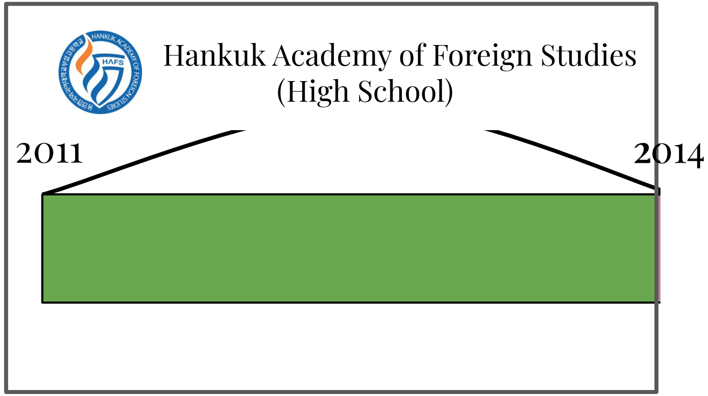
Intro

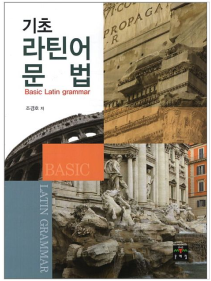
Hazel Kim

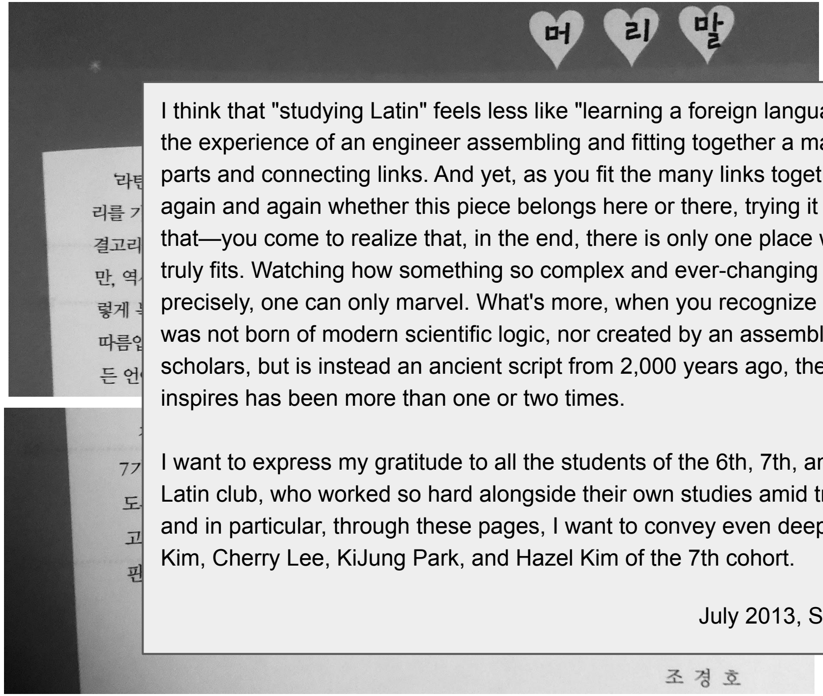
Nov. 30. 2020

Timeline



- **HAFS Latin Honor Society**
    - **Dr. Sancho Kyung-ho Cho**
    - **Korea National High School Textbook of Latin Language and Linguistic Analysis**





I think that "studying Latin" feels less like "learning a foreign language" and more like the experience of an engineer assembling and fitting together a machine made of many parts and connecting links. And yet, as you fit the many links together—pondering again and again whether this piece belongs here or there, trying it this way and that—you come to realize that, in the end, there is only one place where each piece truly fits. Watching how something so complex and ever-changing locks together so precisely, one can only marvel. What's more, when you recognize that this language was not born of modern scientific logic, nor created by an assembly of modern scholars, but is instead an ancient script from 2,000 years ago, the sense of awe it inspires has been more than one or two times.

I want to express my gratitude to all the students of the 6th, 7th, and 8th cohorts of the Latin club, who worked so hard alongside their own studies amid truly busy schedules, and in particular, through these pages, I want to convey even deeper thanks to Andy Kim, Cherry Lee, KiJung Park, and Hazel Kim of the 7th cohort.

July 2013, Sancho Kyung-ho Cho

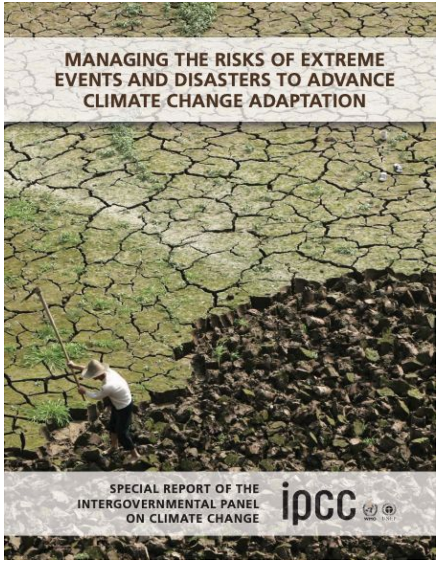
Timeline

Hankuk Academy of Foreign Studies
(High School)

2011 2014

- HAFS Latin Honor Society
    - Dr. Sancho Kyung-ho Cho
    - Korea National High School Textbook of Latin Language and Linguistic Analysis

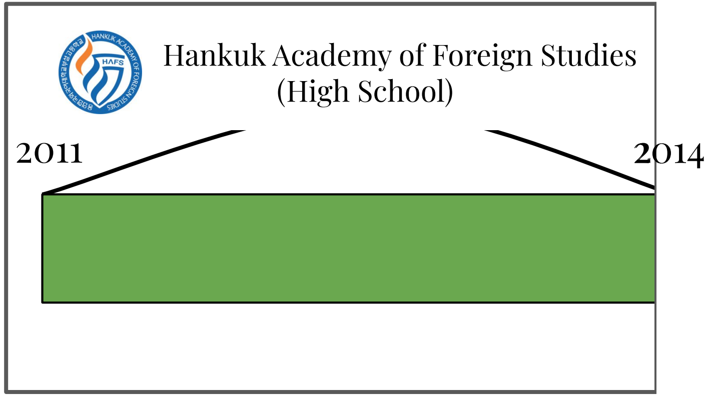
- **IPCC Project**
    - **Seoul National University Graduate School of Public Health**
    - **Korea Meteorological Administration**
    - **Dr. Jae Choe & Dr. Kyung-jin Boo**



… I watched Hazel Hyunji Kim, giving an ambitious presentation about the IPCC project. While all the other students presented on club activities or individual volunteer work they would carry out at their own schools, she proposed an activity to be done ***together*** with the students who had participated in this program. First of all, I thought that collaborating with students from other schools who had far more experience and passion for the environment than I did would in itself be tremendously meaningful. Furthermore, although I thought it would not be easy for students to directly translate a specialized report like SREX, I judged that if many students participated and cooperated, we could certainly manage it. And so I decided to take part in this project.

Hye-jin Jang, Korea Minjok Leadership Academy

Timeline



- HAFS Latin Honor Society
    - Dr. Sancho Kyung-ho Cho
    - Korea National High School Textbook of Latin Language and Linguistic Analysis

- **IPCC Project**
    - **Seoul National University Graduate School of Public Health**
    - **Korea Meteorological Administration**
    - **Dr. Jae Choe & Dr. Kyung-jin Boo**

… with my own conviction and willpower alone, I surely could not have come this far during this busy period as a senior in high school. And the driving force that allowed me to stand before all of you was entirely thanks to the encouragement, humility, and sense of responsibility of my colleagues. Teammates who worked desperately to meet our set deadlines; teammates who never forgot to say "sorry" even when they were late; the way each person stepped up to organize terminology or shared the difficulties they'd faced in order to improve the quality of the translation. These were enough to spur me on whenever I grew weary. I am learning a special sense of ***camaraderie***, ***responsibility***, and the aesthetics of ***cooperation*** that I could never have learned in school.

