Learning machines ID2225

Course examination

# Relational reasoning for learning machines


## Course Coordinator

Magnus Boman <mab@kth.se>

## Author

Óttar Guðmundsson <ottarg@kth.se>



### Abstract

A learning machine (LM) is a novel concept inspired by Alan Turing's vision of Intelligence Machinery. The project Internet psychiatry aims to implement a LM for Karolinska Institutet. One of the challenges of the project is to detect patients at risk based on multiple inputs. This paper suggests two networks, Capsule Networks (CN) and Relational Networks (RN), that could potentially lay down the groundwork for relational reasoning between classifications from all input parameters. Pros and cons of both networks are discussed and RN is concluded as a promising solution to the problem, due to its ongoing development and plug-and-play capabilities.

# Table of contents

# Preface

When I started the second year in my master program in Software Engineering of Distributed Systems, I still did not have any idea of what I wanted to write about in my thesis. My studies were a hybrid of distributed fault-tolerant systems, data science, and neural networks so finding a topic was an interesting task but a difficult one.

After watching professor Magnus Boman's guest lecture in Research Methodology on quality assurance, he introduced his research on epidemiology and it interested me. Some of his research included work on events that I had experienced, for example the Corrupted blood incident [1]. After contacting him and telling him more about my ideas, he mentioned that he was indeed looking for someone to work for him on a project for Karolinska Institutet. The project was called Internet psychiatry, its objective was to design an architecture called a Learning Machine (LM) to monitor patients through an internet platform. This sounded exciting and was relevant to my studies. He wanted me to start as early as December, but my master thesis was not supposed to start until mid-January. Upon further discussions, I learned that he was also teaching the course ID2225 about the subject, where the final task was to write a small project or an essay about the LMs. We both agreed that attending the course would be a good start to my thesis work. I could start the research of a module for his architecture that could in some way fuse or combine information from different data sources to help detect and act on anomalies within the data, so-called "smart flags".

As such, this paper serves as a starting point for my final thesis which name is yet to be decided. It explores the possibility of implementing Capsule or Relational networks or some combination of the two, as a component into a LM. I hope that this work will guide other developers in the near future and be a valuable contribution to the Internet psychiatry project.

# Introduction

An old and unpublished article called “Intelligent Machinery” by Alan Turing et.al [2] described the idea of designing intelligent robots that could wander around the countryside and observe whatever they perceived, continuously learning and improving based on their experience. In the past years, development in Artificial Intelligence (AI) has improved significantly with deep learning architecture using deep neural networks. Convolutional Neural Network (CNN) are capable of extracting features from images and Recurrent Neural Network (RNN) extract the dynamic behavior of a time sequence. These networks are purely a subset of Turing’s vision, as they are only good for a single defined task in a limited domain. Noticeably, most modern AI solutions can only be considered as weak AI like Alexa and Siri [3].

A *learning machine* (LM) is a step towards Turing’s ideas, defined as an “autonomous self-regulating open reasoning machine that actively learns in an unsupervised and decentralized manner, over multiple domains” [4]. Its description is closer to a strong AI or artificial general intelligence (AGI), as it can adapt to new environments and act upon them based on prior knowledge. This type of machine will be implemented and experimented on, in a project called Internet psychiatry.

# Problem statement

Most of the neural network architecture of today has been designed in such a way that it can only be considered as a weak AI. Most networks integrated into production are purposely trained on a well-defined and clean data set, learning how to classify or predict a task within a limited domain. If the definition of a LM has to hold, there are numerous problems with the current approach. One large issue is data relativity and how multiple domains or data sources act upon each other. A LM could be listening to audio, viewing a video, analyzing movement and reading stock market stream, all at the same time. Possibly two of those data sources could be correlated or have a causal relationship that is undetected by the machine. How can such a machine make sense out of all the perceived data derived from multiple domains, in order to see the bigger picture? That requires some sort of relative reasoning but that is easier said than done.

## A short history of logic and reasoning in AI

Mathematical logic is a vital part of computer science as it is often used to prove that a program or algorithm holds under certain conditions. The main benefit of logic is the reasoning for causation where if we define event A as true, then B is as well or will be at some point. The design of distributed systems is no exception as it relies heavily on logic, e.g. systems proved to be eventually correct. However, in this context it is important to note that causation is not the same as correlation.

In the early days of AI, logic-based programs were created to help with decision making by suggesting an answer based on a given input. These types of programs were mainly Decision Support Systems (DSS), Expert Systems (ES) and Knowledge-Based Systems (KBS) [5]. The latter systems were based on a database with facts and rules, a knowledge base that could be domain specific and an inference engine for strategies to combine facts and knowledge such as deductive, inductive and abductive reasoning. Furthermore, systems could be either monotonic or not, meaning that truthfulness of conclusions might change or not based on new input data. Managing these systems was difficult as it involved a lot of manual work, since developers needed to add rules to the database based on facts and knowledge extracted from professionals. Moreover, the systems had to be maintained in an ever-changing world. The systems' biggest flaws were their lack of creativity, imagination, and intuition. Their domain was limited by the computer system they were running on and the developer writing the logic it followed. One of the bigger problems was dealing with multiple contexts when it came to decision making.

