Head1

Head2

Head3

norm

Paper 2

Problem statement

When constructing a Chow-Liu tree, there are situations that can arise where information is lost through the estimation process. This occurs at points where the underlying tree structure of a dataset is not best represented by the particular graph. An example is depicted in \*\*\*Figure\*\*\*

It is noticed that the underlying structure is difficult to restore due to the cyclic nature of the graph.

A suggested remedy for this issue is to combine nodes into a “large node” in order to properly represent the dataset as a tree. \*\*\*Figure\*\*\* (add blurb for it as well)

As a result \*\*\*PAPER 1 \*\*\*\* proposes to created a Large Node Chow-Liu Tree, whereas combination rules are defined in order to construct a large node structure from the results of the traditional Chow-Liu algorithm. This allows the final structure to maintain a tree structure, giving it a strong resistance to overfitting problems.

Definition

As per the \*\*\*PAPER 1\*\*\*, the following notation sets up the LNCLT problem.

~maybe add this notation later if time\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\*\*maybe section on mining association rules

Learning Large Node Chow-Liu Tree

This section will summarize the required definitions and proofs required for the Large Node Chow-Liu Tree.

Combination Transformation

Definition 1: A combination transformation is defined to be a transformation in a tree. structure T. This transformation combines several nodes into a large node and keeps the connection relationship of T

A combination transformation is dicpicted in \*\*\*\*\*\*\*\*\*FIGURE where

Combination Rules

The rules, as defined in the article are as follows:

Rule 1 Sibling rule: The nodes to be combined satisfy that the set of these nodes are sibling relationship, i.e., there exists another node as their common parent.

Rule 2 Parent-child rule: The nodes to be combined satisfy that the set of these nodes can be sorted as a sequence based on a certain node as the root, in which each node is the parent node of its sequent node.

Rule 3 Association rule: The nodes to be combined satisfy that, under a given confidence level and a minimum support, the set of these nodes denoted by A forms an association rule , i.e., AC, where C is the class label.

Rule 4 Bound rule: The nodes to be combined satisfy that the number of these nodes is fewer than a given integer bound K.

Theoretical log likelihood

In order to prove that transformations following Rule 1 and Rule 2 are valid, the log liklihood of the new tree structure must be shown to be greater than or equal to the prior tree structure\*\*\*\*SEE WHY\*\*\*\*. This section will adapt the proof from \*\*\*\*PAPER 2\*\*\*\* in order to show the validity of the proposed algorithm.

Log likelihood

For some dataset and variables, the log likelihood can be written as

Where are dimensional vectors such that is the set .\*\*\*what is P\*\*\*\*. represents parent node for some node . This log likelihood applies when the dataset is fit in accordance to the Chow-Liu algorithm, (is maximized when the dataset is fit as a maximum weight spanning tree, with the weights being the mutual information between nodes)

Rule 2 Log Likelihood

For this proof we consider a tree structure as in \*\*\*\*FIGURE\*\*\*\*\*. We depict some parent structure with child . Node has child , and some structure . Node has some child structre . As a first step, we rewrite the log likelihood formulation to stipulate more clearly this particular tree structure:

If we conduct a combination transformation as per Rule 2 (depicted in \*\*\*\*IGURE\*\*\*\*\*) then we can further write the expanded formula as:

Now we can consider only the different terms between our two equations, \*\*\*insiert numbers). Using the notations in the article:

Simplifying \*\*\*INSET EUQATION\*\*\*\* and using the standar definition of entropy, we obtain:

Using conditional probability and bayes rule on the last term:

Likewise, we can do the same for \*\*\*\*INSERT EQUATION\*\*\*\*\*

Using a similar process\*\*\*\*MAYBE SHOW IF NEEDED FOR SPACE

By comparing \*\*\*insert equations, and utalizing the trait of entopy , we can eliminate the like the like terms, and compare the following:

As a result we can reanoably deduce

And likewise

\*\*\*\*\*\*\*INSERT RULE 1 PROOF IF TIME AND SPACE

==============MENTION RULES and 4, propbs not important / space to do 4

🡪final algorithm in paper? Or maybe just at the end

🡪add a quick summary of the results of the paper as well