Young-im Yu, Yeongbok Girls' High School

There were more than a few moments during the IPCC project when I wanted to give up. I had started it with the dream of contributing to the fight against climate change from a young person's perspective, but when things got difficult, there were many times I lost sight of that dream entirely. Yet looking back now, as I finish the publication and write this reflection, I realize it's precisely because I pushed through each of those moments that I'm able to complete the project so rewardingly. As I pursue other dreams throughout my life, there will surely be difficult stretches along the way. But I have to remind myself that only by ***refusing to give up*** and pressing on can I see them through, just as I did with the IPCC project. I also want to tell Hazel and Da-eun, who comforted me and cheered me on so that I wouldn't give up in my hardest moments, how deeply grateful I am. Like them, I hope to become someone who stands beside the people around me as they chase their own dreams, cheering them on so they never give up.

Sojeong Lee, Hankuk Academy of Foreign Studies

Timeline

Hankuk Academy of Foreign Studies
(High School)

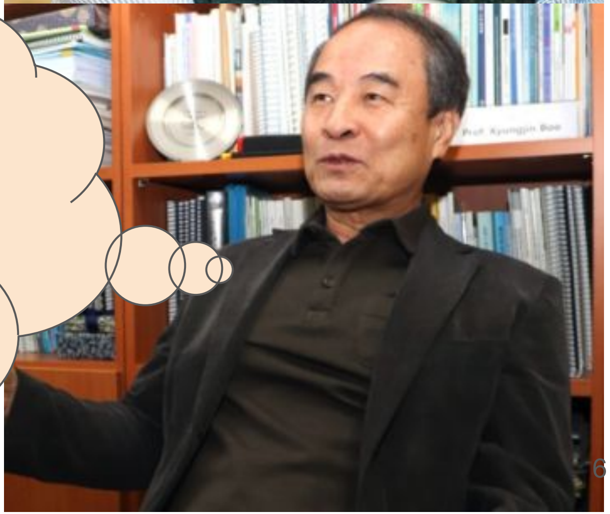
2011

Let's sit down together now and then and talk about this and that. It would be best of all if you could stay on and study here with us, but even if you get into a SKY university, I wonder if you might still come by a couple of times a week to do research together. What do you say?

- HAFS Latin Honors Society
    - Dr. Sancho Kyung-ho Cho
    - Korea National High School Textbook of Latin Language a
      Linguistic Analysis

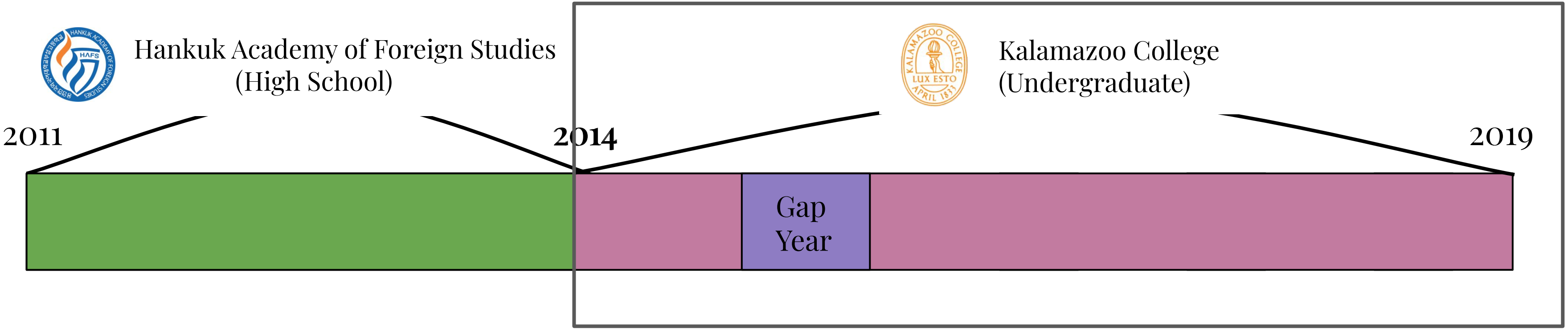
I love your sincere and earnest attitude toward the path of learning. Hazel, you will do wonderfully.

My hope is that you'll always hold on to the heart you started with, and grow into a scholar who gives herself to her work with honesty and sincerity...



- **IPCC Project**
    - **Seoul National University Graduate School of Publi**
    - **Korea Meteorological Administration**
    - **Dr. Jae Choe & Dr. Kyung-jin Boo**

Timeline



Consilience:
The Unity of Knowledge

Master Key

Dr. John Dugas
Political Science

Dr. Alyce Brady
Computer Science

Dr. Andrew Koehler
Violin Performance

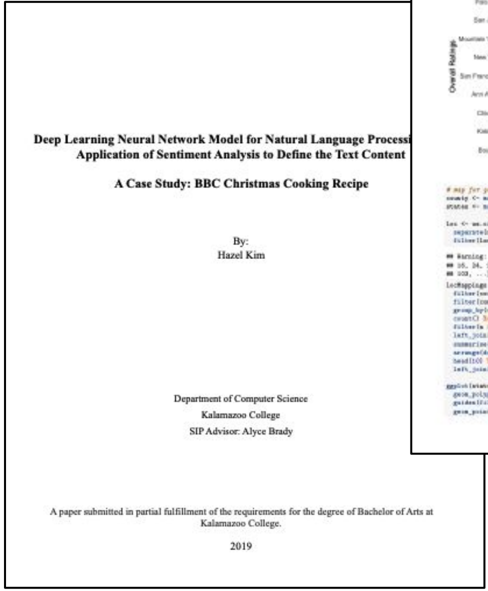
Timeline

Hankuk Academy of Foreign Studies
(High School)

Kalamazoo College
(Undergraduate)

2011 2014 2019

Gap Year



Deep Learning Neural Network Model for Natural Language Processi[ng]
Application of Sentiment Analysis to Define the Text Content

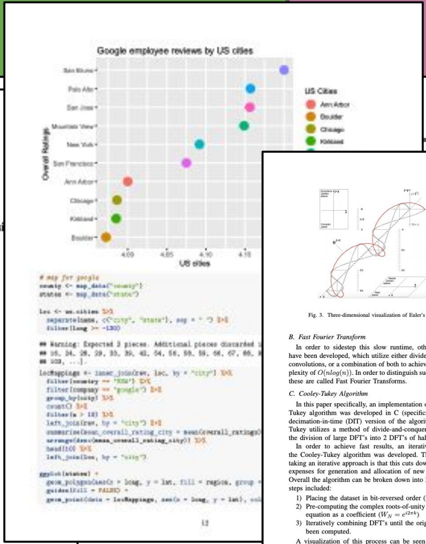
A Case Study: BBC Christmas Cooking Recipe

By:
Hazel Kim

Department of Computer Science
Kalamazoo College
SIP Advisor: Alyce Brady

A paper submitted in partial fulfillment of the requirements for the degree of Bachelor of Arts at Kalamazoo College.

2019



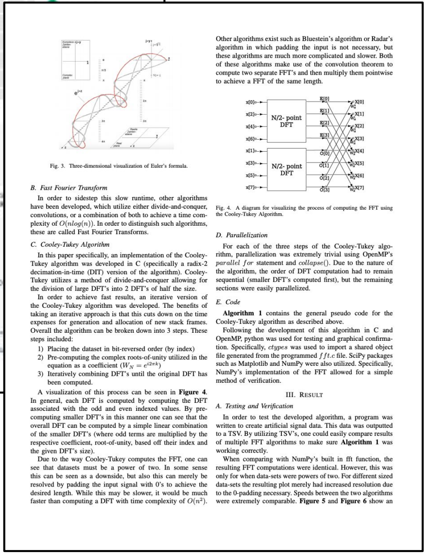


Fig. 3. Three-dimensional visualization of Euler's formula.

B. Fast Fourier Transform

In order to sidestep this slow runtime, other algorithms have been developed, which utilize either divide-and-conquer, convolutions, or a combination of both to achieve a time complexity of $O(nlog(n))$. In order to distinguish such algorithms, these are called Fast Fourier Transforms.