A novel approach using Fuzzy logic [6] had the main goal of reasoning under uncertainty; the facts had a range of answers that were fuzzy as they had an upper and lower bound. The Dempster Schaffer theory [7] was an extension of this concept, using a range of answers and managing to combine evidence from different sources. It combined degrees of belief based on the number and the probability of answers and as such the system was closer to being correlation based than causation based decision making. This method improved the mentioned systems but was still limited due to the nature of their maintainability. Even though KBS were able to switch between domains by switching their knowledge base, it still needed to be configured manually.

## Reasoning in deep neural networks

Reasoning is an important component of AI but it has not been successful for deep neural networks (DNN). The most successful architectures in DNN have focused on correlational analysis for classification. As of today, most neural network architectures used in production are CNN's. These networks capture features of images but not the features relativity and their relations to other data sources like data and knowledge fusion for the deep web [8]. Let us imagine a CNN trained to identify images and one of the classifications is a face. In the training phase the network captures the unique features of a face such as the eyes, nose, and a mouth but does not capture the construction of the face and how the features are related to one another. As such, the network would classify both images seen in Figure 1 as a face.

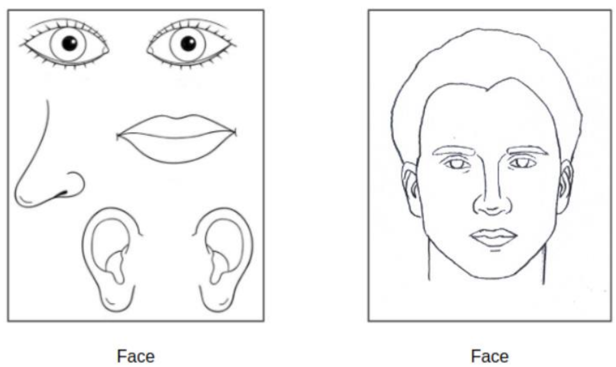



Figure 1: Two images that contain features of a face. Both of them might be classified as a face by a simple CNN that fails to take into account the important spatial hierarchies between features and their lack of rotational invariance. The image obtained from [9].

The results in this bizarre case show that if an image had all of those features scrambled together like a modern art piece, the network would classify this as a face when it clearly is not. It misses the knowledge of how facial features are related to one another and how they define the face as a whole.

The same problem appears when using a combination of different data sources. When a human notices an explosion in the far distance, it first sees the actual explosion and then it hears it as light travels faster than sound. Without a problem, the human brain connects the two together and infers that the sound it just heard a moment later is from the same explosion it saw prior. The incident involves three dimensions of data (visual, audio, and time) that we humans can automatically connect together as we understand their relationship. However, for a LM, combining these multiple inputs and making the correct assessment would prove more difficult.

An even more complex problem would be for the LM assessing the context of the two different scenarios displayed in Figure 2. In the upper scenario the LM classifies the man as being nice and positive while the in the lower scenario he is angry and negative. This is based on multiple inputs received at the same time where the LM connected key components of each source and how they are related to one another.

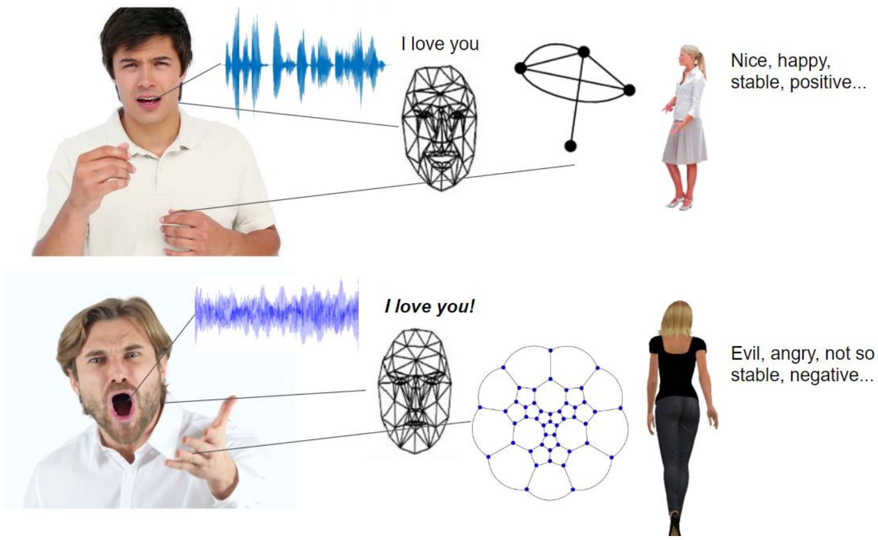



Figure 2: In the upper case, the man is speaking in a calm, soothing voice telling a woman he loves her. His facial expression implies that he is happy and his hand movements are calm and easy. The woman stays to listen to the man. This scenario can be perceived as being “positive”. In the lower scenario, the man yells that he loves the women, however, she walks away from him. His facial expression is angry and his hands are moving in a chaotic way. If a LM would simply classify the context based only on the content of his words both scenarios would be classified the same. However, by classifying the various data sources and assessing them together, the LM can really understand the nature of the situation.

