Center for Applied Intelligent Systems Research (CAISR)

Thorsteinn Rögnvaldsson Antanas Verikas Josef Bigun
Sławomir Nowaczyk Anita Sant'Anna Björn Åstrand
Jens Lundström Stefan Byttner Roland Thörner
Fernando Alonso-Fernandez Martin Cooney Rafael Valencia *

Abstract

Awareness is a broad concept, just like "intelligence", and has many connotations. This paper presents the vision of researchers from Center for Applied Intelligent Systems Research (CAISR) at Halmstad University.

1 Aware systems research definition

Awareness is a broad concept (just like "intelligence") and has many connotations. Relating to research on computing systems, there are two directions: one regarding the logical definition of aware and how to determine if a system is aware or not, another regarding what is required to be aware, i.e. what capabilities are required to be aware. In CAISR we focus on the latter.

Examples along the first direction can be found in, e.g., the papers by Hintikka (1975), Fagin and Halpern (1988) or Modica and Rustichini (1994). They boil down to statements like "Awareness of ϕ if they explicitly know ϕ or they explicitly know they don't explicitly know ϕ" over enumerations of possible worlds, where ϕ is a logical statement that can be true or false. As argued by Devanur and Fortnow (2009), such definitions are of little practical use since there is never endless time to search through all possible objects. It is also relevant to ask if awareness is a purely binary concept. Human awareness works quite differently: we are more aware of recent facts than old facts, even though we know them all. Devanur and Fortnow (2009) suggest a more practical and human-like definition: that awareness of an object is inversely proportional to the time needed to enumerate that object in a certain environment and a context. If you cannot do this within a certain time then you are effectively unaware of the object.

The work by Endsley (1995) is central regarding the second direction: what is required to be aware. Endsley (1995) describes awareness, from a human psychological perspective, as knowledge created through interaction between an agent and its environment, and "knowing what is going on" (see Gutwin and Greenberg, 2002, and references mentioned therein). Similarly, but more recently, Zhao et al. (2012) state, from a computing perspective, that awareness is the ability to perceive, to feel, or to be conscious of events, objects, or sensory patterns, but it may not lead directly to full comprehension. Zhao (2013) refines this into awareness being "a mechanism for obtaining information or materials which are useful for human users, for other systems, or for other parts of the same system, to make decisions". Zhao (2013) further comments that computationally aware systems have been studied for a long time but are often classified based on the event to be aware of. Examples include, context aware, situation aware, intention aware, preference aware, location aware, energy aware, risk aware, chance aware, and so on. Such classifications are not helpful to explore the key properties of aware systems since it divides aware systems research into many subsystems with unclear boundaries between them.

In CAISR we do not pursue research on what it logically means to be aware. We follow the direction of Endsley (1995) and Zhao (2013) and define aware systems research as: *Research on the design of systems that, as autonomously as possible,*

*Authors are at Center for Applied Intelligent Systems Research, Halmstad University, Sweden. Email addresses follow *firstname.lastname@hh.se* pattern

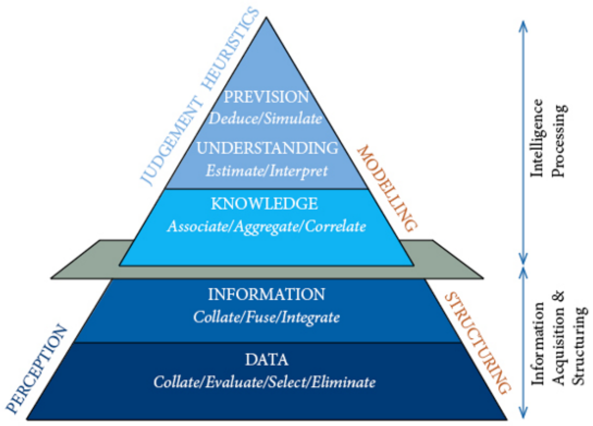

Figure 1: The knowledge pyramid.

can construct knowledge from real life data created through the interaction between a system and its environment. This data necessarily includes streaming data. Such systems should be able to handle events that are unknown at the time of design.