C. Cooley-Tukey Algorithm

In this paper specifically, an implementation of the Cooley-Tukey algorithm was developed in C (specifically a radix-2 decimation-in-time (DIT) version of the algorithm). Cooley-Tukey utilizes a method of divide-and-conquer allowing for the division of large DFT's into 2 DFT's of half the size.

In order to achieve fast results, an iterative version of the Cooley-Tukey algorithm was developed. The benefits of taking an iterative approach is that this cuts down on the time expenses for generation and allocation of new stack frames. Overall the algorithm can be broken down into 3 steps. These steps included:

1) Placing the dataset in bit-reversed order (by index)
2) Pre-computing the complex roots-of-unity utilized in the equation as a coefficient ($W_N = e^{i2\pi k}$)
3) Iteratively combining DFT's until the original DFT has been computed.

A visualization of this process can be seen in **Figure 4**. In general, each DFT is computed by computing the DFT associated with the odd and even indexed values. By pre-computing smaller DFT's in this manner one can see that the overall DFT can be computed by a simple linear combination of the smaller DFT's (where odd terms are multiplied by the respective coefficient, root-of-unity, based off their index and the given DFT's size).

Due to the way Cooley-Tukey computes the FFT, one can see that datasets must be a power of two. In some sense this can be seen as a downside, but also this can merely be resolved by padding the input signal with 0's to achieve the desired length. While this may be slower, it would be much faster than computing a DFT with time complexity of $O(n^2)$.

Other algorithms exist such as Bluestein's algorithm or Radar's algorithm in which padding the input is not necessary, but these algorithms are much more complicated and slower. Both of these algorithms make use of the convolution theorem to compute two separate FFT's and then multiply them pointwise to achieve a FFT of the same length.

Fig. 4. A diagram for visualizing the process of computing the FFT using the Cooley-Tukey Algorithm.

D. Parallelization

For each of the three steps of the Cooley-Tukey algorithm, parallelization was extremely trivial using OpenMP's $parallel\ for$ statement and $collapse()$. Due to the nature of the algorithm, the order of DFT computation had to remain sequential (smaller DFT's computed first), but the remaining sections were easily parallelized.

E. Code

Algorithm 1 contains the general pseudo code for the Cooley-Tukey algorithm as described above.

Following the development of this algorithm in C and OpenMP, python was used for testing and graphical confirmation. Specifically, $ctypes$ was used to import a shared object file generated from the programmed $fft.c$ file. SciPy packages such as Matplotlib and NumPy were also utilized. Specifically, NumPy's implementation of the FFT allowed for a simple method of verification.

III. RESULT

A. Testing and Verification

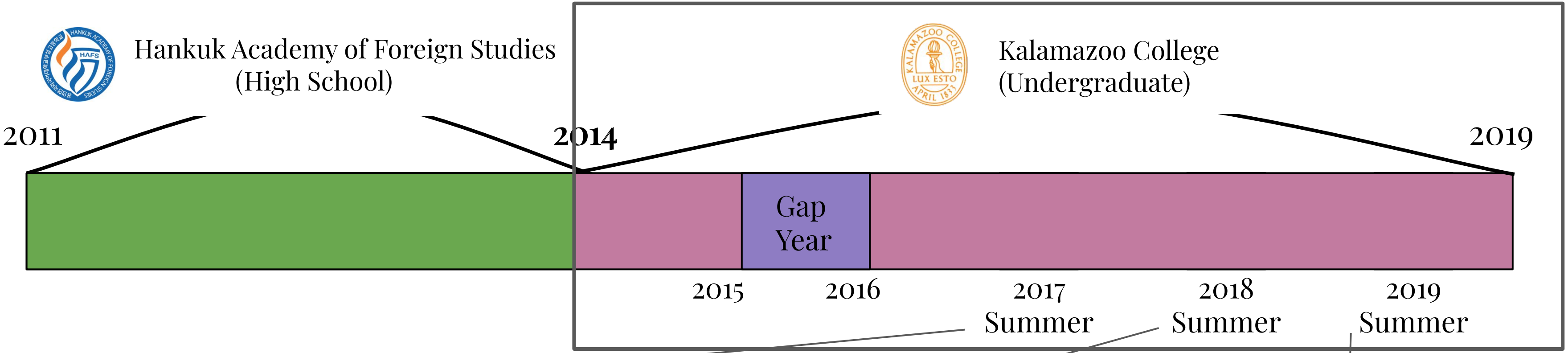
In order to test the developed algorithm, a program was written to create artificial signal data. This data was outputted to a TSV. By utilizing TSV's, one could easily compare results of multiple FFT algorithms to make sure **Algorithm 1** was working correctly.

When comparing with NumPy's built in fft function, the resulting FFT computations were identical. However, this was only for when data-sets were powers of two. For different sized data-sets the resulting plot merely had increased resolution due to the 0-padding necessary. Speeds between the two algorithms were extremely comparable. **Figure 5** and **Figure 6** show an

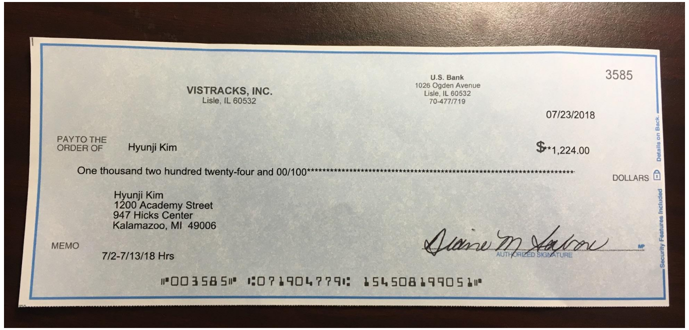
Dr. Alyce Brady
Computer Science

Dr. Eric Nordmoe
Mathematics

Timeline



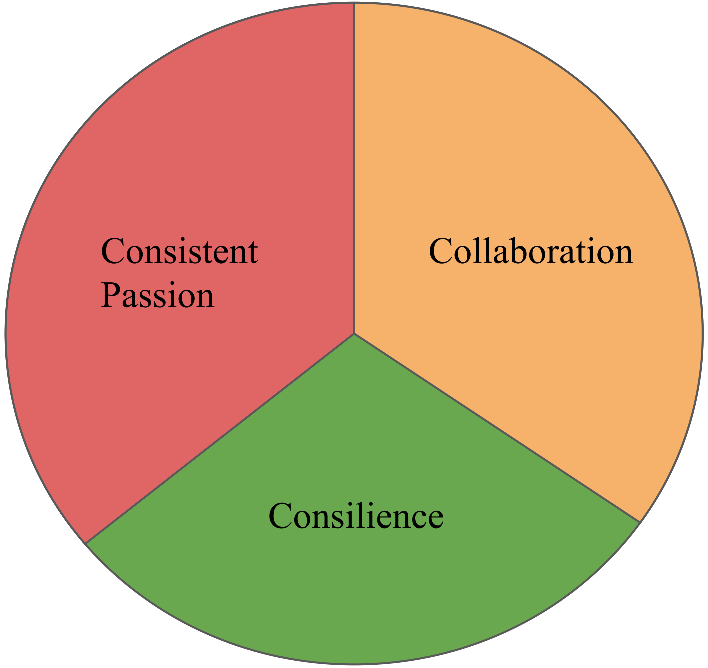
Data Analyst Intern
Seoul Metropolitan Government



Data Engineering Intern
Vistracks Inc. Chicago, Illinois
iOS Mobile Application Service Team

Research Engineering Intern
Seattle, Washington

Hazel’s Values



Research Interest

Natural Language Processing
Machine Reasoning

Future Goals

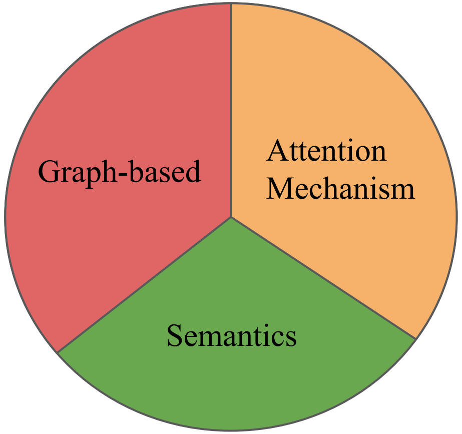
- Level of Artificial Intelligence that I want to reach:
  - Among the social sciences, **law** may come the closest to **a system of formal logic**. To oversimplify, legal rulings involve setting forth axioms derived from precedent, applying those axioms to the particular facts at hand, and reaching a conclusion accordingly. This logic-oriented methodology is exactly the type of activity to which machine intelligence can fruitfully be applied. - AI Will Transform The Field Of Law - Fobes 12/19/2019
    - Contract Review
    - Contract Analytics
    - Litigation Prediction
    - Legal Research
  - “AI technology will ultimately broaden the lawyer’s role from a narrow focus on risk mitigation to more strategic engagement on company initiatives.” - Lawgeex CEO Noory Bechor.