It is important that each individual component of a LM interoperate in order for the LM to make correct assessments of the nature of scenarios. This is in line with Aristotle’s words, “the whole is greater than the sum of its parts” [10] referring to the fact that the value created by the parts working together is higher than just aggregating them individually. That said, we need to look into what type of method could be implemented into a LM to handle multiple data sources and their relativity to one another.

# New sublayer to learning machine architecture

Introducing a sublayer, between the perception and the reasoning layer, that could contribute to a solution to the problem mentioned in great length in the prior chapter. So far, the solution has no name but as the name of this essay suggest, it is somewhere along the lines of *Relational Reasoning*.

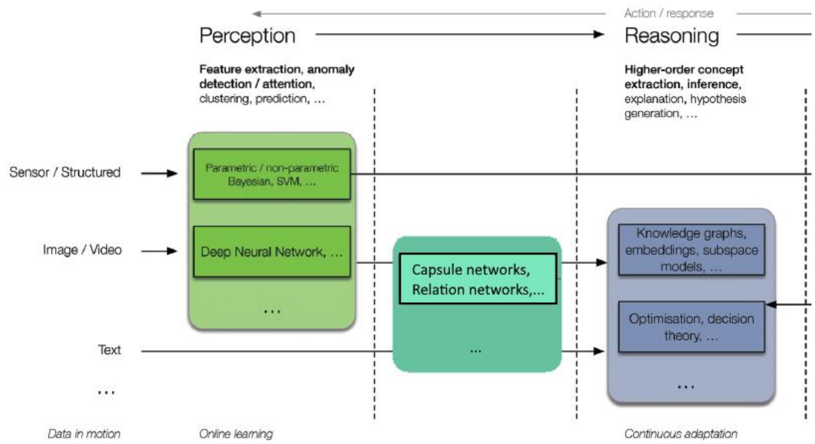



Figure 3: A modified image of the conceptually split architecture of LM.

In the following section, two networks are discussed as a possible part of either the perception and/or the reasoning layer of the LM. By using neural networks, an automatic extraction of important features and relations might be achieved without the manual work needed to configure logic in KBS. The difference between both networks and the traditional classification of CNN is that they map relations to other objects that they have learned to identify. Further research is needed on both networks but they show promising results for a few cherry-picked data sets. The new techniques could provide value for reasoning on classification for different data sources and that subject will be investigated in more depth in my thesis work.

## Capsule networks

In November 2017, Hinton et al. [11] proposed a new type of network that could potentially solve the problems of standard CNN. Introducing a new concept to networks called Capsules, they output a vector instead of a single value or matrix, allowing the possibility to choose a parent in the layer above the capsule that it is sent to. For each potential parent, the capsule network can increase or decrease the connection strength for the identified object called routing by agreement as seen in

Figure 4. This method is supposedly much more effective at adding invariance than the primitive routing introduced by max-pooling.

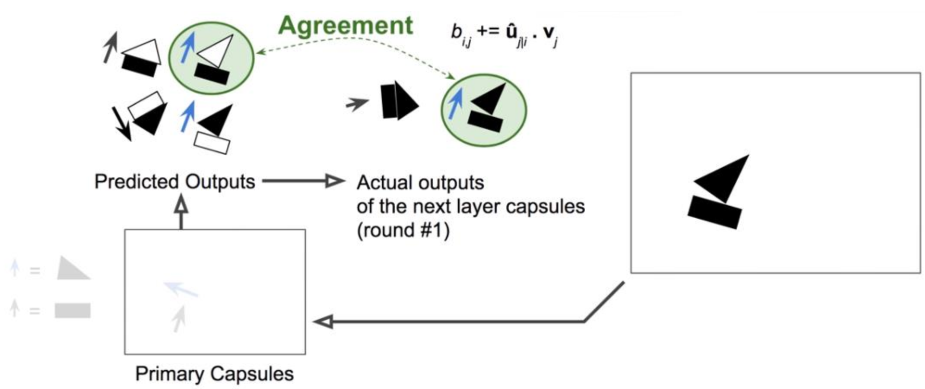



Figure 4: A demonstration of the vector outputs predicting a boat as demonstrated by Aurélien Géron [12].

