Research Proposal

Extraction and graph representation of assumptions in domain-specific images

Unubold Erdenebeleg

unuu0725@gmail.com/erdenebeleg.u.aa@m.titech.ac.jp

Department of Transdisciplinary Science and Engineering, Tokyo Institute of Technology

February 19, 2023

Abstract

In the field of computer vision, very common and successful assumption is 'locality'– nearby pixels in an image influence each other more than the pixels that are further away. In a sense, this assumption of locality helps the AI to specialize in images compared to other types of data. Then, how can we create stronger, more specialized assumption that helps the model in specific image domains, such as global maps, MRIs, or cat pictures? For starters, cats can be captured upside-down, big, or in different poses. On the other hand, global climate maps are always in a consistent format with number of regional features and region-region relationships. This means that we can assume some properties for climate maps, but not for cat pictures. In this research, the author will aim to introduce an **unsupervised** approach to represent such domain-specific assumptions as a graph, and demonstrate its practicality in climate studies.

By using dimensionality reduction techniques on tiles cut from the map, it was proposed to obtain node attributes. Furthermore, through either Markov Random Field or Recurrent Neural Network, the edge properties can be obtained for the graph. Specifically in climate studies, using a RNN could enable encoding natural phenomena in the graph. This will assist climate researchers through recommendation. Moreover, there might be possibilities of improved extreme weather prediction and better understanding of global circulation.

On top of assisting climate researchers in uncovering global scale phenomena, the approach described in this research could be relatively easily extended to other spatially 'consistent' image domains, such as medical imaging. Moreover, even for relatively challenging domains, it still has a potential to offer useful insights to AI researchers when designing learning architectures, by providing potential initial hypotheses based on the assumption for example.

1 Introduction

During my graduate studies, I would like to explore the potential of the extraction, representation, and most importantly, utilization of possible assumptions existing in the data. For this purpose, I would like to further my understanding of probabilistic graphical models, especially Markov random fields and inference problems. Next, after successfully completing the initial phase of my research, I hope to specialize in statistical frameworks for analyzing generalization, such as PAC-Bayes framework, and study the benefits of such assumptions. Considering the advantages of Convolutional Neural Networks over Fully Connected Networks, I hope to demonstrate the benefits of such extended assumptions through this specialization. Lastly, while I admit this ambition is in the future, I am most excited about working on designing learning models that are automatically informed about domain specific assumptions. The potential impact of such application on bringing AI closer to people who don't have enough data, especially smaller businesses, is one of the motivations for my graduate level studies.

I would like to start by addressing two problems in the field of computer vision and environment studies, respectively. Common statistical assumptions such as the locality in the simple grid structures of Convolutional Neural Networks are still quite generic and unable to directly address global scale phenomena that certainly exist the climate data, such as region to region relationships encapsulated in atmospheric circulation. This restriction is not specific to climate maps and similar relationship can be noticed in many different domains, as mentioned before. Furthermore, while Graph Neural Networks allow more complex structures, they still mainly use predesigned graphs which predictably results in restrictions regarding available data or compulsory human interaction. Ultimately, developers have to come up with various potential learning architectures based on their understanding of the data and improve through iteration. On the other hand, most climate studies rely on manual statistical analysis to uncover new phenomena. This requires the researchers to formulate hypotheses often through only observation. This approach, on top of being costly, potentially could leave out patterns that are difficult to be noticed by humans.

2 Literature Review

2.1 Climate studies

While currently there are no literature on machine learning applications for analyzing global climate patterns, Gibson et al. [2017] showed an application of clustering algorithm Self Organizing Maps for detecting extreme weather in the Australian region. However, such application of image clustering on a global scale climate map will be unsuitable, due to characteristics of some regions 'overpowering' others.