The goal with artificial intelligence (AI) research and development is to construct systems that behave intelligently. However, it is standard to assume that human experts define the task that the system should perform and that the collected data used for building systems reflect the "reality". This means that these systems are "designed" or "programmed", which leads to systems that break when the context changes. Our aim is to approach the construction of AI systems that can do "life-long learning"; systems that require less supervision and handle surprising situations. In order to do so, the systems must become more "aware" and able to learn on their own.

The construction of knowledge (going from perception to knowledge) is often represented by the knowledge pyramid (Ackoff 1989), see Figure 1; the higher a system reaches on the pyramid, the more knowledge it has and the more aware it can be. A version of this structure and knowledge pyramid is also how NASA looks at intelligent monitoring of missions (Degani et al., 2009). A fully aware system will have interaction both upwards and downwards in the pyramid, e.g. events higher up in the pyramid will affect choices on what data to collect.

Much of the work on machine learning (ML) and AI has not considered the knowledge creation aspects of intelligent systems. The common approach is to have humans define the problem in significant detail, for example the data characteristics, the representations used, the model used, etc. and the task is to build a machine that replicates the human decision. There are therefore many open research challenges for each of the stages in the knowledge pyramid:

Data: This deals with the collection of data and the representation of it, answering the question "with what" (Zhao 2013). An open question is how an autonomous system should select what data to collect? With streaming data from all sorts of sources, and with data bases of varying quality, how can the system tell what data are (or will be) relevant? The "with what" decision is to a large part done by humans today, which simplifies the learning problem immensely, but it is clearly one of the most relevant questions for autonomous learning. A related, much more researched, question is how to create general features; features that will apply to many problems (e.g. invariant features in images). Furthermore, with endless streams of data (i.e. in the "internet of things" era) it is impossible, even uninteresting, to save all data. It should be possible to save snapshots, compressed, or aggregated representations of the data. These representations should be learned and be general so that they apply to many different tasks. The fact that the working environment of a system may change frequently, requires considering the plasticity-stability problem seriously (Zhao 2013); features that look unimportant today may end up being important tomorrow. A system that is aware should therefore be curious and never stop exploring.

Information: This relates to questions that begin with "who, what, when and how many" (Rowley 2007), creating "events" from the data in the layer below. Examples of operations that are required for this are classification, rearranging/sorting, aggregating, performing calculations, and selection (see e.g. Curtis and Cobham 2005). Much ML research (including that on deep learning models) has been devoted to this stage, and also AI research for e.g. text and language parsing. Important open research questions here regard autonomous clustering and categorization of events. How can events be grouped into categories, e.g. common or uncommon, normal or odd, for later use? A challenge is to do this autonomously, or with only limited interaction with a human (that can provide hints), and in non-stationary environments. There is a significant body of relevant research on deviation detection, change detection and autonomous clustering.

Knowledge: This level is about creating "rules" from the information (rules can be in the form of models and not necessarily in e.g. predicate knowledge form). This always requires combining information from different sources. For example, is an observed "event" from one set of data sources associated with some other event, and can such associations be formulated into rules (and are these rules correct)? One obvious example is the supervised learning setting, where information "events" (input) are matched to correct responses (target) provided by a human expert and encoded into a rule (model). A very relevant research question here is how human generated knowledge, e.g. in the form of text comments in human curated data bases or models of the environment, can be combined with the machine generated information to create rules. Horeis and Sick (2007) have presented one example for incorporating human experts in this process. Another question deals with knowledge representations (knowledge structures); how can knowledge be represented so that it can be used for reasoning and prediction? A set of well-defined, highly-organized yet dynamic knowledge structures is one prerequisite for achieving awareness. A knowledge structure should evolve over time from experience, thus allowing for learning from data and human experts and be capable of taking into account different kinds of initial domain knowledge. Learning from human experts requires automatic methods to transform textual data into conceptual structures (automated ontology learning). Semantic knowledge self-organization is a very important and desirable property of knowledge structures.

Understanding and prevision: (Sometimes this layer is referred to as the "wisdom" layer.) This layer deals with the question "why" or "what will happen"? It is about the ability to project into the future and reason back into the past. An aware system should be capable of extrapolating information into the future, and be able to estimate and evaluate the consequences of certain actions based on previous observations. In robotics, this can be predicting paths. In other fields it tends to mean reasoning, e.g. ontology-based reasoning. An active research field here is the autonomous creation (learning) of ontologies that can be used for reasoning (Zhou, 2007; Barforush & Rahnama 2012).

