Large streams of data warn cars, banks and oil drillers

April 16 2014

Better warning systems that alert motorists to a collision, make banks aware of the risk of losses on bad customers, and tell oil companies about potential problems with new drilling. This is the aim of AMIDST, the EU project coordinated by Aalborg University and software company Hugin Expert A/S.

A German car maker, a Spanish bank and a Norwegian IT company in the oil industry, all with customer issues and huge amounts of data that newly developed software will translate into smart decisions. The project also involves Norwegian and Spanish universities, and the goal is to create a tool that is flexible enough to encompass a wide variety of industries, explains Anders L. Madsen, CEO, Hugin Expert A/S, Research and Technology Development Manager for the AMIDST project (Analysis of MassIve Data STreams) that has a budget of nearly DKK 30 million (EUR 3.9 million):

"With cars, it's about predicting when other motorists cut in front when changing lanes. The bank case has 1000 different pieces of information on each of four million customers where we will identify loss patterns. And the oil case will use extensive measurements from sensors to predict when something is about to go wrong during a drilling operation. These are three extremely good and very specific examples with vastly different demands, all of which will show what this type of software can do when we are able to scale it up to work with massive streams of data," says Anders L. Madsen.

Hugin Expert A/S already develops software for decision support systems so scaling up will both benefit new and existing users who, for example, use the technology for credit and insurance risk assessment, environmental monitoring and for monitoring processes in the industry.

"We're bringing the tool to a new level so it's not overwhelmed by large amounts of data. It will be developed to handle much larger tasks with the available hardware. Hardware is very important, and the algorithms must not be limited by it; they must be scaled to run on the hardware that is available," says Anders L. Madsen.

From supercomputer to pocket device

Aalborg University is represented by Thomas Dyhre Nielsen, Associate Professor in the Department of Computer Science, in the research group for Machine Intelligence. His focus is the models and algorithms the project will develop:

"Machine learning is one of the major challenges we face when working with data streams and trying to make predictions based on them. The models we develop will primarily be learned on the basis of the data we receive from the partners in the project. The models will be part of an overall framework that can embrace the three use cases in the project, even though they represent different types of data and run on platforms with different resources," says Thomas Dyhre Nielsen.

As a researcher, he has access to supercomputers with lots of processing power, but one of the challenges in addressing massive data streams is to make do with the resources that are available:

"Supercomputers and faster regular desktop computers allow for things we could not previously do. But for some models we may need to develop smart algorithms that can make a trade-off between precision

and computational complexity in order to ensure that the algorithms can run in realistic time on the available devices," says Thomas Dyhre Nielsen

Hugin Expert's CEO, Anders L. Madsen, mentions the test with cars as an example of the necessity of balance because the extensive sensor data from Daimler vehicles will really put the technology and software to the test:

"It's all about increasing safety when you're driving at 100 km/hour with cars around you. The sensors collect 22 million measurements per hour. With new data every 40 milliseconds, it doesn't help that you need 60 milliseconds for calculations. The system must be able to keep up. But the processing power of a car is so limited that it makes your phone look like a ninja computer in comparison. We have to take this into account."

Life and money at stake

A solution that can help prevent motorists from colliding at high speeds inherently has the potential to save lives. The EU has set a target of reducing the number of road fatalities by 50 percent from 2011 to 2020. Along with other initiatives, AMIDST can play a role.

In the other sub-projects, it is first and foremost money that is at stake. But lots of it. The financial solution for the bank Cajamar in Spain could, according to estimates from the bank, save DKK 56 million (EUR 7.5 million) a year if it lives up to expectations for predicting and forestalling losses on bad customers. Verdande Technology, the Norwegian IT supplier that develops specialized software for the oil industry, expects to save end users expensive downtime with new drilling.

Cars, banks and oil drilling are good examples, but according to Anders

L. Madsen, it could just as well have been other industries with other challenges related to data streams:

"Banks and insurance companies are drowning in data. But it could also be the medical field, production company processes, or data embedded in printers, automobiles and submarines. We hear a lot about Big Data, but this is different because we work with streams of structured data where we know what we're measuring. Our challenge is to analyze the data stream and translate it into a graphical representation," says the Hugin CEO.

Provided by Aalborg University