2.2 Image representation

There have been numerous different takes on representation learning on images as whole. Some of the most promising approaches rely on architectures with a 'bottleneck', which is used for eliminating insignificant details from the data. A noteworthy example is Invariant Information Clustering by Ji et al. [2018], which is based on maximization of mutual information between a datapoint and its image under random transformation (image flipping, shearing, etc.). While IIC is an extremely suitable approach for creating representations for the whole image, encoding grid to grid relationships is still problematic. Nevertheless, it would still be sensible to utilize information maximization for 'capturing' assumptions.
On the other hand, Noroozi and Favaro [2016] demonstrated the importance of spatial context through their Jigsaw solving model, Context Free Network. However, a high significance of the **absolute position** of a tile would carry no semantic meaning, making it an undesirable feature for the CFN. For our purpose, absolute position of the tile is still quite meaningful if associated with a particular appearance. But perhaps more importantly, CFN is not explainable for researchers in climate studies, whereas a graph can be easily visualized.
Lastly, there have been couple of studies related representing images as graphs. Namely, Han et al. [2022] developed a model called Vision GNN that is capable of representing an image as a graph where each node represents a grid. However, in this method, edges created between nodes are based on the distance between the nodes. Specifically, they have to be within each other's K nearest neighbors. Such connections that are solely based on similarity would not be able to infer any phenomena.

3 Research Ideas

Keeping the above general picture and my previous research experience in mind, I am currently planning to start by creating graph representation of the grid-to-grid relationships of the global climate map data. To clarify, in this graph (*Fig 1*), nodes' features will be optimized to represent a specific grid, whereas the edges will represent the relations of them. On top of the potential impact on the climate studies field that I mentioned above, the advantages of working with climate data are as follows.
There are already well known global phenomena connecting grids located far away; hence, guaranteeing the existence of connections in the graph. Furthermore, the same well known phenomena also allows us to validate our representation graphs. Lastly, following the common practices of climate studies, we can start with a simplified scenario, by creating node pseudo-features using simple statistics.

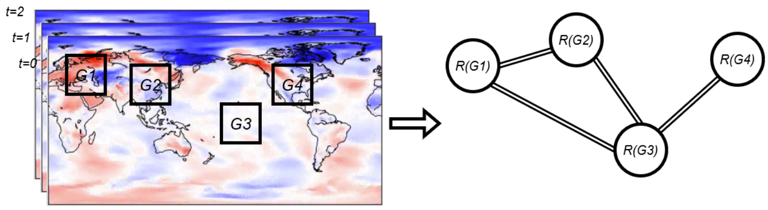
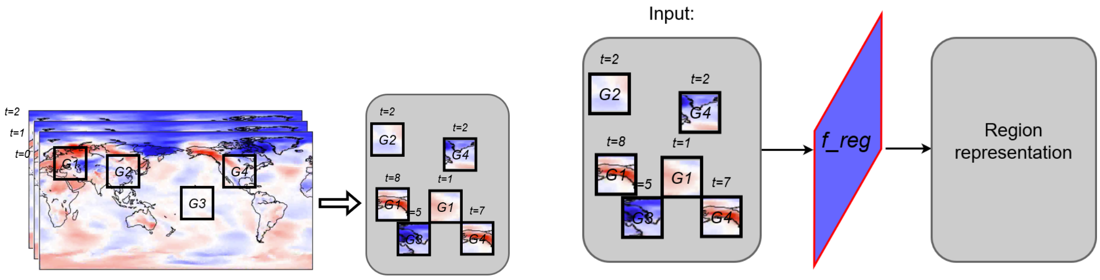


Figure 1: General objective

Note that compared to analyzing the whole map only through clustering, we will be able to guarantee a statistical importance for the grid and encode region to region influence.

3.1 Capturing assumptions: Probabilistic modeling



(a) Dataset

(b) f_{reg} outputs representation of regions

Figure 2

Steps of this approach are as described below:

1. Each datapoint (See *Fig 2*) will be hard clustered using an arbitrary function f_{reg} that's designed to represent a region. As mentioned above, choosing simple statistics (mean, median, etc.) is also an option for node attributes at this stage for initial testing. However, keep in mind that choosing a information maximization based approach seems the most sensible.

2. By sampling from all available climate maps (the whole image), we will be able to create a Bayesian Network or Markov Random Field, whichever is more suitable. Intuitively, we will essentially eliminate connections between independent regions.