In all levels is uncertainty a key aspect. The uncertainty in the data should be propagated to the information level, where it is transformed into an uncertainty in the information, and then on to the knowledge level, etc..

An important point that perhaps is not obvious in our perspective on aware systems is the life-long learning although it is implicit in the "unknown at the time of design". We are not approaching problems where all data is available at once; we are studying systems where learning takes place over time, typically with streaming data. There are already excellent efforts towards automated data mining or model building, e.g. the recent "Feature Lab" by Kanter & Veeramachaneni (2015) or the KXEN system that is now part of SAP (Fogelman-Soulie & Marcade, 2008). These build on the idea that all data is available and the question is how the relationships in this data should be best modeled.

In all levels in the knowledge triangle is the human role and interaction with the systems an important research question. Human can play part in all steps of the knowledge creation, leading to a semi-unsupervised knowledge creation, e.g. providing clues on interesting data representations, clustering events, providing external data, giving feedback on suggested structures, etcetera. What is important is how machine and human create knowledge together, not like in the traditional AI or ML form where humans provide expertize that the machine is expected to replicate. We refer to this as joint human-machine learning.

2 Meeting societal challenges

We list some examples below that are particularly relevant for CAISR, using headings from the EU Horizon 2020 framework program, that tie to activities within CAISR.

2.1 Health, demographic change, wellbeing

Improving individuals' health patterns: The development in wearables has inspired a vision of using self-tracking for personalized (and improved) health. This means wearable devices that log our activities and interact with us, in order to help us improve our lifestyle (eat better food, exercise more, etc.) and, with time, decrease the load on the health care system. Numerous apps in this field are being introduced daily on the market, there are lots of user data being sent to servers all the time, and there are national projects aiming at storing long-term data for individuals. Khosla claims that "In fifteen years, data will transform diagnostics, to the point where automated systems may displace up to 80-percent of physicians' standard work" (Khosla, 2014). Even though the statement is about medical diagnosis and not proactive health patterns per se, it certainly applies also to promoting healthy behaviors. Combining life-logging data with health record data will be, to say the least, challenging (missing data, erroneous data, etc., humans are entering quite a lot of it) but the large quantity of data means that with time we should be able to make good analysis and provide good advice.

Healthy and active ageing: Wearables also apply to healthy and active ageing. However, what is equally important is in-home monitoring (ambient assisted living). Elderly living homes can be equipped with sensors and the sensor data (streams) analyzed and used for providing security solutions, emergency solutions, and assistive services. This offers a decreased cost for elderly living services, while maintaining a high quality of service. Aware systems research is about developing methods for autonomously analyzing such streams of data and construct knowledge about the individuals' living patterns. Without such knowledge it is difficult (impossible) to design a working (and simple) intelligent service for elderly living. In a critical text on ambient intelligence, two Philips researchers (Reddering & Scholten, 2003) express what they consider the most important challenges for ambient intelligence if it should ever become a useful technology: to construct knowledge systems that are simple, that can cope with the diversity and unpredictability of human needs, and that learn to ask and value the unknown.

2.2 Secure, Clean and Efficient Energy

Better models of energy use and demands: Energy production is to an increasing level produced by small-scale renewable energy sources (solar, wind, biofuel). The volatility of wind and solar (and in the future: wave) generation creates problems in balancing the demand with the generation of energy and operating conventional power plants in part load. This requires better prediction of energy demands. On a consumer scale this may be possible to achieve by combining ambient intelligence with smart power meters, but modelling this will require autonomous knowledge creation since the data and variation is so large. Ambient intelligent systems technology can also be used to learn the inhabitants' living patterns and provide feedback in order to improve (lower) the energy consumption. It is difficult to imagine how this can be done without autonomous knowledge creation.

2.3 Smart, green and integrated transport

Improving vehicle uptime: Transport is intimately linked to economic growth. Transport cost and efficiency are intimately linked to vehicle uptime, i.e. the reliability of the vehicle fleet. Transportation vehicles are becoming more and more advanced, with huge amounts of data streaming on the controller area network on-board the vehicles. More and more data are also logged in databases about maintenance operations and vehicle setups. Aware systems will be important for better maintenance solutions that build upon this data. Maintenance of vehicles is not optimal today; there are erroneous repairs done, there are on-road breakdowns that could be prevented, and there many commercial transport vehicles that don't pass the national annual inspections. All this can be improved with systems that allow logging on-board (streaming) data, fleet wide comparisons, and connecting on-board signatures with repair histories. Making sense from these data, learning over the life-time of the fleets, requires autonomous knowledge creation.