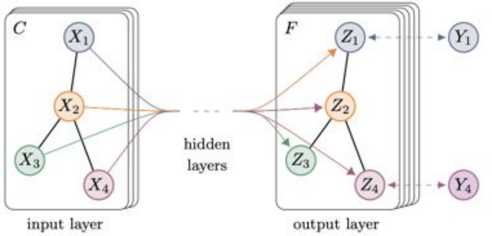
Future Goals

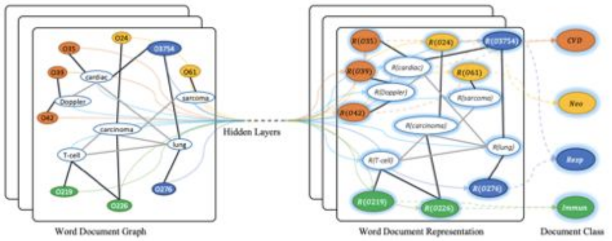
- Areas that I want to get involved:
  - Latent Alignment for Semantic Analysis
  - Interpretable and Generalizable Deep Learning
  - Logic
  - Linguistic Perspective + Cognitive Science
    - Language serves essentially for the expression of thought. - Noam Chomsky
    - Language is the jewel in the crown of cognition. - Steven Pinker

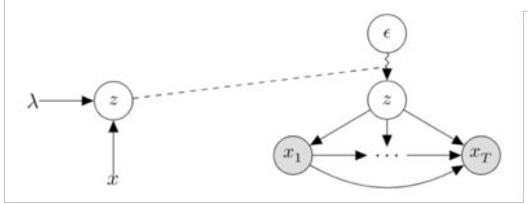
Previous Research

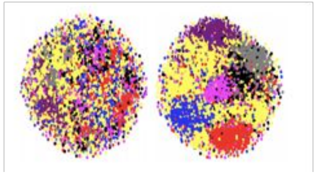
- Graph-based (Heterogenous)
- Attention Mechanism
- Semantics











References from https://arxiv.org/pdf/1812.06834.pdf, and https://arxiv.org/pdf/1809.05679.pdf

- **Preserving semantic knowledge** for better text analysis: Use of Graph's Topological Structures.
- **Latent Semantic Analysis**: An NLP Technique for conceptual analysis which produces concepts to analyze relationship between a set of documents and the terms they contain.

Previous Research

1. Gated Attention Graph Network
2. Graph Bert for Text Classification
3. Distributional Semantics

Previous Research 1: Gated Attention Graph Network

Research Objectives

- Convolutional Operator + Attention Mechanism + Different Importance
    - Graph Aggregator

Previous Research 1: Gated Attention Graph Network

Data Exploration

Node Features

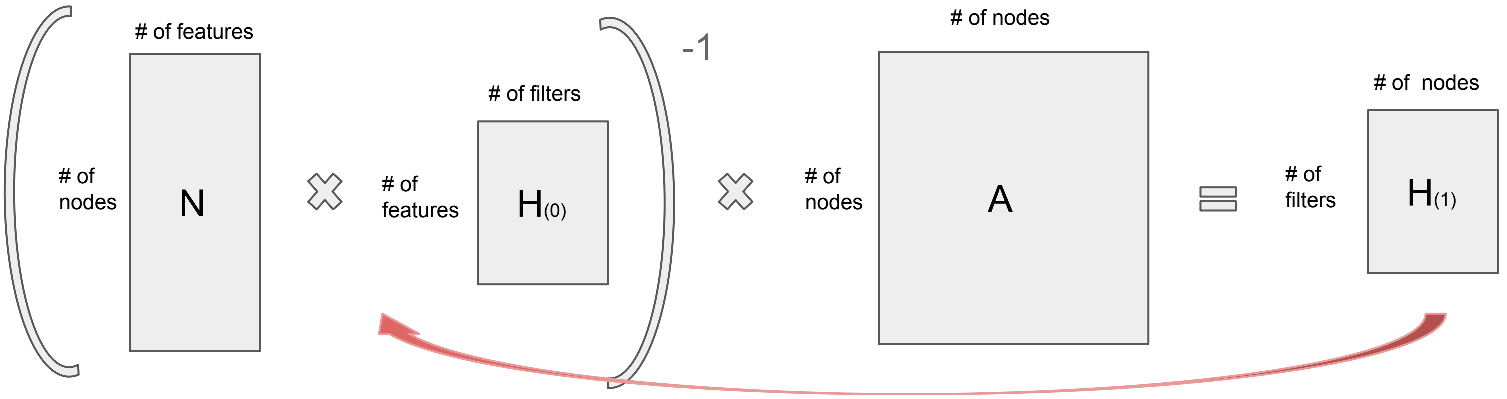
- Row: node size = 12772
- Column: feature categories = 1256

Edges (size = 12772)

- Movie-Director: 4661 stored elements
- Director-Movie: 4661 stored elements
- Actor-Movie: 13983 stored elements
- Movie-Actor: 13983 stored elements

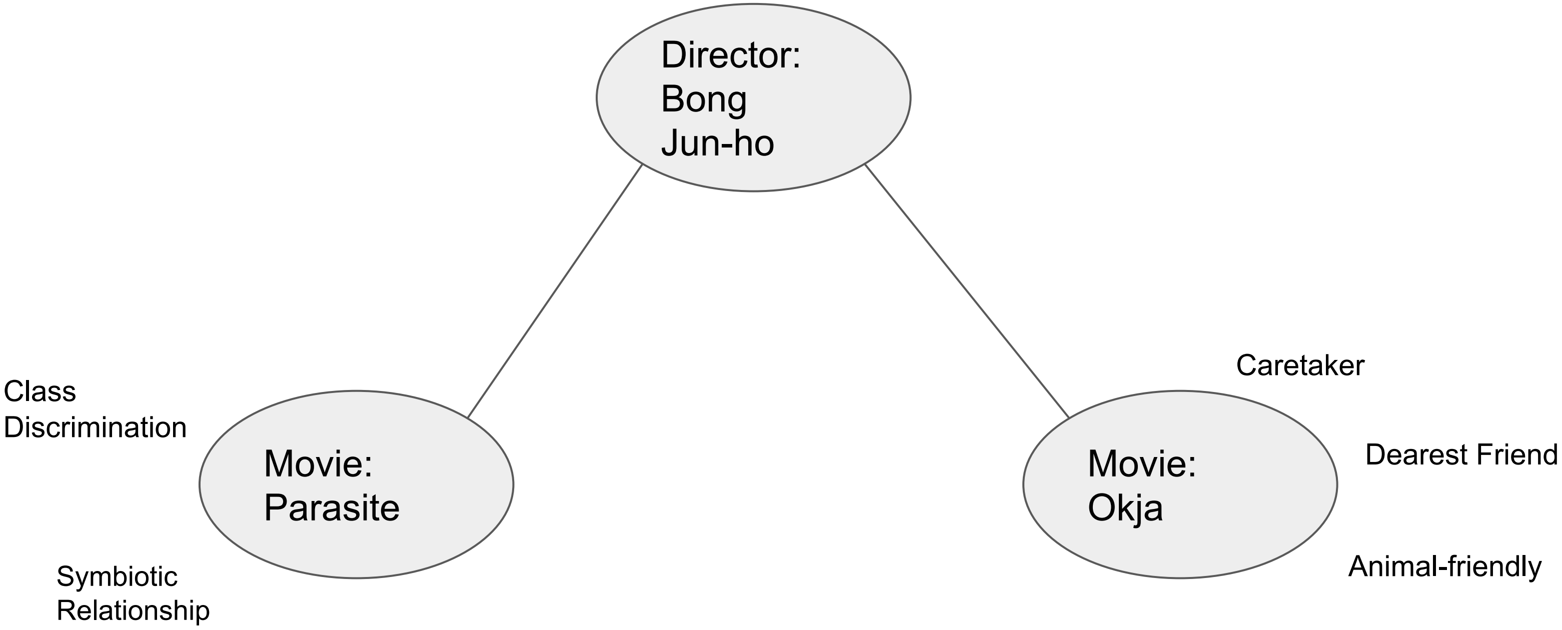
Labels (0: drama, 1: comedy, 2: action)