By outputting vectors of the identified features in Figure 4, the capsules pick the parent with the strongest connection for the first iteration. In further iterations, it manages to combine higher level routed features to form an object that will finally be classified by the network.

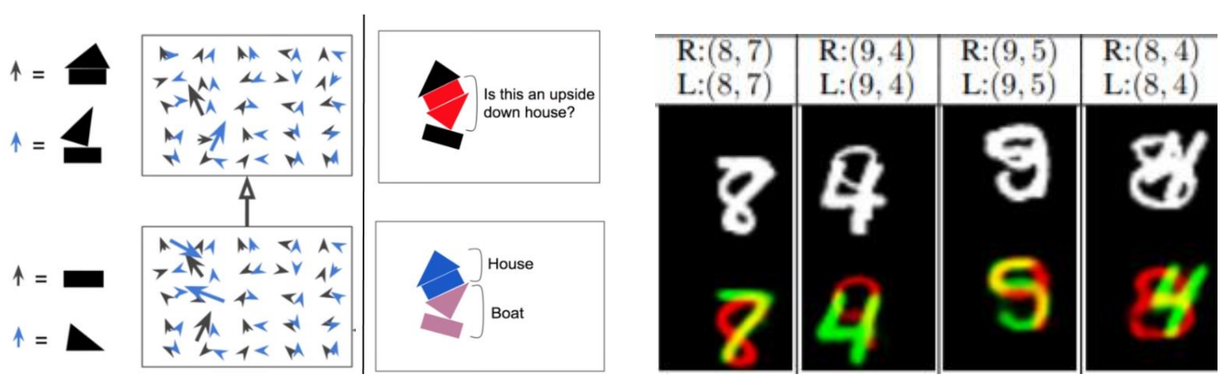



Figure 5: To the left, one of the problems is presented in a simplified case. If two features overlap that might confuse traditional networks, the routing algorithm still manages to capture both objects as the agreement factor for a house and a boat is stronger than an upside-down house. On the right, two digits have been stacked on top of each other presenting a cluttered problem. The capsule network successfully captures the features of both objects and colors the different relative features to their object.

### Pros and cons

The new concept that capsule networks present have numerous benefits, making it unique compared to an MLP and CNN. In the original paper, an accuracy of 99.8% is achieved on the MNIST dataset [13] which is the state of the art and got close to 90% on CIFAR10 [14] by increasing the number of primary capsules and routing rounds. Fewer data points are needed to train the network, as data augmentation methods like image rotation and flipping are not needed since the network captures rotation, thickness, and skewing. As the routing by agreement technique considers the preserved position and the pose of features, it can account for overlapping objects as demonstrated in Figure 5. The downsides are that this structure does not work well on complex data as seen on CIFAR10. One of the big issues is that capsules try to account for everything in the image [15], so when the backgrounds are too varied, the network does not manage to capture a good model to account for all the backgrounds. The network has not been tested on ImageNet and bigger images, which might result in disappointing results just as it did with CIFAR100 where it reached only 18% accuracy [16]. Furthermore, the training is considerably slower due to the inner loop of the algorithm when applying the routing by agreement.

## Relational networks

In the summer of 2017, Deepmind released a paper about a new type neural network called Relation Network (RN), capable of relational reasoning, meaning they could learn to understand relations between objects [17] which is an important characteristic of intelligence. An example of a relational question can be identifying an object without explicitly mentioning it. The authors provide a great example in Figure 6 to demonstrate the concept.

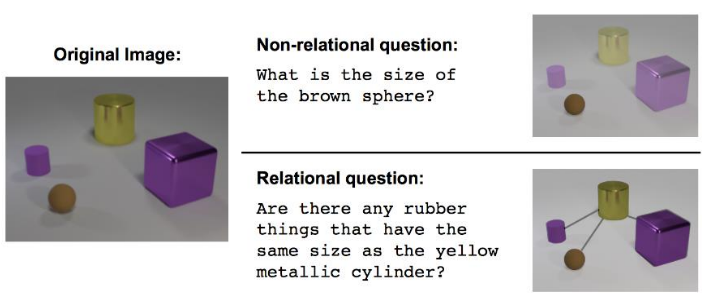



Figure 6: Two questions asking about objects from the CLEVR dataset [18]. The difference between a regular question and a relational one is made clearer in the context of the image.

As mentioned before, the problem with CNN is that they only capture the individual features of images, while the RN manages to capture relations between those features in a different manner than the former proposed architecture using capsules. The architecture behind the relational networks is defined as fθ for O is

$$\mathrm{RN}(O) = f_\phi\left(\sum_{i,j} g_\theta(o_i, o_j)\right)$$

where the O is a set of objects where relations between the objects should be learned. The gθ is a network that takes in two objects as a parameter and outputs the relation that we are interested in. For all objects in the set, the total gθ is calculated for all pairs and summed up. In the paper, the network is constructed as a module that can accept any encoded objects to learn their relations. Furthermore, the module can be combined with other types of neural network architecture such as CNN or LSTM as they provide in the paper as seen in Figure 7.

