This space is reserved for the EPiC Series header, do not use it

Evaluation of Axiom Selection Techniques

Qinghua Liu1, Zihao Wang2, Zishi Wu2, and Geoff Sutcliffe2

1 System Credibility Automatic Verification Engineering Lab of Sichuan Province, Southwest Jiaotong University, China, [qhliu@my.swjtu.edu.cn](mailto:qhliu@my.swjtu.edu.cn)

2 University of Miami, USA, zxw526@miami.edu,ry04ert39@miami.edu,geoff@cs.miami.edu

**Abstract**

Evaluation of Axiom Selection Techniques

# Introduction

GEOFF:

Intro about axiom selection. Most evaluation by running ATP, the ”proofs in the pudding”. Takes time, propose Quantitative metrics based on selection. Our methods. Evaluation vs Vampire and E. Section on selection techniques - cuttion (eg Isabelle) and projection (eg SInE).

# Selection Metrics

GEOFF: Description of metrics.

# Our Selection Techniques

GEOFF: Intro

QINGHUA: Qinghua’s distance

## Infinity Cut

1. Qinghua’s infinity cut

## A(nother) Machine Learning Approach

1. Qinghua’s ML?

## Our Selection Techniques

In this chapter, 3 methods of axiom selection based on the graph was proposed. First, using spectral clustering to select about half of the needed axioms. Second, based on the first step, using local search to find path from the conjecture to one of it’s infinity distance axiom. The local search algorithm is a greedy strategy by finding the optimal solution only based on the current state. There are local beam search algorithm is implemented. There are 2 experiments. The first one is finding a random path which each node is visited its nearest neighbor. The process starts from conjecture to one of the axiom that is the infinity distance from the conjecture to them. And second one is find all paths from the conjecture to infinity distance node connected to the conjecture. After step 2 and 3, a spectral cluster algorithm which include more clusters is implemented. Based on the path from the step 2 and 3, each path node’s cluster set node will be added to the path set.

## Graph cut based on the spectral clustering

The graph will be divided into 2 clusters by using spectral clustering. Most of the partitions can include all need axioms in one cluster.

Based on the graph weighted by the dissimilarity, a new graph with the same amount of vertices and edges are generated by weighted by similarity by using the Gaussian Radial basis function kernel showed below:

S=e-D

In the formula, S is the similarity matrix, D is the dissimilarity matrix. After generating the similarity matrix, the spectral clustering is implemented by using python sklearn based on the similarity matrix.

The algorithm of spectral clustering of sklearn is illustrated below:

input: similarity

output: an array construct of integers which means each element’s cluser

Step 1: Generate degree matrix D from the similarity matrix

Step 2: Generate Laplace matrix L

Step 3: Calculate normalized Laplace matrix L1 by using Ncut algorithm

Step 4: Calculate L1’s eigenvalue and each eigenvalue’s eigenvector f

Step 5: Use k-means based on the normalized eigenvector

The steps illustrated above are packaged in the sklearn library. After spectral clustering, the majority automated theorem proof (ATP) task can separate in to 2 parts which has 1 part includes all the needed axioms. For ATP tasks which includes all the needed axioms in one cluster, a subgraph will be generated for greedy search. Otherwise, the original graph will be used for the greedy search.

## Local beam search for axioms selection

The local beam search uses greedy search strategy which is find the local optimal state based on the current state. However, the local beam search not only kind find one local optimal state, but also keeps track k states. In this study, the definition of the local optimal solution are the minimum dissimilarity axiom nodes connected to the current node. The local beam search is implemented based on the breadth first search (BFS).

In this study, the BFS only considers the neighbors which has the minimum dissimilarity. The BFS start at the conjecture and finish once reach an infinity dissimilarity node of the conjecture (terminate node). There are 2 experiments. The first is randomly find a minimum dissimilarity node if there are more than 1 local optimal solutions. The second is search all minimum dissimilarity nodes for all local optimal solutions.

In the first experiment, BFS started at the conjecture and stop when reach an infinity dissimilarity node connected with the conjecture. A FIFO queue is implemented for the BFS search. The pseudocode is illustrated below.