- Train: 300
- Validation: 300
- Test: 2339



Previous Research 1: Gated Attention Graph Network

Node Features



Previous Research 1: **Gated Attention Graph Network**
Evolution of GNN

- Graph Convolutional Network (GCN), 2017
  - Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *International Conference on Learning Representations (ICLR)*, 2017.
- GraphSAGE, 2017
  - W. Hamilton, Z. Ying, and J. Leskovec. Inductive representation learning on large graphs. In *NIPS*, pages 1025–1035, 2017a.
- Graph Attention Network (GAT), 2018
  - P. Velicˇkovic´, G. Cucurull, A. Casanova, A. Romero, P. Lio`, and Y. Bengio. Graph attention networks. In *ICLR*, 2018.
- Gated Attention Network (GaAN), 2018
  - Zhang, Jiani, et al. "Gaan: Gated attention networks for learning on large and spatiotemporal graphs." *arXiv preprint arXiv:1803.07294* (2018).

Previous Research 1: Gated Attention Graph Network

Graph Aggregator

Develops a localized parameter-sharing operator on a set of neighboring nodes to aggregate a local set of lower-level features. (Using a receptive field of the aggregator)

- Its design determines a model's ability to capture the structural information of graphs.
- Functions that are permutation invariant and can be dynamically resizing
- i.e. either pooling over neighborhoods or computing a weighted sum of the neighboring features.

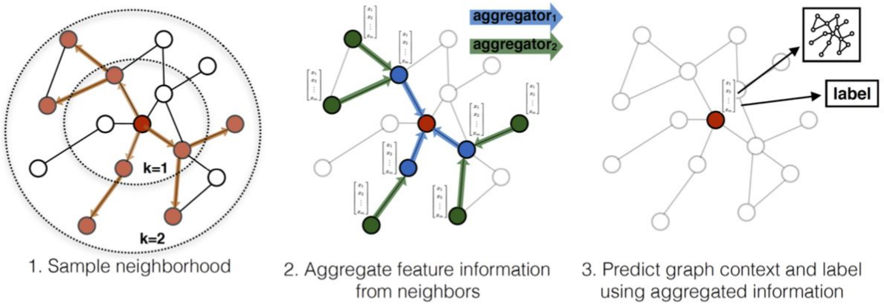


Figure 1: Visual illustration of the GraphSAGE sample and aggregate approach.

W. Hamilton, Z. Ying, and J. Leskovec. Inductive representation learning on large graphs. In *NIPS*, pages 1025–1035, 2017a.

Previous Research 1: Gated Attention Graph Network

How to Learn Representations of Nodes in a Graph

Algorithm 1: WL-1 algorithm (Weisfeiler & Lehmann, 1968)

Input: Initial node coloring $(h_1^{(0)}, h_2^{(0)}, ..., h_N^{(0)})$
Output: Final node coloring $(h_1^{(T)}, h_2^{(T)}, ..., h_N^{(T)})$
t ← 0;
repeat
  for $v_i \in \mathcal{V}$ **do**
    $h_i^{(t+1)} \leftarrow \text{hash}\left(\sum_{j \in \mathcal{N}_i} h_j^{(t)}\right)$;
  $t \leftarrow t+1$;
until *stable node coloring is reached*;

$$h_i^{(l+1)} = \sigma\left(\sum_{j \in \mathcal{N}_i} \frac{1}{c_{ij}} h_j^{(l)} W^{(l)}\right)$$

- c_{ij} is an appropriately chosen normalization constant for the edge (v_i, v_j)
- $h_{(l)}$ now to be a vector of *activations* of node i in the l_{th} neural network layer
- $W_{(l)}$ is a (i) layer-specific weight matrix

<u>Replace</u> the hash function with a neural network layer-like differentiable function with trainable parameters.

Brendan L. Douglas. The Weisfeiler-Lehman method and graph isomorphism testing. *arXiv preprint arXiv:1101.5211*, 2011.
Boris Weisfeiler and A. A. Lehmann. A reduction of a graph to a canonical form and an algebra arising during this reduction. *Nauchno-Technicheskaya Informatsia*, 2(9):12–16, 1968.

Previous Research 1: Gated Attention Graph Network

Different Graph Aggregators

GCN

Scales linearly in the number of graph edges.

Learns hidden layer representations that encode both local graph structure and features of nodes.

GraphSAGE

(SAmple + AggreGatE)

Compute only the representations that are necessary to satisfy the recursion at each depth.

Operates by sampling a fixed-size neighborhood of each node, and then perform a specific aggregator over it.

- Mean Aggregator
- Pooling Aggregator

GAT

Uses a sampling algorithm to select a small subset of the nodes and edges.

To obtain sufficiently expressive power to transform the input features into higher level, the model adds learnable linear transformation (attention mechanism).

Does not compute additional value vectors and uses a fully-connected layer to compute φ_w.

GaAN

Adopts the key-value attention mechanism and the dot product attention.

Uses convolutional network that takes the center node and neighboring node features to generate the gate values.

Adding gates would not produce too many additional parameter. (Computationally Inexpensive)

Previous Research 1: Gated Attention Graph Network

GaAN

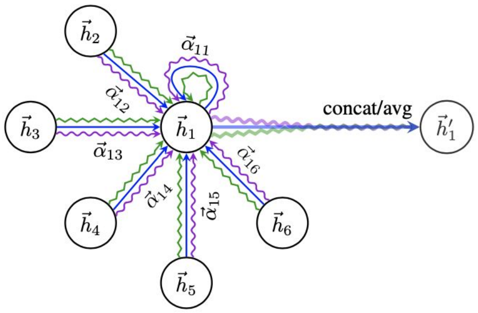
- G vector: **combination of average pooling and max pooling.**

$$\mathbf{y}_i = \mathrm{FC}_{\theta_o}(\mathbf{x}_i \oplus \Big\Vert_{k=1}^{K} (g_i^{(k)} \sum_{j\in\mathcal{N}_i} w_{i,j}^{(k)} \mathrm{FC}^h_{\theta_v^{(k)}}(\mathbf{z}_j))),$$

$$\mathbf{g}_i = [g_i^{(1)}, ..., g_i^{(K)}] = \psi_g(\mathbf{x}_i, \mathbf{z}_{\mathcal{N}_i}),$$

$$\boxed{\mathbf{g}_i = \mathrm{FC}^{\sigma}_{\theta_g}(\mathbf{x}_i \oplus \max_{j\in\mathcal{N}_i}(\{\mathrm{FC}_{\theta_m}(\mathbf{z}_j)\}) \oplus \frac{\sum_{j\in\mathcal{N}_i}\mathbf{z}_j}{|\mathcal{N}_i|})}$$

  - θ_m maps the neighbor features to a dm dimensional vector before taking the element-wise max.
  - θ_g maps the concatenated features to the final K gates.
  - By setting a small dm, the subnetwork for computing the gate will have negligible computational overhead.