3 CAISR research questions

The CAISR focus is on research questions that are general across application areas, across research groups and relevant for external partners. We describe them below in relation to the levels in the knowledge pyramid. They are further refined, with specific contributions we have made and intend to make in CAISR, at the end of this section and in the following section.

Data: The research questions we explore here are how to select what data to collect and how to find general and robust representations of data. This can be by learning representations, by designing representations, or by searching through sets of representations and estimating how good (interesting) they are. This means research on how to autonomously engineer features, or ways to learn representations. Work on automatic feature engineering has been presented by Kanter & Veeramachaneni (2015), Cheng et al (2011), and Paulheim & Furnkranz (2012). Bengio et al. (2013) present a review on learning representations. It also means research on the general applicability of representations, at least for certain types of signals like images, see e.g. Bigun (2006) for a discussion. It means research on measures to determine how interesting a particular representation is, which is related to (but not equal to) measuring interestingness of rules, see e.g. Zhang et al. (2009) for a review of the latter. The plasticity-stability problem mentioned above is very important and something that is usually not handled. There are many practical issues here related to e.g. dealing with missing, flawed or erroneous data. This corresponds to using feedback from higher levels in the knowledge pyramid; what data is expected based on the type of event? Another important issue is how human expertize can be combined with machine work, i.e. how the machine data exploration can be done in interaction with humans. One more important issue is curiosity; an aware system must continuously search for interesting things.

Information: The research questions we explore here are how to do (semi-)autonomous deviation detection and autonomous clustering of events, as well as the maintenance of such categorizations, e.g. dealing with concept drift, seasonal variations, application changes, and so on. Clustering is still something of an art, see von Luxburg et al. (2012), and certainly a challenge to do well in an unsupervised manner and for general types of problems. Iverson (2008) has suggested, and patented, one general data driven solution intended for system maintenance. Angelov (2013) suggests fuzzy clustering as a method to design general purpose cluster structures. But there is still lots of room for improvement, or as von Luxburg et al (2012) put it: "Depending on the use to which a clustering is to be put, the same clustering can either be helpful or useless". It is also important to incorporate humans in the loop; can humans provide initial suggestions for categories, can humans give feedback on suggested categorizations, etc.?

Knowledge: The research questions we explore here are how to associate events from different data sources, including human generated data. An important part is also how this knowledge should be represented. With reallife data, the information provided will (inevitably) be connected with uncertainty, and a question is how to handle the combination of two information sources that both may be uncertain. In this context it is important to also consider how a human can be incorporated to build this knowledge, in a semi-supervised way.

Understanding and prevision: We aim to be able to predict the progress of observed events, and explain why certain things have happened. However, we will be using hand-made ontologies (at least initially), and not do research on the generation of ontologies.

Aware systems research is a systems science, i.e. there are many parts to the system and the results need to address several parts in the knowledge triangle and tie them together. To enable this, we aim to build demonstrators to showcase what this means, with sets of tools for all levels (at least for three levels). These tools will be parts in toolboxes for aware systems. One demonstrator will be the intelligent home environment (the aware system for ambient assisted living). Another one, funded mainly by projects outside of the Knowledge Foundation CAISR funds, will be the self-aware vehicle fleet for increased uptime. A very likely demonstrator is the mobile based decision support system for persons diagnosed with a chronic condition. A possible one is the aware fork-lift truck in a warehouse.

4 CAISR research projects

In this section we summarize the aware systems contributions made in CAISR during the first four years and the proposed contributions during the second half. Contributions have been made mostly on the data level and the information level, and to a very small part on the knowledge level. The data level research has been on constructing robust and general features (HMC2) and/or semi-supervised feature construction (AIMS and SA3L). The information level research has been on using unsupervised clustering to group observations (AIMS and SA3L). The knowledge level research has throughout been on using human supplied labels for the clusters, or observations. The planned contributions during the second half are on the data, the information, and the knowledge levels.