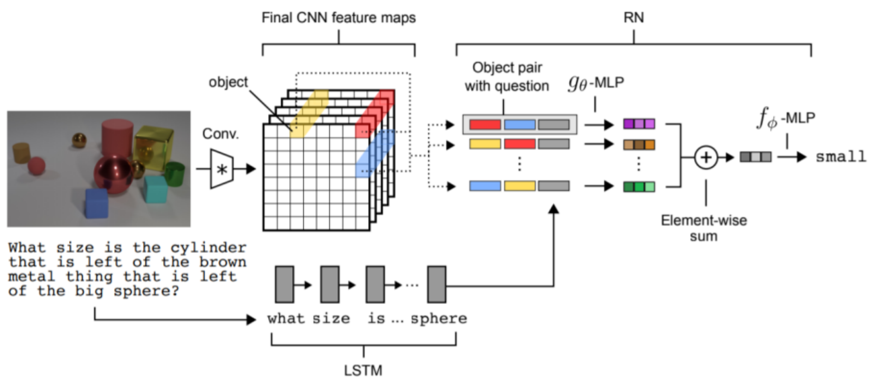
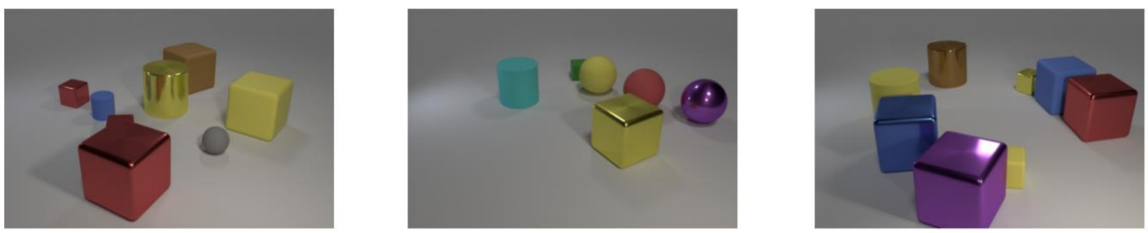



Figure 7: The structure of the network presented in the paper to answer text-based queries about an input image.

In the implementation, the CNN is used to learn features of the images while the LSTMs are used to understand the meaning of a query that is presented to the model. By combining the two networks together inside the RN, an end-to-end neural network can learn relations between the encoded objects in an image and answer queries about their relations.

## Pros and cons

According to the paper, RNs show promising result in how good they are at learning how to infer relations and do so in a data efficient manner. By designing them as modules that can be plugged into external architecture, they are quite flexible and can easily be used for CNNs and LSTMs. The combination of both of them, plus the RN, scores overall 95.5% accuracy for different datasets - scoring higher than humans do in some of the presented test cases. The most interesting thing is that this combination manages to connect text to an image, which might work for other encoded data sources as well. Further research is needed on this network and its usability with other types of networks such as Restricted Boltzmann Machines (RBM) [19] or Autoencoders (AE) [20]. The failed test cases were also demonstrated and hinted that some improvement is needed, see Figure 8. The authors also demonstrate the network's flexibility as seen in Figure 9, capturing relations of moving objects in a video.



| | What shape is the small object that is in front of the yellow matte thing and behind the gray sphere? | What number of things are either tiny green rubber objects or shiny things that are behind the big metal block? | What number of objects are blocks that are in front of the large red cube or green balls? |
|---|---|---|---|
| *RN:* | cylinder | 1 | 2 |
| *GT:* | cube | 2 | 3 |

Figure 8: Test cases where the network failed to answer the correct question.
In its defense, the questions are actually quite hard to read.

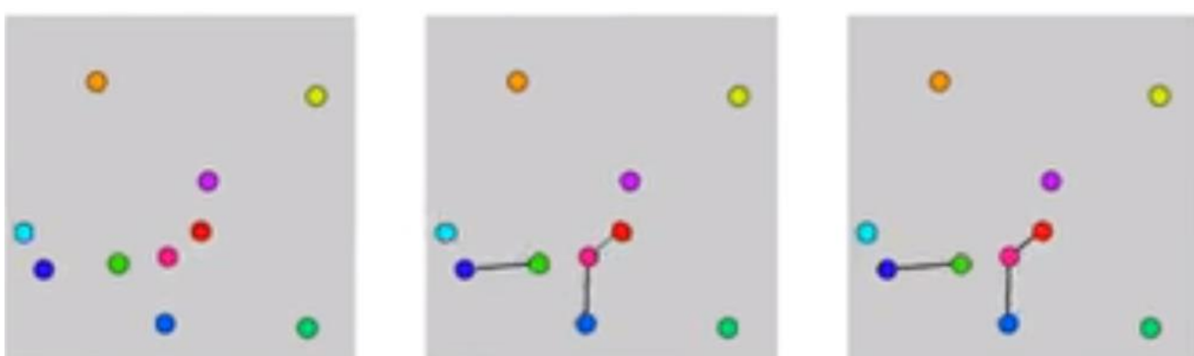


Figure 9: On the far left, the input data of moving dots are presented to an RN. In the middle, the predicted connectivity between dots that move together are linked and the ground truth is displayed on the far right. This suggests that RN is able to reason in different context, not only still images and text. The image is taken from David Raposo's Youtube channel [23].