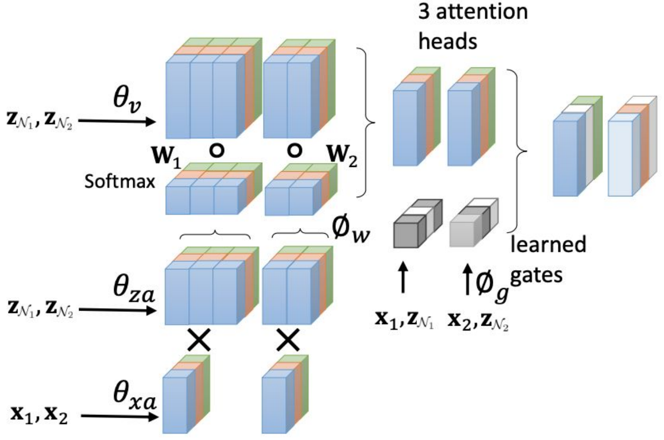


Figure 1: Illustration of a three-head gated attention aggregator with two center nodes in a mini-batch. $|\mathcal{N}_1| = 3$ and $|\mathcal{N}_2| = 2$ respectively. Different colors indicate different attention heads. Gates in darker color stands for larger values. (Best viewed in color)

Previous Research 1: Gated Attention Graph Network

Take Advantage of GaAN

- Sampling a small number of neighborhood size instead of considering all.
- Various attention alignment available.

Table 4: Comparison of the test F1 score on the Reddit and PPI datasets with different sampling neighborhood sizes and attention head number K. S_1 and S_2 are the maximum number of sampled neighborhoods in the 1st and 2nd sampling steps. 'all' means to sample all the neighborhoods.

Models	Reddit					PPI	
	#Param	S_1, S_2 25,10	S_1, S_2 50,20	S_1, S_2 100,40	S_1, S_2 200,80	#Param	S_1, S_2 all, all
2-Layer FNN	1.71M	73.58±0.09	73.58±0.09	73.58±0.09	73.58±0.09	1.23M	54.07±0.06
Avg. pooling	866K	95.78±0.07	96.11±0.07	96.28±0.05	96.35±0.02	274K	96.85±0.19
Max pooling	866K	95.62±0.03	96.06±0.09	96.18±0.11	96.33±0.04	274K	98.39±0.05
Pairwise+sigmoid	965K	95.86±0.08	96.19±0.04	96.33±0.05	96.38±0.08	349K	98.39±0.05
Pairwise+tanh	965K	95.80±0.03	96.11±0.05	96.26±0.03	96.36±0.04	349K	98.32±0.18
Attention-only-K1	562K	96.15±0.06	96.40±0.05	96.48±0.02	96.54±0.07	168K	96.31±0.08
Attention-only-K2	571K	96.19±0.07	96.40±0.04	96.52±0.02	96.57±0.02	178K	97.36±0.08
Attention-only-K4	587K	96.11±0.06	96.40±0.02	96.49±0.03	96.56±0.02	196K	98.09±0.07
Attention-only-K8	620K	96.10±0.03	96.38±0.01	96.50±0.04	96.53±0.02	233K	98.46±0.09
GaAN-K1	620K	96.29±0.05	96.50±0.08	96.67±0.04	96.73±0.05	201K	96.95±0.09
GaAN-K2	629K	96.33±0.02	96.59±0.02	96.71±0.05	96.82±0.05	211K	97.92±0.05
GaAN-K4	645K	**96.36±0.03**	**96.60±0.03**	96.73±0.04	**96.83±0.03**	230K	98.42±0.02
GaAN-K8	678K	96.31±0.13	**96.60±0.02**	**96.75±0.03**	96.79±0.08	267K	**98.71±0.02**

Zhang, Jiani, et al. "Gaan: Gated attention networks for learning on large and spatiotemporal graphs." *arXiv preprint arXiv:1803.07294* (2018).

Previous Research 1: Gated Attention Graph Network

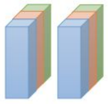
Implementation: Gated Attention Aggregator

- Class Attention(nn.Module)

$$\alpha_{ij} = \frac{\exp\left(\mathrm{LeakyReLU}\left(\vec{\mathbf{a}}^T[\mathbf{W}\vec{h}_i \Vert \mathbf{W}\vec{h}_j]\right)\right)}{\sum_{k\in\mathcal{N}_i}\exp\left(\mathrm{LeakyReLU}\left(\vec{\mathbf{a}}^T[\mathbf{W}\vec{h}_i \Vert \mathbf{W}\vec{h}_k]\right)\right)}$$

- Class GCN Layer: forward()
- Class GCN Block: layers.append()
- Class GCN Net: forward()

3 attention heads

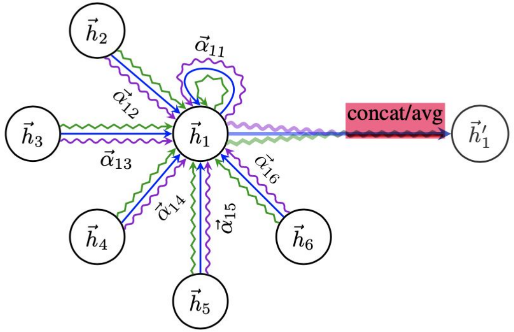


$$\mathbf{y}_i = \mathrm{FC}_{\theta_o}(\mathbf{x}_i \oplus \boxed{\Big\Vert_{k=1}^{K} (g_i^{(k)}} \sum_{j\in\mathcal{N}_i} w_{i,j}^{(k)} \mathrm{FC}^h_{\theta_v^{(k)}}(\mathbf{z}_j))),$$

$$\mathbf{g}_i = [g_i^{(1)}, ..., g_i^{(K)}] = \psi_g(\mathbf{x}_i, \mathbf{z}_{\mathcal{N}_i}),$$

g vector operation

```
out = torch.sum(out, dim=0)
out = torch.div(out, num_head)
out = self.elu(out)
```



arXiv:1710.10903

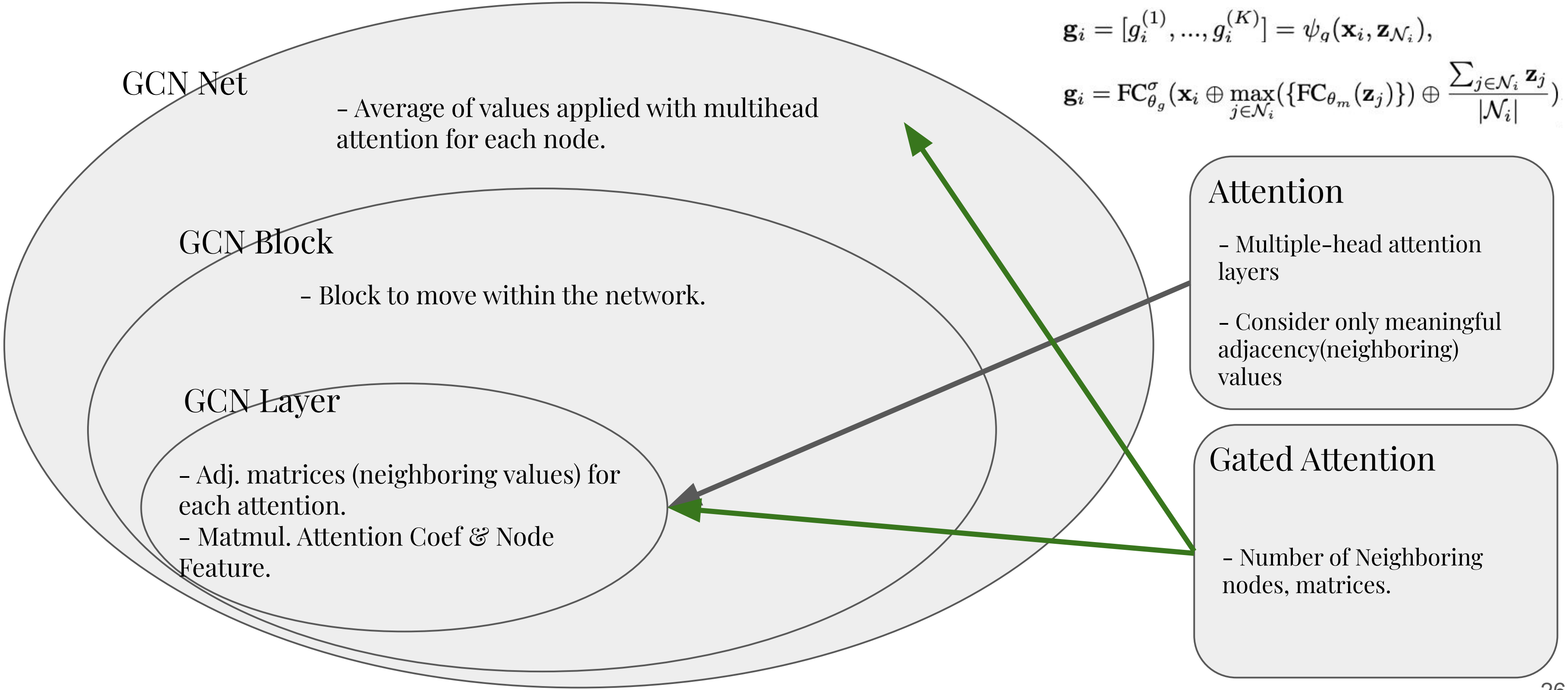
Previous Research 1: Gated Attention Graph Network