4.1 AIMS (Automatic inventory and mapping of goods)

A robot acquires semantics by linking its world model with human knowledge. A challenge is to find the appropriate level of abstraction before linking the human knowledge and the robot world model. The scientific results include a semi-supervised approach for semantic mapping, introducing human knowledge after unsupervised place categorization, in combination with adaptive cell decomposition of an occupancy grid map. The system autonomously builds a high level spatial model of the world by adopting generic features on the data (occupancy map) and instantiates it, without prior knowledge of the environment. Semantic inference is done on the derived instantiates and semantics are provided as labels accompanied with their functionality and inter relations between them. The proposed adaptive cell decomposition method interprets occupancy maps to bring out underlying spatial characteristics in the data (environment). The information is stored in two corresponding data structures in a format readily usable for humans and machines. Additional knowledge is created by subsequent semantic labeling by using human constructed templates. In addition, we have presented a canonical geometric-semantic model (adjustable according to different scenarios), along with a method for generating and matching these models into the latent structure of the map. The result is a geometric-semantic map there semantics (corridors, pallet rack cells) are encoded into the model through the choice of landmarks (pallet rack pillars).

4.2 HMC2 (Human Motion Classification and Characterization)

The methods and algorithms developed for prediction of physiological parameters of an athlete from EMG data cover have on the data and information level concerned segmentation/structuring of raw signals so that we get robustness of segmentation results in case of changing signal variability. The system assesses signal variability on the information level and makes necessary adjustments in the data level to obtain adequate results of the structuring. Accurate predictions of physiological parameters obtained from trainable models using the extracted features allow assessing state/condition of the athlete, providing short-term advice and gaining knowledge for long-term evaluations and future planning. The reference conditions are obtained by separate measurements.

The methods and algorithms developed for estimation of the fundamental parameters in human locomotion make uses of existing knowledge about human walk, known from research in physiology. This makes the features robust and possible to transfer from the laboratory settings into the real-world. Robust estimation of the fundamental gait events, assessed longitudinal over long time spans, allow generation of new knowledge on e.g. how variability in movement patterns is influenced with successful medication, enabling feedback systems for medication level control and evaluation of long term effects in treatments.

4.3 SA3L (Situation Aware Ambient Assisted Living)

The SA3L project is concerned with developing methods and tools for detection and interpretation of potentially dangerous situations in the home of elderly people. The situations are inherently difficult to specify and generalize due to the diversity of homes, behaviors and the numerous ways of deviating. Thus, a system for detecting deviations based on manually specified rules is difficult (impossible) to build and maintain. We approach this problem by learning the activity patterns in the home.

In the data layer, a new method for representing time-dependent patterns of binary sensor deployed in-homes was proposed and used for modelling human activity patterns. To address the plasticity-stability problem were no prior assumptions made regarding the relevant sensors or the spatio-temporal relations between sensors.

The project has contributed to the modelling of human in-home activity patterns with an unsupervised approach to compare, cluster and relate (in space and time) similar behaviors. Deviations from such normal models are distinguished by indirect if-then rules and thereby contributing to both the information and knowledge layers of the knowledge pyramid. The method has been shown to work both for simulated and real data (in a demonstrator environment).A focus has also been on building up a realistic smart home simulator, in cooperation with Ulster University.

4.4 MoveApp

The goal of the MoveApp project was to develop mobile and wearable systems to support self-management of chronic conditions characterized by motor symptoms. Research activities have focused on the data level of the information pyramid. In particular to find appropriate representations for accelerometer data; representations that facilitate power-efficient, on-line processing of the data onboard the smart watch for long periods of time. Data have been collected on subjects but the study is not analyzed yet.

4.5 Situation aware safety systems (SAS2)

The goal is to develop a system that goes beyond the requirements defined in the current safety standard for automated guided vehicles, and state-of-the-art, by introducing additional functionalities for safe detection of other objects and situations listed, identifying objects and estimating the trajectory of objects in the environment. Such a system would be more situation-aware since actions of moving agents could be foreseen and concerns can be made based on objects' identities (e.g. human or other truck/AGV).