One of the noticeable problems with RNs is that for each recognized object, the network can only answer questions about the relations of that individual object. It cannot infer relations of relations, such as object A relations to object B, and object B relations to object C, to find out that A leads to C. In the proposed network, object A has to have a direct relation to object C to answer anything about it, as it cannot imply the relations out of its own established relations. However, a new type of network called Recurrent Relation Network (RRN) [21] effectively solves this limitation, performing multiple reasoning steps between objects. This allows relations and actions to be propagated from object to object to form a chain of interactions. The RRN can be seen as an extension of RNs, solving RN's problems and reaching 96.6% accuracy on complex problems such as Sudoku. Further development has been conducted on the RN creating a network called multiRN that uses Broadcasting Convolutional Network [22], hinting that the RN is gaining attention and a larger fan base. This paper will not discuss RRN nor multiRN in more detail, however such an in-depth discussion could be of value to the LMs as they seem to capture and resolve more complex reasoning problems.

# Discussions

The purpose of this paper is to explore the potential use of the mentioned neural network architectures that might be a beneficial component for a LM. The networks could help the LM capture, reason and understand how multiple data sources work together. The first proposed network based its reasoning on the routing by agreement algorithm that proved successful for overlaid objects and cluttered images, while the second one focused on reasoning between networks, being able to combine classification for different data dimensions. By incorporating some sort of model that can reason about features of features for single data input and further reason about different classifications from multiple inputs, it might help to identify and understand the wider context of all data observed.

## A component to learning machines

Even though both types of networks seem to be proper candidates to a LM there are still arguments about which layer they belong to and if they work for the definition proposed in the original paper.

As demonstrated in Figure 3, both models could provide to be a good fit somewhere between the perception and the reasoning layer. The reason why it is hard to decide the layer is that logically, they use some sort of perception as an input to reason. This needs to be further investigated, to see how far the networks stretch their functions conceptually. Also, both papers did not introduce any method for online training as both networks are trained on batches of data. This is a problem for a LM as the definition states that it should be continuously learning, meaning that it cannot be trained on a single data set and be ready to take on the world. Some modifications would be needed for both architectures to work, to be able to process and improve upon newly perceived data for stochastic training.

## Combine the networks with more networks

Both networks provide a new method to capture reasoning of neural networks in their own unique way. The benefit of the RN is that it can be integrated with other types of networks, so combining it with multiple capsules might be an interesting approach. As the capsule network might be able to provide relational reasoning for a single input to the LM, the RN could serve as a wrapper for all other data inputs. Furthermore, a LM can have multiple types of neural networks installed as components. It would be interesting research to test the RN with multiple types such as RBM or AE and see if that has any benefits at all. If successful, the RN could serve as overlay for all neural networks' components.

## Research potential for other data sources

The capsule network only demonstrated its effectiveness for the MNIST dataset and mentioned its problems regarding ImageNet, suggesting that capsules should mainly consider images as an input. It did not mention using the architecture for any other type of data such as audio or text, meaning that there is a whole lot of work to be done. In the paper, the features of an image captured by a CNN are mapped together. These features are used as an input to the primary capsule, which manages the vector directions for classifications. In modern text classification such as Word2Vec [24], the relations of words are also captured using vectors. Maybe capsules can serve as a new technique for other data classification, such as NLP. If not, data points can still be plotted into images, like audio files as spectrograms, and trained using capsule networks. For the RNs, both examples demonstrated the usage of combining text queries with detected features of an image.

The authors mention that it is indeed plug-and-play for different networks, so combining them with e.g. audio to time series or financial data to twitter statistics might be an interesting approach.

## The black box problem

One of the problems using neural networks, in general, is that it is hard for humans to understand why the network came to its conclusion. Matrix multiplication and mathematical equations are not visually engaging nor intuitive, but luckily researchers have recently started to focus more on understanding the concept by visualizing how a network learns [25] or by plotting up the high dimensional space [26]. One of the major advantages of the logic and reasoning methods used in the early DSS and KBS is that the inferring method was often parsed in a tree, which is easier for us humans to interpret. By doing so, it is clearer seeing a formula that answers something like A -> B ^ C = D (where variables have some sort of meaning) rather than looking at multiple matrices with floating numbers. If the reasoning is supposed to be implemented into LM, some sort of interface with explainable visuals should be developed to understand why the machine reasoned the way it did.