While being relatively simple, cheap, and explainable, this approach cannot make connections through time. In other words, if a temperature increment in one region ends up causing a strong wind in different location 3 months later, we will not be able to encapsulate that phenomena, no matter how consistent that happens. **Therefore, this approach is more suitable when the time is not important (e.g. medical images).**

3.2 Capturing assumptions: Information maximization

To address the shortcoming above, the following learning architecture is proposed. In this case, the function f_{reg} will output a representation vector, as opposed to hard clustering. After stacking the outputted vectors along time, we can use a Recurrent Neural Network f_{rel} that will aim to maximize the mutual information between two randomly chosen grids. From there, by setting a threshold, **we will be able to connect the grids that seem to influence each other.**

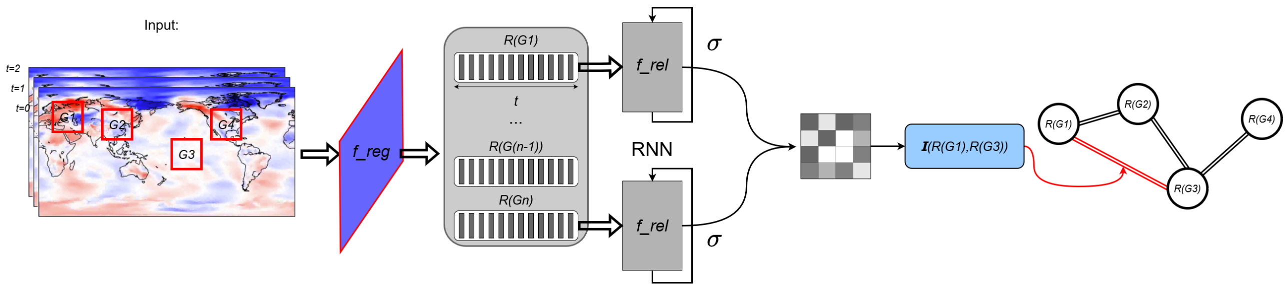


Figure 3: Information maximization approach

It is under consideration whether to train f_{reg} and f_{rel} simultaneously or separately for this approach. In fact, it might even be sensible to allow the end-user to specify weights for regional and relational importance.

3.3 Connection to learning theory

Take a graph representing grids/tiles in the nodes and relational features between the grids in the edges. While such graph is a useful tool out of the box (e.g. identifying global scale phenomena), it's also useful when designing machine learning models. Aside from more clear benefits in node prediction tasks of the application domain, one could also use it to reduce the redundancies in the shallow layers of NNs. For example, one can propose following simple hypotheses (and many more) based on the graph by partially connecting layers based on the receptive fields of the outputs.
In a sense, this can be seen as an extension to the assumption of locality which reduces redundancy by combining the grids that are highly related to one another. Since we are removing connections from fully connected layers, the resulting complexity of the hypotheses

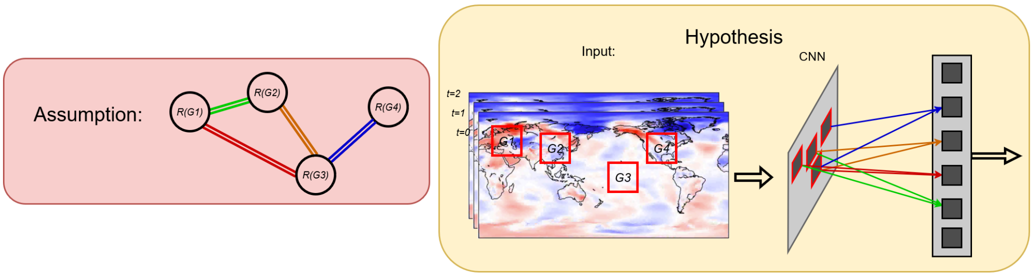


Figure 4: Example hypothesis (model) based on the assumption graph. Note that the parameters of the model can vary.

should be reduced. Furthermore, since lattice structure (for CNN) can also be represented by using graphs, it should be possible to quantize the complexity of our proposed hypotheses, which is very useful when selecting algorithms. With my limited knowledge, I believe this direction should be explored in order to make AI machines domain-aware.