The main research question is how to detect, estimate the trajectory of, and identify objects (and agents) in a warehouse environment, such that actions based on this information lead to safer and more efficient (in terms of productivity) AGV operation. The approach is to use a multi-layer map where each layer is more adapted to the specific purpose. We foresee at least three levels: semantic map (used for reasoning), geometric map (the layout of the environment used for e.g. planning) and spatial-temporal map (used for reaction and obstacle avoidance). A challenge is how interactions between different layers is done and how feedback from higher levels can, for example, be used to improve accuracy, consistency of different layers in the map; how to detect the difference between foreground (static and dynamic obstacles) or background model (static objects).

4.6 Long term multi-layer mapping

Mapping is a classical problem in robotics and is currently well understood how to solve this problem in static environments. However, in the long-term mapping problem it is not clear enough how to deal with changes in the environment and how to manage the scalability of the lifelong mapping process. Arrange maps in different layers depending on data, information and knowledge content (and/or application, e.g. localization, monitor events) helps in adding scalability. However, all these maps must be maintained, i.e. updated to account for changes in the environment. We can refer to this as lifelong mapping, to enable lifelong situation awareness.

The scientific contribution are: a long-term multiple-layer mapping system that accounts for changes in the environment and scalability issues to maintain global maps, acquired along a typical industrial vehicle operation, i.e. approximately 20 km of total path length; a multi-layer map architecture with a useful connotation in the industrial world; a mapping technique to model the dynamics in the environment; a map maintenance strategy to maintain compact but informative maps efficiently.

4.7 Intelligent environments supporting ageing at home

This project is a continuation of the SA3L project, and it aims at building models of human behavior patterns that can be generalized over different environments and individuals. Future capabilities of AAL involve reaction to an otherwise normal sensor reading if it happens concurrently with something else, preventing or alerting of unwanted events or abnormal patterns without the need of constant attention of an operator. Additionally, the user should be able to interact or to "speak" to the system, telling that s/he is aware of the situation (perhaps thanks to the alert). Early detection of diseases is another example of a promising application by monitoring activity patterns of elder people who wish to live in their own homes as long as possible, e.g. detecting a reduction in physical/sleeping activity or a fall. Research questions here include how to autonomously learn the habits of inhabitants, and thus be aware of their activities/status, and how to interact with the inhabitants.

In more detail in regard to the knowledge pyramid, data representations will be required, e.g., to keep track of individuals who interact with the robot. For the information layer, event recognition will occur in a simplified manner after event detection: e.g., an interacting person will be identified or marked as new if a face is detected, a touch such as a hug will be recognized or marked as new if a touch is detected. For the knowledge level, we will manually find features and parameters which could be useful from the literature, our own ideas, and watching interactions (e.g., a person might talk more if the interaction is proceeding well) as a starting point; parameters will be adapted by the robot during interactions based on evaluating success.

4.8 Knowledge discovery and data mining life logging data for fertility

The main contributions of this study are related to the data and information layer of the pyramid. One important challenge in this application stems from the fact that each individual is unique, presenting individual cycles and individual hormonal levels. Therefore, models based on an average representation of the fertility cycle are not accurate for the majority of individuals. The research focus is on developing models that capture domain knowledge, such as information about ovary fertility cycles, but automatically adapt to each individual, providing more accurate information. This requires designing robust or adaptable features.

4.9 Knowledge creation from data streams and service operations

We extend our collaboration with Toyota Material Handling Europe, from autonomous vehicles into the uptime and predictive maintenance of forklift trucks. The planned research contributions relate to all the stages in the knowledge pyramid. In the data level we will continue our research concerning automatically evaluating which signals that are most interesting to monitor, and how to find appropriate representations for various types of data. In the information level of the pyramid we need methods for grouping systems, automatically, into clusters based on various aspects such as configuration, usage or condition. The fleet of vehicles will be heterogeneous, and we will need to investigate new methods for describing operation (both the commonalities and differences). It is a challenge to find methods that can automatically, or with minimal support from human experts, decide on the most interesting configurations to focus on. This is additionally complicated by issues such as concept drift, seasonal variations, application changes, and so on, which all need to be automatically detected and taken into account. In the knowledge level, we will contribute with algorithms related to how machines and humans can jointly create knowledge from information.

The awareness requires analyzing different clues when looking for faults, and in this project we will develop an integrated framework for using several approaches: identification of "normal" or "expected" behavior and detecting deviations from that; finding out common patterns between issues that have proven problematic in the past and looking for new situations that are similar; characterizing events and time points when the condition changes for some reason. By building at all stages of the knowledge pyramid we will also develop online and incremental algorithms for decision support based on very little data for early detection of issues. Those algorithms will be looking at streams of data and processing information as soon as it arrives, giving initial warnings as soon as possible and then refining those decisions as more information becomes available.