# Relevance to the KI project

The KI project is called Internet Psychiatry, and in the simplest description, is about installing a LM into a room full of psychiatrists. The LM will help the staff at Karolinska Institutet to monitor patients through a platform to detect symptoms for mental illness e.g. anxiety and depression. This platform has these patients to perform multiple tests through a period of twelve weeks to see if they are getting better or worse. The main goal for the LM is to detect anomalies or obscure behaviors of the patients, and notify the staff if observed patients require immediate attention.

The next step in the project is to start designing the architecture for the LM, as all of the data has already been gathered and cleaned up. This data is a combination of multiple parameters such as patients messages with the psychiatrists, a test they are to perform, self-evaluation on how they are feeling and so on. If this is connected back to the example shown in Figure 2, the same solution is required to understand the bigger picture of all data inputs, as in both cases the inputs to the LM can be as different as they are many. The reason why capsules or RN might be of value to this project is that the patient can write a message to a psychiatrist that might be related to the outcome of his self-evaluation. Or, in the case of a message not being in relation to his self-evaluation, his condition might be critical, having him holding misbeliefs or being severely depressed. This could be an anomaly that the LM could pick up, and notify the Karolinska staff about the patient. Both of the proposed networks might serve as a solution to manage these parameters for the KI project, or hint at similar methods where merging multiple classifiable inputs is essential.

# Conclusion

This paper introduced and assessed two newly developed neural network architectures that are components of a LM. The investigation served as a preparation for my thesis work, which is part of a larger project called Internet psychiatry. The discussion focused on the description and speculation of the two networks and stating their pros and cons. As it involves no experimental results, I do not try to conclude with which network is better. The network’s testing on the projects data and building the architecture for the LM is left for future work.

I assume RN might be a starting point for such experimental work as RNs are being actively development and they are able to adapt to different types of encoded data. The ability to be plugged with or into new networks in the form of a module makes RNs adaptable with other components, which is one of the major concepts of a LM. The capsule networks have yet to solve the problem of reasoning of complex data as they are still quite primitive and have only been successful on selected data sets and for simple images, however, they could potentially be used as a subcomponent.

Future work for the project should involve the creation and evaluation of both models and they should run over multiple data sources for the LM to give a better insight into the advantages and disadvantages regarding relativity. This would be especially valuable for researchers continuing the development of the project’s LM. Finally, I hope that my work will be useful for developers considering relational reasoning and be a valuable contribution to the Internet psychiatry project.

## Reference(s)

[1] Boman, M. and Johansson, S. (2007) Modeling Epidemic Spread in Synthetic Populations - Virtual Plagues in Massively Multiplayer Online Games *Proc Digital Games Research Association (DiGRA),* Tokyo, September 2007. [Online]. Available: http://www.digra.org/wp-content/uploads/digital-library/07311.21412.pdf. [Accessed: 17-Dec-2018].

[2] A.M. Turing. Intelligent machinery. In B. Meltzer and D. Michie, editors, Machine Intelligence 5, pages 3–23. Edinburgh University Press, Edinburgh, 1969. Written on September 1947 and submitted in 1948 to the National Physical Laboratory

[3] "What's the Difference Between Weak and Strong AI?," *Machine Design*, 15-Feb-2017. [Online]. Available: https://www.machinedesign.com/robotics/what-s-difference-between-weak-and-strong-ai. [Accessed: 17-Dec-2018].

[4] M. Boman, M. Sahlgren, O. Gornerup, D. Gillblad, "Learning Machines" The 2018 AAAI Spring Symposium Series, 2018 [Online]. Available: https://www.aaai.org/ocs/index.php/SSS/SSS18/paper/download/17482/15508. [Accessed: 04-Dec-2018]

[5] M. K. Bergman, "Knowledge-based Artificial Intelligence," AI3::: Adaptive Information, 17-Nov-2014. [Online]. Available: http://www.mkbergman.com/1816/knowledge-based-artificial-intelligence/. [Accessed: 17-Dec-2018].

[6] "Fuzzy logic - New World Encyclopedia." [Online]. Available: http://www.newworldencyclopedia.org/entry/Fuzzy_logic. [Accessed: 17-Dec-2018].

[7] M. Beynon, B. Curry, and P. Morgan, "The Dempster–Shafer theory of evidence: an alternative approach to multicriteria decision modeling," Omega, vol. 28, no. 1, pp. 37–50, Feb. 2000. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S030504839900033X. [Accessed: 17-Dec-2018].

[8] Dong, Xin Luna, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Kevin Murphy, Shaohua Sun, and Wei Zhang. „From Data Fusion to Knowledge Fusion", 1. March 2015. [Online]. Available: http://arxiv.org/abs/1503.00302. [Accessed: 04-Dec-2018]

[9] "Capsule Networks Explained | Kendrick Tan." [Online]. Available: https://kndrck.co/posts/capsule_networks_explained/. [Accessed: 17-Dec-2018].

