Large Scale Hierarchical Classification(LSHTC)

Massand Sagar Sunil

Y . Kushal

Motivation

- Wikipedia

  Assigning categories for new documents, recently edited documents.
- Quora Question Tags

  An application of LSHC is automation of the process of choosing relevant topics for a question on Quora.
- Both these tasks are crowd-sourced at present - rely on human experts.

Introduction

Hierarchy can be thought of as a tree or a DAG

- Tree – Every child category has only one parent super category
- Directed Acyclic Graph – A child category can have more than one parent super category

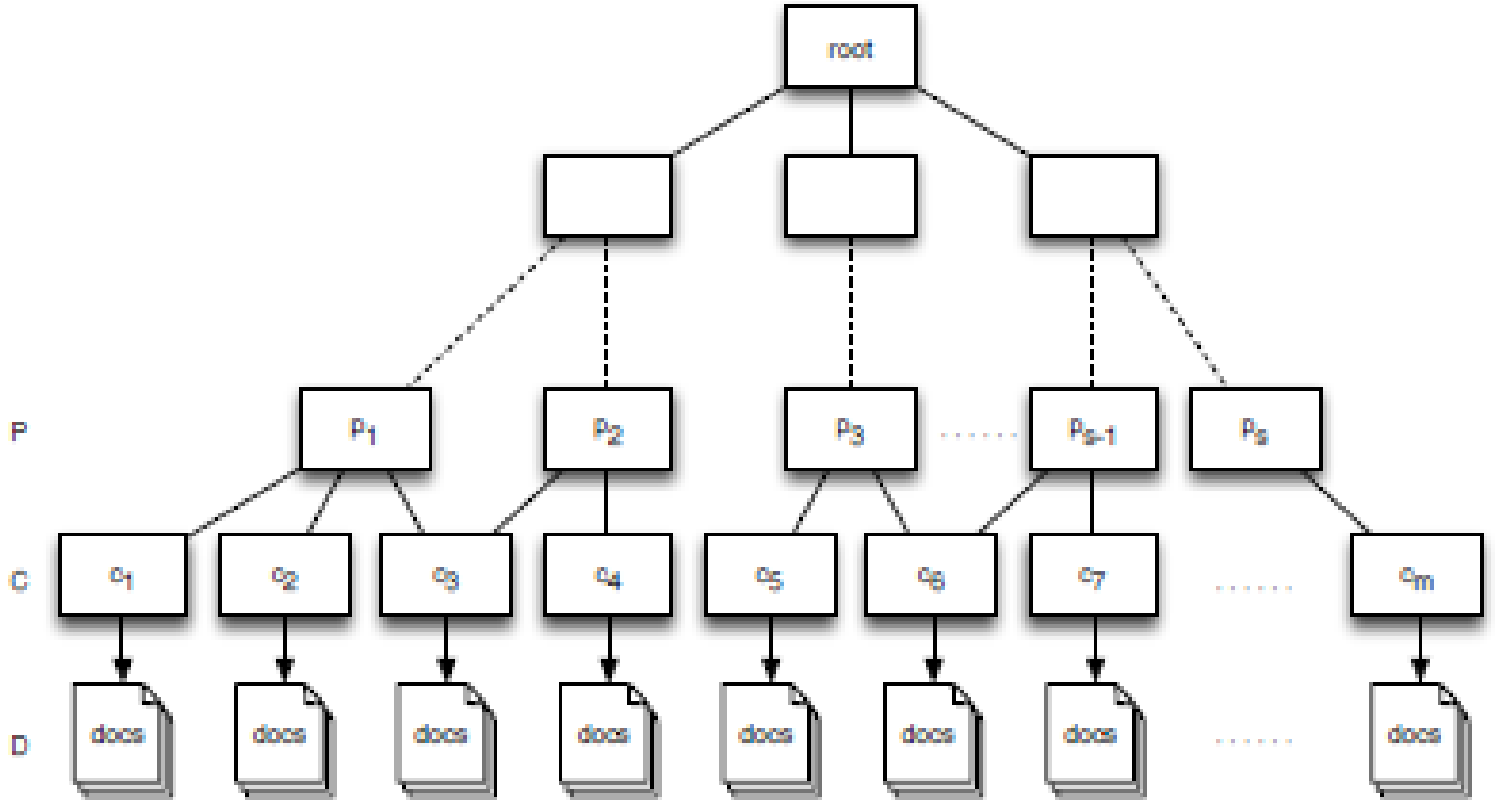


Fig. 1. Hierarchical Categorization Structure.

A simple tree hierarchy

Image taken from short paper “An Optimized K-Nearest Neighbor Algorithm for Large Scale Hierarchical Text Classication” by Xiaogang Han et al.

Introduction

- Every document needs to be categorized as one of the leaf categories.
- It is assumed that it comes under all its ancestor categories.
- Can classification be done using the semantics of the hierarchical relationships ?