4.10 Image analysis

There is a substantial research activity in CAISR directed at image analysis, mostly with a focus on the data and information levels. The most prominent research contribution regards the design of robust, general, translation, rotation and scale invariant features for images.

Knowing where humans are (presence) or who they are (identity) can be fundamental in aware intelligent systems not only for individual events (e.g. "person A is here"), but for the construction and understanding of bigger pictures in critical situations, for example "who is where", "is the right person in the right place (and not wounded)?", or "is a (maybe harmful) person where should not be?" Analysis of human activity over a certain period (something which involves continuous detection) is also necessary to provide information about what a person is doing, either over a short period (activity event) or a longer period (activity pattern).

Image analysis-based environmental monitoring, e.g. studies of long-term changes in aquatic ecosystems, assessment of water quality parameters, is another application area of aware intelligent systems addressed in our research. Semi-supervised training applied to build a model for machine-learned ranking allows assessing automatically the quality of intermediate results produced by the system itself and greatly reduces the risk of propagating intermediate errors into next processing steps.

4.11 Healthcare technology

Laryngology is a healthcare area with great potential for application of aware intelligent systems.Our research here touches on all four levels of the knowledge pyramid. A number of representations of varying granularity and complexity are created from raw voice data and then used for adaptive modelling, which enables having a data point specific model for each data point to be processed. Knowledge extracted from voice data is enriched with knowledge gathered in a set of association rules elicited from subjective self-evaluations/self-assessments using affinity analysis. The set of very simple rules together with self-organized 2D maps extracted for voice data-based models help clinicians gaining comprehensible insights (understanding) concerning specific cases as well as trends and important associations.

4.12 Self monitoring systems

Together with Volvo Technology we have been pursuing ideas on self-monitoring for several years now, focusing on the area of uptime and predictive maintenance. Due to the distributed nature of the vehicle fleet and communication limitations, we have developed methods for modelling the data and automatically finding the most interesting signals to focus on. A city bus is a very complex system and automatically building information about its operation, from the data, is beyond current state-of-the-art. We have published new algorithms for analyzing behavior of the fleet of vehicles and detecting anomalous individuals. Based on outliers detected this way, in combination with historical repair information, we have created knowledge about the state of health of a given vehicle, as well as whether and when it requires a workshop visit.

We have also looked at combining more sources of data. This brings its own issues related to, among other things, awareness of the reliability of each source and ways to resolve inconsistencies between them.

In addition, we have started the BIDAF project, in collaboration with SICS (project coordinator) and University of Skovde, about realizing the promise of advanced, near real-time analytics on uncertain data with high volume and velocity through machine learning techniques. Key challenges include development of a computational platform; machine learning algorithms suitable for handling massive data; analytics methodology for automatic creation of information, knowledge and understanding.

4.13 Robotics

The AIR project (Action and Intention Recognition in Human Interaction with Autonomous Systems) is a project in cooperation with Skovde University (project leader), Orebro University, and the Viktoria Institute. The CAISR contributions in the AIR project explore the use of novel forms of data (e.g. using breath sensors to allow a robot to detect a person's location without using possibly intrusive sensors capable of identifying individuals), as well as deriving rules to generate typical classes of behavior (such as playful motions) and predicting how a person will perceive such behavior over time to achieve good interactions. For the latter, knowledge creation will involve deriving an initial model from expert knowledge refined by data from real interactions (hand-coding by humans) and adapting the model autonomously to accommodate individual preferences by the robot.

References

[1] Hintikka, J., "Impossible Possible Worlds Vindicated", Journal of Philosophical Logic, 4, pp. 475-484 (1975).