Network Implementation Structure

$$\mathbf{y}_i = \mathrm{FC}_{\theta_o}\Big(\mathbf{x}_i \oplus \big\Vert_{k=1}^{K} \big(g_i^{(k)} \sum_{j\in\mathcal{N}_i} w_{i,j}^{(k)} \mathrm{FC}^h_{\theta_v^{(k)}}(\mathbf{z}_j)\big)\Big),$$

$$\mathbf{g}_i = [g_i^{(1)}, \ldots, g_i^{(K)}] = \psi_g(\mathbf{x}_i, \mathbf{z}_{\mathcal{N}_i}),$$

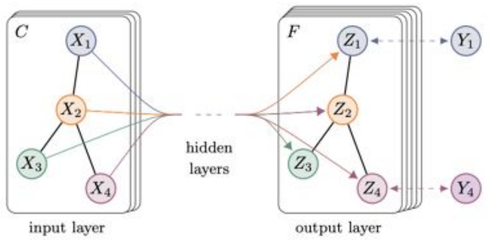
$$\mathbf{g}_i = \mathrm{FC}^{\sigma}_{\theta_g}\Big(\mathbf{x}_i \oplus \max_{j\in\mathcal{N}_i}(\{\mathrm{FC}_{\theta_m}(\mathbf{z}_j)\}) \oplus \frac{\sum_{j\in\mathcal{N}_i}\mathbf{z}_j}{|\mathcal{N}_i|}\Big)$$

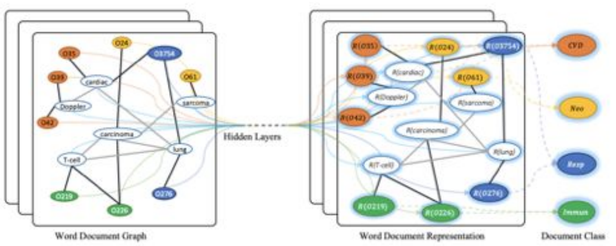


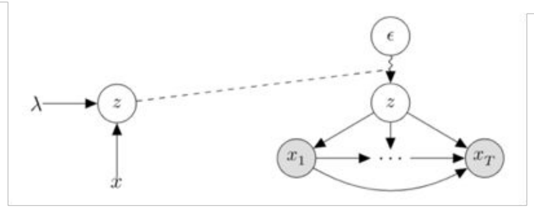
Previous Research 1: Gated Attention Graph Network

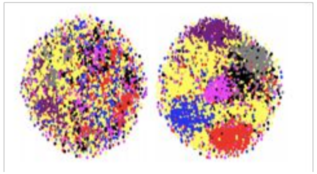
Challenges

- **Preserving semantic knowledge for better text analysis**: Use of Graph's Topological Structures.
- **Latent Semantic Analysis**: An NLP Technique for conceptual analysis which produces concepts to analyze relationship between a set of documents and the terms they contain.
- **Yet, lack of explainability for the latent space.**









References from https://arxiv.org/pdf/1812.06834.pdf, and https://arxiv.org/pdf/1809.05679.pdf

Previous Research 3: Distributional Semantics

Proposal: Distributional Semantics, why important?

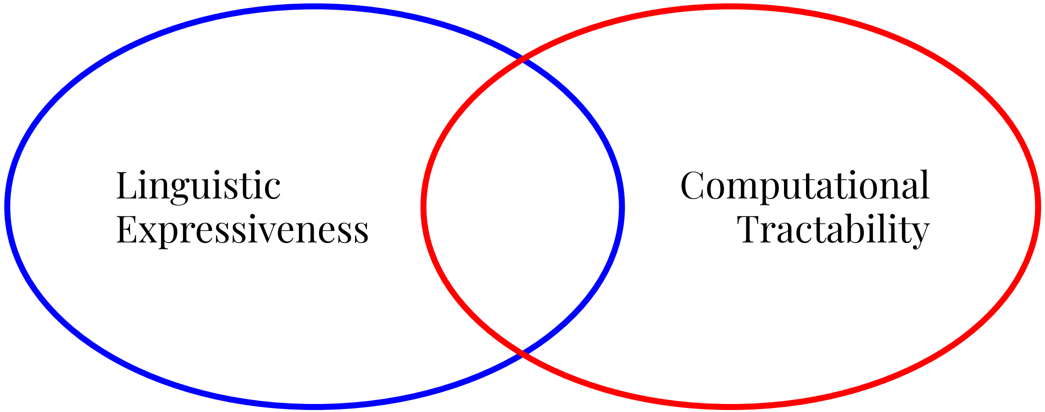
- Individual Corpora
  - "Meaning is negotiated between two people."
  - Words from every speaker with every possible use are not actually realistic in terms of what an individual understands a meaning of a word.
  - We crucially don't know about negotiation of meaning between speakers when we get a very large text corpus.
  - Corpora which are individuated so that we know what an individual speaker hears and says as opposed to a very large, huge text corpora we have. (Variations between speakers)

- A longitude of Corpora
  - After serious scientific investigation on individual corpora.
  - Corpora collected from individual speakers help work on syntax, speaker variation.

- Proposal by Prof. Copestake
  - In 2013 Berlin, https://www.youtube.com/watch?v=b40Up7KPseA

Previous Research 3: Distributional Semantics
Idea Overview

- While linguistic insights can guide the design of model architectures, future progress will require **balancing** the often conflicting demands of <u>linguistic expressiveness</u> and <u>computational tractability</u>.



Previous Research 3: Distributional Semantics

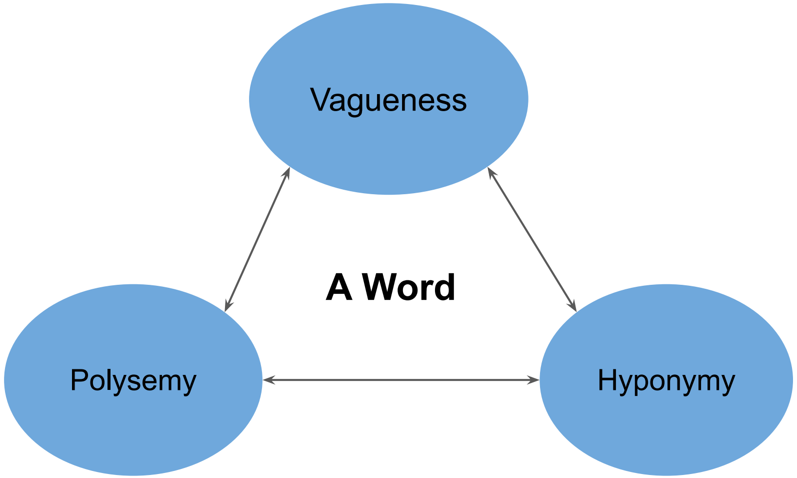
Meaning and the World

- Grounding
  - Process of connecting language to the world
  - If the meanings of words are defined only in terms of other words, these definitions are circular (Harnad, 1990)
- Approaches
  - Train a distributional model as normal, then combine it with a grounded model.
  - Find correlations between distributional and sensory features.
  - Joint learning
    - Define a single model, whose parameters are learnt based on both corpus data and grounded data.
    - Latent Dirichlet Allocation (LDA) Model
    - Skip-gram model
- Pure distributional models look for word co-occurrence patterns, while joint models prefer co-occurrence patterns that match the grounded data.

