Spectral Clustering for Axiom Selection

Zishi Wu

Department of Computer Science
University of Miami

May 17, 2020

UNIVERSITY
OF MIAMI

Outline

Introduction

Definitions

- What is Automated Theorem Proving (ATP)?
- *Logical formulae* are statements about a domain:
- Show that the *conjecture* is a *logical consequence* of the axioms.
- Example:
  - Axiom 1: *All men are mortal*.
  - Axiom 2: *Socrates is a man*.
  - Conjecture: *Socrates is mortal*.

Logical Consequence

Logical Consequence

- Every model of the axioms is a model of the conjecture.
- A set of axioms has a *model* if there is an *interpretation* (assignment) of boolean values to the axioms such that the conjunction of the axioms evaluate to *True*.
- If we list all interpretations of N formulae on a truth table, we get 2^N rows.
- The faster method is to show that the union of the axioms and the negation of the conjecture is *unsatisfiable*. $Ax \cup \neg C = \emptyset$
- In other words, if no model of the axioms is a model of the negated conjecture, then all models of the axioms are models of the conjecture.

Problem Statement

Problem Statement

- A problem consists of one conjecture to be proven, and a large number of axioms to be considered.
- But the solution set(s) usually consist of a few axioms.
- How do we select the needed axioms?

Data Summary

MPTPTP2078 Dataset

- **MP** = Mizar Problems [Urb06]
  - Library of problems in formalized mathematics, in first-order logic
  - Part of the Mizar System, which contains of a proof checker, and a language for writing mathematical definitions and proofs.
- **TPTP** = Thousands of Problems for Theorem Provers [Sut17]
- **MPTP** = Mizar Problems for Theorem Provers [Urb06]
- Two versions of each problem
  - Bushy = smaller version (3 to 40 axioms, 1 to 15 needed)
  - Chainy = larger version (10 to 500 axioms, 2 to 119 needed)

Related Work

Automated Theorem Proving (ATP) Systems

- E [Sch13]
- Vampire [KV13]

Methodology

Data Representation

- NOTE: this should be a comprehensive summary of the entire methodology, not details of one part
- Qinghua designed a dissimilarity metric [LXH17]
- Problem is converted into an undirected fully-connected graph
    - Vertices $V = \{\text{Axioms} \cup \text{Conjecture}\}$
    - Edges $E =$ dissimilarity weights between vertices

Graph Theory

Spectral Graph Theory [Chu97]

- Adjacency matrix A consists of similarity values between vertices
- Degree matrix D is a diagonal matrix where the i^{th} element is the sum of the elements of the i^{th} column of A
- Un-normalized Graph Laplacian matrix
  - $L = D - A$
- Normalized Graph Laplacian matrix contains *features* of the graph
  - $L_{norm} = I - (D^{-1/2}\ L\ D^{-1/2})$

Example Graph

Calculate Normalized Laplacian Matrix

- Adjacency, Degree, and Un-normalized Graph Laplacian

$$A = \begin{bmatrix} 0 & 3 & 0 \\ 3 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad D = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad L = \begin{bmatrix} 0 & 3 & 0 \\ 3 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

- Normalized Graph Laplacian

$$L_{norm} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 3 & 0 \\ 3 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Selection Method

Spectral Clustering [vL07]

- Given graph $G = (V, E)$
- Partition vertices in V into k clusters: $C_1, C_2, ..., C_k$
- Denote the cluster containing the conjecture as C_C
- Problem may have more than one set of solutions
- Succesful conjecture cluster only needs to contain one solution

Spectral Clustering Algorithm

A Tutorial on Spectral Clustering [vL07]

1. Construct a weight matrix W (i.e. adjacency matrix A)
2. Compute the normalized Laplacian matrix L_{norm} from W
3. Compute the first k eigenvectors $v_1, \ldots, v_k$ of L_{norm} and construct a feature matrix U from those eigenvectors
4. For $i = 1, \ldots, n$, let p_i be the feature vector for the i^{th} vertex, corresponding to the i^{th} row of U
5. Cluster the vertices based on their feature vectors into k clusters: $C_1, C_2, \ldots, C_k$

Spectral Clustering Algorithm

K-Means Initialization Problem

5. Cluster the feature vectors into k clusters: C1, C2, . . . , Ck
    - Each run of k-means chooses a different set of initial centroids for the *k* clusters
    - Results in different clusterings each run
    - We need a deterministic way of clustering that doesn't change over multiple runs of k-means

F. R. K. Chung.
Spectral Graph Theory.
American Mathematical Society, 1997.

L. Kovacs and A. Voronkov.
First-Order Theorem Proving and Vampire.
In N. Sharygina and H. Veith, editors, *Proceedings of the 25th International Conference on Computer Aided Verification*, number 8044 in Lecture Notes in Artificial Intelligence, pages 1–35. Springer-Verlag, 2013.

Q. Liu, Y. Xu, and X. He.
New terms metric based on substitutions.
In *2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, pages 1–6, 2017.

S. Schulz.
System Description: E 1.8.
In K. McMillan, A. Middeldorp, and A. Voronkov, editors, *Proceedings of the 19th International Conference on Logic for Programming, Artificial Intelligence, and Reasoning*, number 8312 in

Lecture Notes in Computer Science, pages 477–483. Springer-Verlag, 2013.

G. Sutcliffe.
The TPTP Problem Library and Associated Infrastructure. From CNF to TH0, TPTP v6.4.0.
Journal of Automated Reasoning, 59(4):483–502, 2017.

J. Urban.
MPTP 0.2: Design, Implementation, and Initial Experiments.
Journal of Automated Reasoning, 37(1-2):21–43, 2006.

Ulrike von Luxburg.
A tutorial on spectral clustering.
Statistics and Computing, 17(4):395–416, Dec 2007.