[2] Fagin, R. and Halpern, J.Y., “Belief, Awareness, and Limited Reasoning”, Artificial Intelligence, 34, pp. 39-76 (1988).
[3] Modica, S. and Rustichini, A., “Awareness and Partitional Information Structures”, Theory and Decision, 37, pp. 107-124 (1994).
[4] Devanur, N., and Fortnow, L, “A computational theory of awareness and decision making”. In Proceedings of the 12th Conference on Theoretical Aspects of Rationality and Knowledge, pp 99-107 (2009)
[5] Endsley, M.R., “Toward a Theory of Situation Awareness in Dynamic Systems”, Human Factors, 37, pp. 32-64 (1995).
[6] Gutwin, C., and Greenberg, S., “A Descriptive Framework of Workspace Awareness for Real-Time Groupware”, Computer Supported Cooperative Work, 11, pp. 411-446 (2002).
[7] Zhao, Q, Hsieh, C-H, Naruse, K, and She, Z., “Awareness Science and Engineering” (editorial), Applied Computational Intelligence and Soft Computing, 2012 (2012).
[8] Zhao, Q., “Computational Awareness: Another Way towards Intelligence”, Computational Intelligence (Eds. Madani et al.), Book Series Studies in Computational Intelligence, 465, pp. 3-14 (2013).
[9] Ackoff, R.L., “From data to wisdom”, Journal of Applied Systems Analysis, 16, pp. 3-9 (1989).
[10] Degani, A., Jorgensen, C, Iverson, D., Shafto, M. and Olson, L., “On Organization of Information: Approach and Early Work”, Technical Report NASA/TM2009215368, NASA AMES (2009)
[11] Rowley, J., “The wisdom hierarchy: representations of the DIKW hierarchy”, 33, pp. 163-180 (2007).
[12] Curtis, G., and Cobham, D., Business Information Systems: Analysis, Design and Practice, FT Prentice Hall, Harlow (2005).
[13] Horeis, T., and Sick, B., “Collaborative Knowledge Discovery & Data Mining: From Knowledge to Experience”, IEEE Symposium on Computational Intelligence and Data Mining 2007 (CIDM 2007), pp. 421-428 (2007)
[14] Zhou, L., “Ontology learning: state of the art and open issues”, Information Technology Management, 8, pp. 241-252 (2007).
[15] Barforush, A.A., and Rahnama, A., “Ontology learning: revisited”, Journal of Web Engineering, 11, pp. 269-289 (2012).
[16] Kanter, M. J., and Veeramachaneni, K., “Deep Feature Synthesis: Towards Automating Data Science Endeavors”, International Conference on Data Science and Advanced Analytics (DSAA), pp, 1-10 (2015)
[17] Fogelman-Souli, F., and Marcad, E., “Industrial Mining of Massive Data Sets”, Mining Massive Data Sets for Security, (NATO ASI Workshop), IOS Press, pp. 44-61 (2008).
[18] Khosla, V., “20-percent doctor included: Speculations and musings of a technology optimist” (2014)
[19] Reddering, K., and Scholten, L., “Understanding Human Cultures”, Chapter 1.2 in The New Everyday: Views on Ambient Intelligence (Eds. Aarts, E., and Marzano, S.), 010 Publishers, Rotterdam (2003)
[20] Cheng, W., Kasneci, G., Graepel, T., Stern, D., and Herbrich, R., “Automated feature generation from structured knowledge”, Proceedings of the 20th ACM Conference on Information and Knowledge Management, ACM, pp. 1395-1404 (2011).
[21] Paulheim, H., and Frnkranz, J., “Unsupervised generation of data mining features from linked open data”, Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics (WIMS12), ACM, pp. 31:131:12 (2012).
[22] Bengio, Y., Courville, A., and Vincent, P., “Representation Learning: A Review and New Perspectives”, IEEE Transactions on Pattern Analysis and Machine Intelligence (special issue Learning Deep Architectures), 35, pp. 1798-1828 (2013).
[23] Bigun, J., Vision with direction, Springer Berlin (2006).
[24] Zhang, Y., Zhang, L., Nie, G. and Shi, Y., “A Survey of Interestingness Measures for Association Rules”, 2009 International Conference on Business Intelligence and Financial Engineering, IEEE (2009).
[25] von Luxburg, U., Williamson, R. C. and Guyon, I., “Clustering: Science or Art?”, Journal of Machine Learning Research (JMLR): Workshop and Conference Proceedings, 27, pp. 65-79 (2012)
[26] Iverson, D. L., “Inductive monitoring system constructed from nominal system data and its use in realtime system monitoring”, Patent US 7,383,238 B1 (2008).
[27] Angelov, P., Autonomous Learning Systems, John Wiley & Sons (2013).