4 Initial Research Project

In short, I am beyond thrilled about investigating **the effects** of such domain-specific assumptions. However, I also realize that assumption representation stage requires considerable amount of time. Therefore, while it can be sped up based on the supervisor's directions, my current plan is as follows:

1. **Choosing a dataset.** For this research, I hope to use real-world measurement data. However, there are concerns with availability of enough data points, spanning over sufficiently long period of time and including sufficient variables. Therefore, using simulation data is another possibility. Considering my familiarity with the CESM model by Danabasoglu et al. [2020], I plan to finish obtaining the correct data within 1 month of joining the program.
2. **Capturing assumptions: Probabilistic modelling.** Due to the simplicity, I believe this approach can be tested relatively quickly and cheap to reproduce. However, I hope to receive direction from my future supervisor before deciding whether or not to skip this step. Including the time needed for learning, testing, and validating, this step is estimated to take roughly 4 months.
3. **Capturing assumptions: Information assumption.** While the effectiveness of this approach seems promising, it could require more time and resources. Depending on the university's facilities, current plan is to finalize in no more than 5-7 months.
4. **Application to other domains (Optional).** Depending on the situation, I would like to demonstrate applications in other domains as well. I believe this stage would be immensely helpful for internship opportunities. Since medical images, head shots, and finger prints have similar 'consistency' as climate image, I am currently considering those domains. However, while these application scenarios could offer surprising new insights, there is no apparent strategy for validation, as opposed to climate map data.
5. **Generalization and statistical learning theory.** Since there can be numerous proposed hypotheses based on the obtained graph, concepts of statistical learning theory, such as complexity term, are very important for providing AI with domain-awareness.

Thus, I hope to invest enough time in training before making a concrete plan for this stage. While I already started learning using various sources and papers, I intend to deepen my understanding through graduate level courses and interaction with the leading researchers of the field. Additionally, directions and insights from my future supervisor would be truly invaluable for verifying my idea and making a concrete research plan.

References

Peter B. Gibson, Sarah E. Perkins-Kirkpatrick, Petteri Uotila, Acacia S. Pepler, and Lisa V. Alexander. On the use of self-organizing maps for studying climate extremes. *Journal of Geophysical Research: Atmospheres*, 122(7):3891–3903, 2017. doi: https://doi.org/10.1002/2016JD026256. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016JD026256`. pages 3

Xu Ji, João F. Henriques, and Andrea Vedaldi. Invariant information clustering for unsupervised image classification and segmentation, 2018. URL `https://arxiv.org/abs/1807.06653`. pages 3

Mehdi Noroozi and Paolo Favaro. Unsupervised learning of visual representations by solving jigsaw puzzles, 2016. URL `https://arxiv.org/abs/1603.09246`. pages 3

Kai Han, Yunhe Wang, Jianyuan Guo, Yehui Tang, and Enhua Wu. Vision gnn: An image is worth graph of nodes, 2022. URL `https://arxiv.org/abs/2206.00272`. pages 3

G. Danabasoglu, J.-F. Lamarque, J. Bacmeister, D. A. Bailey, A. K. DuVivier, J. Edwards, L. K. Emmons, J. Fasullo, R. Garcia, A. Gettelman, C. Hannay, M. M. Holland, W. G. Large, P. H. Lauritzen, D. M. Lawrence, J. T. M. Lenaerts, K. Lindsay, W. H. Lipscomb, M. J. Mills, R. Neale, K. W. Oleson, B. Otto-Bliesner, A. S. Phillips, W. Sacks, S. Tilmes, L. van Kampenhout, M. Vertenstein, A. Bertini, J. Dennis, C. Deser, C. Fischer, B. Fox-Kemper, J. E. Kay, D. Kinnison, P. J. Kushner, V. E. Larson, M. C. Long, S. Mickelson, J. K. Moore, E. Nienhouse, L. Polvani, P. J. Rasch, and W. G. Strand. The community earth system model version 2 (cesm2). *Journal of Advances in Modeling Earth Systems*, 12 (2):e2019MS001916, 2020. doi: https://doi.org/10.1029/2019MS001916. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001916`. e2019MS001916 2019MS001916. pages 6