in which humans examine the data. The goal is to enable them to quickly see all of the interesting, new kinds of observations, without having to examine every single one. Of course, if they don't examine every single one, you can never be absolutely sure that something wasn't overlooked, so we do a lot of tests on a variety of data sets with known labels indicating the human-assigned classes, where we can check to see if DEMUD found them all.

Question 2: How does the distinction of classes change across different fields?

The concept of a "class" varies by field and even by individual. Some people may be interested in dividing images into large categories (animal, vegetable, mineral) while others might want to identify individual dog breeds. This granularity can be controlled through the DEMUD parameter k. The k value indicates how many principal components DEMUD will use to represent the data it has seen. If k is small, DEMUD will only learn very general trends and identify many components as novel. If k is large, DEMUD will more precisely memorize fine details, and it will only identify minor components as novel. The user can control what kind of outliers DEMUD should be sensitive to by changing the value of k. Sometimes it takes multiple runs to decide what a good value of k is for a given problem.

Question 1: Is there a limit to the number of new features DEMUD can detect?

Perhaps you meant "number of new classes." (We use "features" to refer to the representation DEMUD is given, like pixel values in images or measurements such as height, weight, size, etc., depending on the type of items.) DEMUD is iterative, so it will keep going as long as the human user wants to see more selections.

Question 2: What are the limitations of DEMUD in regards to visual input?

The primary limitation is in terms of the representation used. I showed a comparison of using raw pixels vs. SIFT features vs. CNN-based features as one example of how the representation affects performance. The better we can capture image content in terms of some kind of numeric features, the better DEMUD will perform.

Question 1: Is there any area of science that you think machine learning will not be able to contribute to?

Machine learning is a tool. I do not think it is a magic wand that can solve all problems and answer all questions. Human ingenuity is needed to decide how to apply it. “Machine learning” also encompasses a range of different capabilities, from classification to regression to novelty detection (as in this talk). It can help in a variety of different ways.

Question 2: Even with the use of training data sets, how can we be certain that we can trust the results found via machine learning?

That question motivates the area of machine learning that focuses on “explainable” machine learning, in which the machine learning system needs to be able to provide a justification or reason for its decisions. The explanation must be human-comprehensible to be useful. This is the same standard to which we hold other humans; we require that they can communicate why they made a particular choice. If we understand and accept the reason, then we feel we can trust the decision.

Question 1: Is anyone working on incorporating optimization algorithms for vector sizing and manipulation, to minimize/manage impacts of information loss?

It seems likely. I don’t have any pointers to specific investigations to recommend.

Question 2: The graph showing how efficiently classes are identified and added to the model gave me pause. How do you ensure there's not an artificially inflated number of classes generated? In other words, if the model detects 100 base classes in a dataset of 200,000, with 20 actual classes (10 of which were known before and 10 of which are legitimately novel), how do you train it that some “novelty” it detected actually isn't novel but assigned to another class?

We were comparing it to human-assigned class labels, so we have a correct answer to compare to for those plots. If we feel confident that it has done the right thing in a known setting, then we can expect it will also perform well on new data sets.

Question 1: What machine learning systems should be present on a space probe that were to visit Alpha Centauri?

Machine learning could assist with detecting potential biosignatures (e.g., specific elements or minerals or something more generic like a system being out of “chemical equilibrium” that can suggest that some kind of life (or other) activity is active.

Question 2: Can Machine Learning systems be programed to evolve on their own?

If this topic interests you, I recommend looking up information about “genetic programming” or “reinforcement learning” in which a system can develop and refine its own algorithms to solve a problem.

Question 1: What is the biggest hurdle to overcome to get the next big advancement in machine learning technology?

That is a very broad question. Personally, I think having machine learning system that can provide explanations for their decisions can greatly increase human trust and therefore wider use of machine learning methods. This is an area of ongoing investigation and improvement. There are also always open questions about how to increase computational efficiency so more data can be processed more quickly.

Question 2: It seems like machine learning is good for data processing after it's collected, is there any way to use it for the ata collection process itself?

Yes, the results of analyzing one batch of data can inform how the next data is collected (e.g., the AEGIS system on the Mars Science Laboratory rover examines images to identify rocks and select the next target for the ChemCam laser spectrometer).

Question 1: Can DEMUD detect temporal changes within an image?

Temporal changes would be detected by analyzing a sequence of images. It would be possible to configure a DEMUD experiment in which before/after pairs of images are provided as inputs (or a multi-image sequence). DEMUD would then have the opportunity to detect changes within those inputs.

Question 2: Can DEMUD recognize the same image that was taken at different lighting conditions?

Most likely, the first time it observes a change in illumination, it would identify that as novel. Then it would learn to ignore that difference as being unimportant in future selections.

Question 1: After the neural network system for mars pictures has ran through many different images, would the unique photos selected become less and less unique due to the network growing in knowledge, and ultimately become less useful?

Yes, the initial selections will have the most dramatic or largest amount of novelty. DEMUD provides a ranking. As you go down the list, the amount of novelty decreases. The user can stop the process when they determine that the “novel” content is no longer of true interest (e.g., it could just be bits of noise or bad pixels within an image).