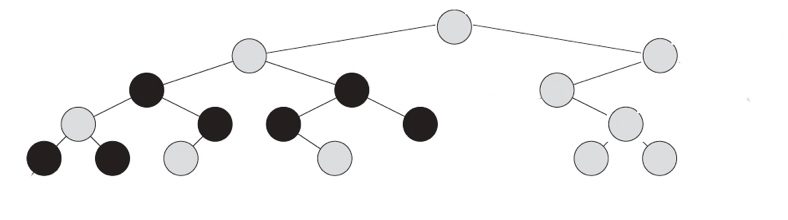


In the pseudocode, s is the start vertex. White means the node haven’t be visited yet. Gray means the node is been visiting now. Black means the node’s all neighbors have been visited. The attribute d means the level of the vertex. Lines 1–4 mark every vertex except the start vertex, set u.d to be infinity for each vertex u, and set the parent of every vertex to be NIL. Line 5 paints s gray,. Line 6 initializes s.d to 0, and line 7 sets the predecessor of the source to be NIL. Lines 8–9 initialize Q to the queue containing just the vertex s.

The randomness happened due to the order of enqueue is random. After dequeue the first node u, its child v will recode its parent to be u. After an terminate node is visited. The backtrack process will be implement by recursively search node and its parent. The path is saved in a set without consider the order.

The second experiment is finding all greedy path if one node has more than 1 minimum dissimilarity neighbors. The first step is finding a maximum level of the terminate node N. The second step is doing BFS until u.d=maximum level. All black nodes are added to the path set.

This method might select more unnecessary axioms. There are 3 cases in different branch paths based on the maximum level nodes. The 3 cases are illustrated in the Figure 1.



case 3

case 2

case 1

Figure 1: different branch paths

The root represents the conjecture and the black nodes represent the terminate node.

In case 1, the node u with u.d=maximum level and u is a terminate node. This kind of path is seen as a select enough path.

In case 2, the node u with u.d=maximum level, u is not a terminate node, and the terminate node t.d<maximum level. Such this path is defined as redundancy path. Some nodes’ level greater than t.d can be pruned in the path to improve the selectivity.

In case 3, node u with u.d=maximum level and the path hasn’t find a terminate node yet. This path keeps uncertainty and will be kept in the axiom set. Moreover, there is a theorem that if the terminate node hasn’t been visited when reach the maximum level, the terminate node can never be found along this path. The theorem can be proved by using contradiction.

Proof: Suppose there exist a terminate node t after the maximum level. After BFS, t.d> maximum level. This case is contradict with de definition of maximum level.

To improve the robustness of the axiom selection, a deep spectral cluster method is proposed to add more axioms to the path for both experiment 1 and 2.

## Deep spectral clustering to improve the robustness

Based on the axiom path set, a deep spectral clustering by using amortized analysis implemented to improve the selectivity of the axiom selection. The amortized idea is that suppose one node is connected with k neighbors which are necessary to prove the conjecture. In this study, k equal to 5 defined by randomness. Based on the subgraph get from the binary spectral clustering, a new spectral clustering with number of clusers defined by total number of nodes in subgraph divided by k. For each node in the path, it’s cluster will be found and the neighbors will be added to the path set.

## Result of the random greedy path selection with deep spectral clustering

Two datasets (Bushy and Chainy) from the MPTP 2078 is used to evaluate the algorithm. There are both 325 tasks in Bushy and Chainy dataset. There are 4 features to evaluate algorithm: the selection evaluation (All score), selected enough number (Enuf), selection evaluation on selected enough tasks (Enuf score) and selectivity. The definitions of the 4 features are illustrated below.

The all score is the average accuracy of successful selected tasks.

accuracy= the number of need axioms/ the number of selected axioms

If there are some needed axioms not in the selected axioms, then accuracy is 0.

Enuf is the number of tasks that the accuracy is not 0.

Enuf score is the average accuracy based on Enuf tasks.

the selectivity is the average of the number of selected axioms/total number of axioms in the task.

The evaluation result is illustrated in the Table 1.

Table 1 : evaluation result

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Bushy | | | | Chainy | | | |
| Algorithm | All score | Enuf | Enuf score | selectivity | All score | Enuf | Enuf score | selectivity |
| Random path | 0.39 | 274 | 0.46 | 0.54 | 0.03 | 201 | 0.62 | 0.4 |
|  |  |  |  |  |  |  |  |  |

Evaluation of Axiom Selection Techniques Liu, Wang, Wu, Sutcliffe

## Our Selection Techniques

1. Zishi’s way

# Evaluation Results

Section on evaluation 1. The test set(s)... Should we add tptp based set? 2. The results 3. The conclusions

Data on MPTPTP2078, Number of problems in test set, how selected (proofs, hence already solved, but possibly with axiom selection), numbers of different adequate subsets, average ratio nntp/all

# Conclusion

GEOFF: 1. Future correlate metrics with ptover performance (or do now!)

2