[10] philosiblog, “The whole is more than the sum of its parts,” *philosiblog*, 17-Mar-2016. [Online]. Available: https://philosiblog.com/2016/03/17/the-whole-is-more-than-the-sum-of-its-parts/. [Accessed: 17-Dec-2018].

[11] Sabour, Sara, Nicholas Frosst, and Geoffrey E. Hinton. „Dynamic Routing Between Capsules“. arXiv:1710.09829 [cs], 26. október 2017. [Online]. Available: http://arxiv.org/abs/1710.09829. [Accessed: 04-Dec-2018]

[12] Aurélien Géron, *Capsule Networks (CapsNets) – Tutorial*. [Online]. Available: https://www.youtube.com/watch?v=pPN8d0E3900&list=WL&index=81&t=0s. [Accessed: 17-Dec-2018].

[13] “MNIST handwritten digit database, Yann LeCun, Corinna Cortes and Chris Burges.” [Online]. Available: http://yann.lecun.com/exdb/mnist/. [Accessed: 17-Dec-2018].

[14] “CIFAR-10 and CIFAR-100 datasets.” [Online]. Available: https://www.cs.toronto.edu/~kriz/cifar.html. [Accessed: 17-Dec-2018].

[15] E. Xi, S. Bing, and Y. Jin, “Capsule Network Performance on Complex Data,” *arXiv:1712.03480 [cs, stat]*, Dec. 2017. [Online]. Available: https://arxiv.org/pdf/1712.03480.pdf. [Accessed: 17-Dec-2018].

[16] R. Mukhometzianov and J. Carrillo, “CapsNet comparative performance evaluation for image classification,” *arXiv:1805.11195 [cs, stat]*, May 2018. [Online]. Available: https://arxiv.org/ftp/arxiv/papers/1805/1805.11195.pdf. [Accessed: 17-Dec-2018].

[17] Santoro, Adam, David Raposo, David G. T. Barrett, Mateusz Malinowski, Razvan Pascanu, Peter Battaglia, and Timothy Lillicrap. „A simple neural network module for relational reasoning“. arXiv:1706.01427 [cs], 5. June 2017. [Online]. Available: http://arxiv.org/abs/1706.01427 [Accessed: 04-Dec-2018]

[18] “CLEVR: A Diagnostic Dataset for Compositional Language and Elementary Visual Reasoning.” [Online]. Available: https://cs.stanford.edu/people/jcjohns/clevr/. [Accessed: 17-Dec-2018].

[19] A. Oppermann, “Deep Learning meets Physics: Restricted Boltzmann Machines Part I,” *Towards Data Science*, 27-Apr-2018. [Online]. Available: https://towardsdatascience.com/deep-learning-meets-physics-restricted-boltzmann-machines-part-i-6df5c4918c15. [Accessed: 17-Dec-2018].

[20] “Unsupervised Feature Learning and Deep Learning Tutorial.” [Online]. Available: http://ufldl.stanford.edu/tutorial/unsupervised/Autoencoders/. [Accessed: 17-Dec-2018].

[21] R. B. Palm, U. Paquet, and O. Winther, “Recurrent Relational Networks” *arXiv:1711.08028 [cs]*, Nov. 2017. [Online]. Available: https://arxiv.org/abs/1711.08028. [Accessed: 17-Dec-2018].

[22] S. Chang, J. Yang, S. Park, and N. Kwak, “Broadcasting Convolutional Network for Visual Relational Reasoning” *arXiv:1712.02517 [cs]*, Dec. 2017. [Online]. Available: https://arxiv.org/abs/1712.02517. [Accessed: 17-Dec-2018].

[23] David Raposo, *dynamic physical system, spring connections (late training)*. [Online]. Available: https://www.youtube.com/watch?v=5rd21AINarw. [Accessed: 17-Dec-2018].

[24] S. Banerjee, “Word2Vec — a baby step in Deep Learning but a giant leap towards Natural Language Processing,” *Medium*, 12-May-2018. [Online]. Available: https://medium.com/explore-artificial-intelligence/word2vec-a-baby-step-in-deep-learning-but-a-giant-leap-towards-natural-language-processing-40fe4e8602ba. [Accessed: 17-Dec-2018].

[25] N. Wolchover, “New Theory Cracks Open the Black Box of Deep Learning,” *Quanta Magazine*. [Online]. Available: https://www.quantamagazine.org/new-theory-cracks-open-the-black-box-of-deep-learning-20170921/. [Accessed: 17-Dec-2018].

[26] Google Developers, *A.I. Experiments: Visualizing High-Dimensional Space*. [Online]. Available: https://www.youtube.com/watch?v=wvsE8jm1GzE. [Accessed: 17-Dec-2018].