Question 2: Do you believe this technology can suffice when it comes to data gathering on mars, or would you consider gathering data remotely to be a large hindrance compared to what could be achieved in person?

DEMUD provides an analysis of data that was gathered by another process, which could be an instrument or a human. It could be used in either scenario to help detect novelty within the data that was collected.

Question 1: What is an efficient method you use to validate the machine learning algorithm did not miss potential features of interest?

That is a good and open question. We do a lot of tests on a variety of data sets with known labels indicating the human-assigned classes, where we can check to see if DEMUD found them all. For a new (unlabeled) data set, one can spot-check by examining the remaining images that were not selected by DEMUD.

Question 2: Why does DEMUD show some green color in the "what's new" image for the dust removal tool?

Because the image was "less red" than DEMUD expected, given the previous images. If you remove red from an image, you get an enhanced green/blue appearance.

Question 1: The process of novelty detection seems computationally intense. How easy/ effective would it be to do it on-board the spacecraft?

DEMUD is currently only used to analyze data after it is transmitted to the Earth. We envision future operational scenarios in which it could be used onboard spacecraft to help prioritize the data for transmission (e.g., most novel first) or even to help decide which rocks or other targets to investigate next. On my laptop, it takes a few seconds per suggestion for a data set with 7000 images. For use onboard, we would compress and optimize the algorithm to fit within available computational and memory resources.

Question 2: When the number of classes is low (say for a body like Bennu), is there still any advantage in using DEMUD?

The definition of "class" is subjective; for example, one person might divide the insects by their species, while another person would divide them by individual variations or pose. On Bennu, one person might consider it to be all one class, while another would be interested in minor compositional variations. The user would control this through the DEMUD "k" parameter. The k value indicates how many principal components DEMUD will use to represent the data it has seen. If k is small, DEMUD will only learn very general trends and identify many components as novel. If k is large, DEMUD will more precisely memorize fine details, and it will only identify minor components as novel. The user can control what kind of outliers DEMUD should be sensitive to by changing the value of k. Sometimes it takes multiple runs to decide what a good value of k is for a given problem.

Question 1: How did you get interested in machine learning?

I took a class on artificial intelligence, and the piece that seemed to make to make systems truly intelligent is their ability to learn and improve over time. Machine learning is the study of how to make that possible.

Question 2: Can DEMUD algorithm be modified from just detecting objects, and what further can be done in the algorithm?

DEMUD detects novelty; given images, it detects novel images (which are often specific objects because that is what we take pictures of), while given spectrometer readings, it detects novel composition or other spectral features.

Question 1: Can machine learning be used for problems in tectonics?

What kind of problems would you like to investigate in tectonics? Can they be phrased as a classification task (assigning labels to observations), a regression problem (predicting a numeric value like the amount of strain or slip), or unlabeled problem (like novelty detection, the subject of my talk)? If so, machine learning can be of use.

Question 2: How can machine learning-related errors be reduced?

We have methods to conduct an error analysis on data where the correct answer is known. When we identify patterns in errors, we develop strategies to correct them, which could mean adjusting the algorithm, collecting new data, or correcting human errors in how the training data was labeled.

Question 1: In your presentation, it looked like most of the novelty detections were features at similar image scales. How does DEMUD handle scale differences and detecting both small and large anomalies?

DEMUD analyzes the content of an image. Scale isn't a factor except in how the representation (features) are constructed. The similar scale in the examples I showed is due to the fact that the images themselves tend to be taken at similar scales, for a particular data set (e.g., the insect data set was collected using the same microscope with the same field of view).

Question 2: How would you use DEMUD to identify changes on a spacecraft (wheel damage) and surface environment changes caused by wind?

DEMUD selects images that contain novel content, so in the examples you mentioned, you would give DEMUD (a) many images of wheels, and it would identify the ones with novel content (e.g., wheel damage) or (b) many images of the surface, and any unusual surface features would be selected. Wind-sculpted features might be quite subtle and difficult to detect; wheel changes would be more straightforward!

Question 1: What does the “Eigenbasis” in the DEMUD acronym mean?

DEMUD uses a Singular Value Decomposition (SVD) to model the data it has seen. Another way to describe this is by “eigenbasis modeling” since the SVD breaks the data down into eigenvectors that represent common patterns or trends in the data.

Question 2: How long does it take to train a neural network before it can be put to work on a dataset and be expected to produce reliable results?

The answer depends on the complexity of the data set; simple data sets require less training while more complicated ones (e.g., with many different classes or more pixels per image) require more training.

Question 1: Is machine learning working for the figures of thermal chemical convection for the Earth's mantle?

I do not know how to answer this question.

Question 2: 2. Is there any specific number defining the number for “big data”, hundreds, thousands, millions?

“Big data” can be “big” in terms of size (number of items) or rate (number of items coming in per second) complexity (difficulty for humans to examine each item and understand it). There is no fixed number, but generally, we say “big data” when it exceeds the ability of a single human to keep up (or would take an excessively long time for a human to go through it).