# Conversion from non-mandatory leaf-node to mandatory leaf-node:

Assume there is a non-mandatory leaf-node hierarchical category DAG as shown in figure 1. Each node has a unique integer node id. Samples are named as letters from ‘a’ to ‘n’.

To convert this non-mandatory leaf-node DAG to a mandatory leaf-node DAG, a new child node is added to each non-root node. The new child nodes are leaf nodes and have node ids as the negative of the node id of its parent’s. These new child nodes are named as “dummy nodes” of their parent nodes.

The samples that only belongs to one node is assigned to its dummy nodes. Hence, all samples are categories into dummy nodes which are leaf-nodes. And the resulting DAG is mandatory leaf-node DAG. An example conversion is shown in figure 2. All dummy nodes are in red.

To get the original path, we simply remove the last node of the predicted path, i.e. [0, 2, 11, -11] to [0, 2, 11].

0

1

2

3

11

12

21

31

32

{a, b}

{c}

{d, c, f}

{g, h}

{i, j}

{k}

{l, m, n}

{}

Figure 1 Original non-mandatory leaf-node category DAG

0

1

2

3

11

12

21

31

32

{a, b}

{c}

{d, c, f}

{g, h}

{i, j}

{k}

{l, m, n}

-1

-11

{}

-12

-21

-2

-31

-32

-3

Figure 2 Converted mandatory leaf-node DAG

# Node-path distance and path distance definition:

0

1

2

3

11

12

21

31

32

111

121

211

1211

Figure 3 Sample DAG

The distance, *d*, from a node *i* to a path *p* is defined as following:

Take the distance from node 11 to the red path in figure 3 as example. The *shortest distance from the node to a node on the path* is 1, which is from node 11 to node 1. The *distance from the intersection node to the end node* is 2 which is from node 1 to node 121. Hence the total distance is 2\*1+2 = 4.

The distance, s, between path *p1* and path *p2*, where *p2* is the target path, is defined as following:

, if there is no *ith* element in path, *None* will be returned. If *p1[i]* is None, *p2[i]* will be used in *d(p1 [i],p2)*.

For example, the similarity between [0, 1, 11] and [0, 1, 12, 121] is:

0\*(3) + 0\*(2) + 1\*(4) + 1\*(0) = 4.

# Results of applying the non-mandatory leaf node conversion:

The test is done in the “Art Craft and Sewing” category to save time.

|  |  |  |
| --- | --- | --- |
|  | Original classification | Converted classification |
| Layer 1 Accuracy | 1.00000 | 1.00000 |
| Layer 2 Accuracy | 0.68367 | 0.92857 |
| Layer 3 Accuracy | 0.39796 | 0.48639 |
| Layer 4 Accuracy | 0.22959 | 0.26531 |
| Layer 5 Accuracy | 0.18291 | 0.17979 |
| Layer 6 Accuracy | 0.11713 | 0.11713 |
| Layer 7 Accuracy | 0.04167 | 0.07692 |
| Average distance between the predicted and actual path | 27.43537 | 25.59694 |