Previous Research 3: Distributional Semantics
Semantic Challenges in Linguistic Perspective

- Lexical Meaning

No sharp cutoff between concepts

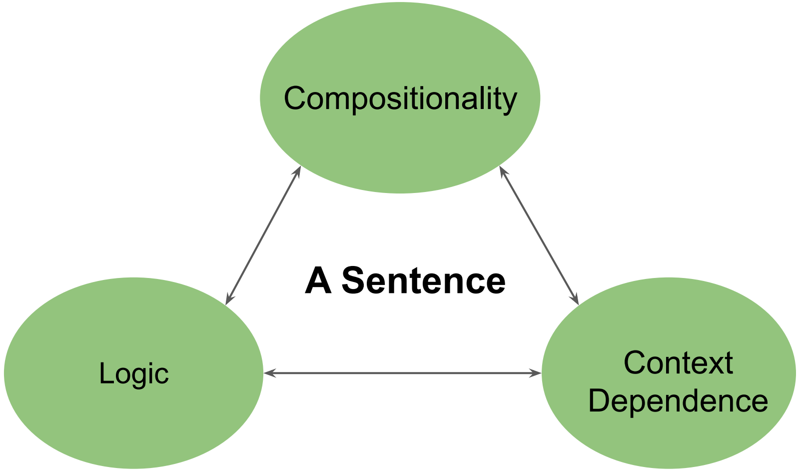


With related senses among distinct

One term that has more general meaning than another

- Sentence Meaning

Ability to derive the meaning of a sentence from its part, and to generalize to new combinations



Complex thoughts and chains of reasoning

The meaning of an expression depending on the context

Previous Research 3: Distributional Semantics

Sentence Meaning: <u>Logic</u>

- A truth-conditional functions that are compatible with first-order logic (Emerson and Copestake, 2017b; Emerson, 2020b)

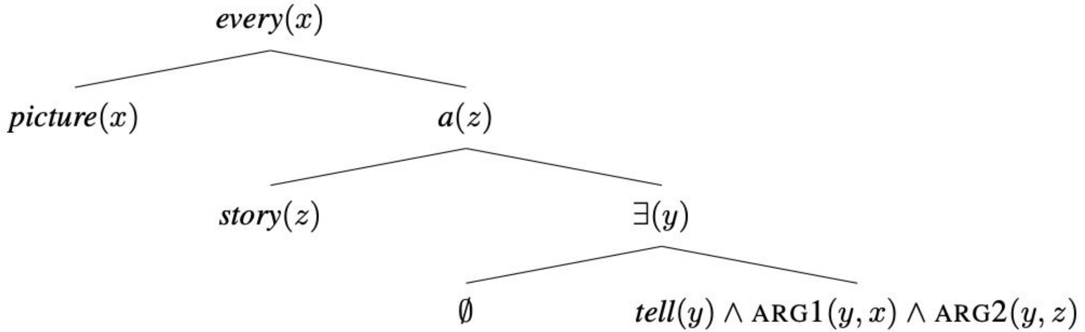


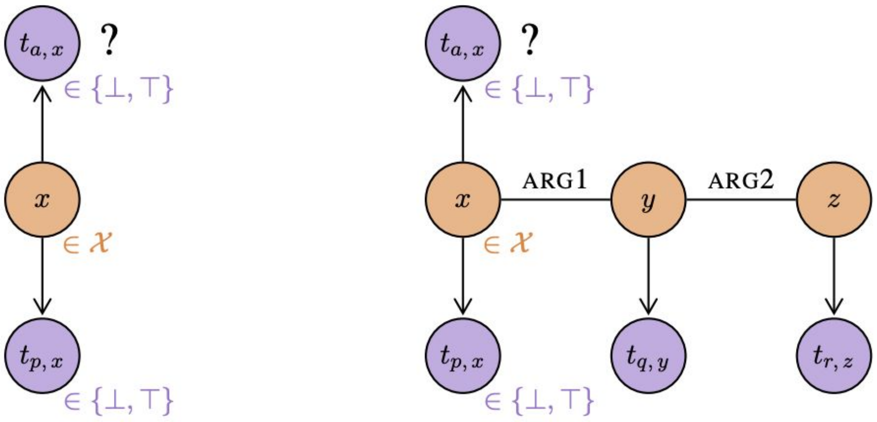
Figure 3: Fully scoped representation of the most likely reading of *Every picture tells a story*. Each non-terminal node is a quantifier, its left child its restriction, and its right child its body. We assume that event variables are existentially quantified with no constraints on the restriction.

Emerson, Guy & Copestake, Ann. (2017). Semantic Composition via Probabilistic Model Theory.

Previous Research 3: Distributional Semantics

Sentence Meaning: Context Dependence

- **Calculate occasion meanings** by conditioning on the context (Emerson and Copestake, 2017b)



(a) Inferring if a is true of x, when there are no further pixies in the situation, and given the truth value for the predicate p.

(b) Inferring if a is true of x, when x is part of a larger situation, and given the truth value for one predicate for each pixie in the situation.

Figure 4: Examples of graphical models for inference. In both cases, we want to find the probability of the predicate a being true of a pixie x, given some other information about the situation.

Emerson, Guy & Copestake, Ann. (2017). Semantic Composition via Probabilistic Model Theory.

Previous Research 3: Distributional Semantics

The Aim of Distributional Semantics

- Requires structure beyond representing meaning as a point in a space
    - It seems desirable to represent the meaning of a word as a region of space or as a classifier, and to work with probability logic.
    - However, there is a trade-off between <u>expressiveness</u> and <u>learnability.</u>
        - Promising neural architectures for working with structured data
            - Dependency Graphs
            - Logical Propositions
        - Promising new techniques to mitigate computationally expensive calculations in probabilistic models
            - Amortized variational inference, used in the Variational Autoencoder

References

- Asai, Akari, et al. "Learning to retrieve reasoning paths over wikipedia graph for question answering." *arXiv preprint arXiv:1911.10470* (2019).
- Spinelli, Indro, et al. "Efficient data augmentation using graph imputation neural networks." *arXiv preprint arXiv:1906.08502* (2019).
- Wang, Haoyu, et al. "Do Multi-hop Readers Dream of Reasoning Chains?" *arXiv preprint arXiv:1910.14520* (2019).
- Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).
- Reimers, Nils, and Iryna Gurevych. "Sentence-bert: Sentence embeddings using siamese bert-networks." *arXiv preprint arXiv:1908.10084* (2019).
- Deng, Yuntian, et al. "Latent alignment and variational attention." Advances in Neural Information Processing Systems. (2018).
- Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *International Conference on Learning Representations (ICLR)*, 2017.
- W. Hamilton, Z. Ying, and J. Leskovec. Inductive representation learning on large graphs. In *NIPS*, pages 1025–1035, 2017a.
- P. Velicˇkovic´, G. Cucurull, A. Casanova, A. Romero, P. Lio`, and Y. Bengio. Graph attention networks. In *ICLR*, (2018).
- Brendan L. Douglas. The Weisfeiler-Lehman method and graph isomorphism testing. *arXiv preprint arXiv:1101.5211*, 2011.
- Boris Weisfeiler and A. A. Lehmann. A reduction of a graph to a canonical form and an algebra arising during this reduction. *Nauchno-Technicheskaya Informatsia*, 2(9):12–16, (1968).
- Zhou, Jie, et al. "Graph neural networks: A review of methods and applications." *arXiv preprint arXiv:1812.08434* (2018).
- Zhang, Jiani, et al. "Gaan: Gated attention networks for learning on large and spatiotemporal graphs." *arXiv preprint arXiv:1803.07294* (2018).
- Xue, Lanqing, Xiaopeng Li, and Nevin L. Zhang. "Not All Attention Is Needed: Gated Attention Network for Sequence Data." *arXiv preprint arXiv:1912.00349* (2019).
- Emerson, Guy & Copestake, Ann. (2017). Semantic Composition via Probabilistic Model Theory.
- Emerson, Guy. "What are the Goals of Distributional Semantics?." *arXiv preprint arXiv:2005.02982* (2020).