A Good Classifier

- Should be using the hierarchy structure of the categories
- Accuracy
- Should be computationally efficient
  - Training time
  - Test time per query
  - Space to store the model

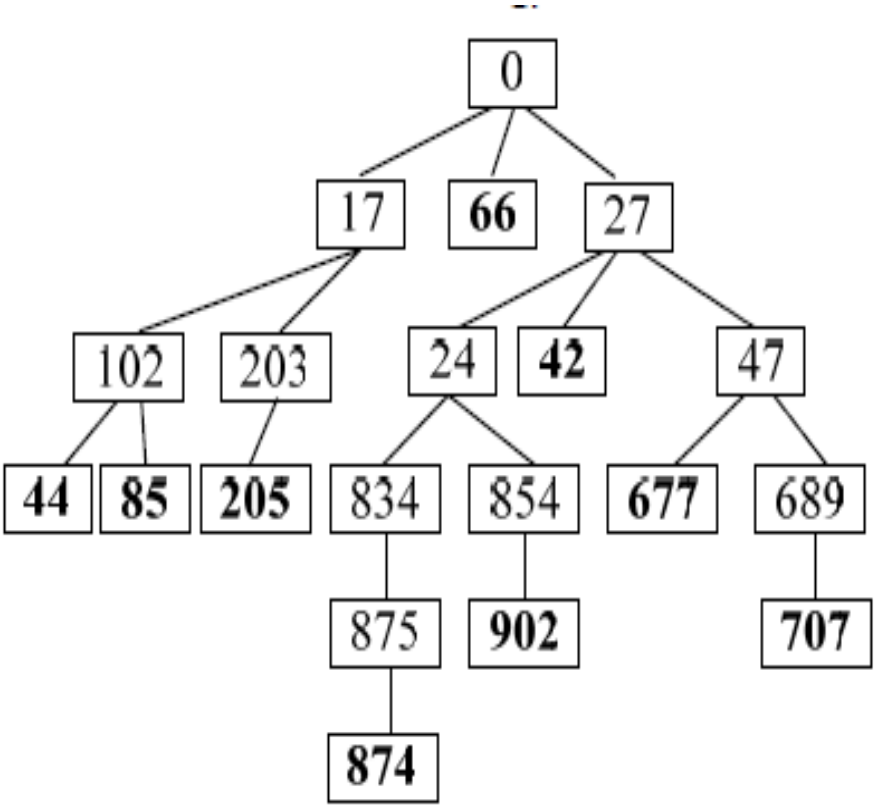
Document as a vector

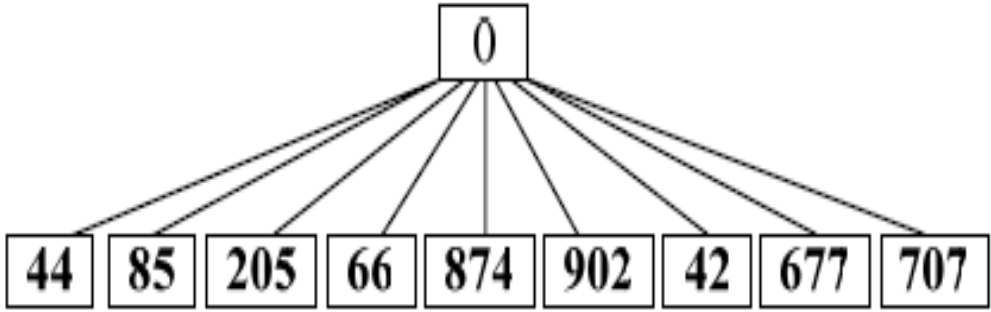
- We use the bag of words model – document is a bag of the words(terms) in it.
- A document is a vector with keys corresponding to all the words in the dictionary.
- For similarity measures we can use one of the distance measures – Euclidean, Manhattan , cosine similarity, Chebyshev etc.

Flat Classification

- Ignores the hierarchy
- The problem is reduced to ***O(k)*** one vs rest problems – where *k* is the number of leaf categories.

Flat Classification





Hierarchy Structure

Flat Classifier

Image taken from paper *"Deep Classification in Large-scale Text Hierarchies" by* Gui-Rong Xue1

Using K-NN algorithm

- This is the most popular method in practice
- Most of the methods using this algorithm do not use the hierarchy structure or use only a few levels of the structure
- We discuss an example from a short paper –

"*A k-NN Method for Large Scale Hierarchical Text Classication at LSHTC3*" by Xiaogang Han et al.

Using K-NN algorithm

- Store all of the training set document vectors for the model.
- Extract the top k similar documents – using similarity measure of choice.
- Assign labels using *weighted scoring* between the k- neighbours.
- The parameters(decided empirically) for the scoring can be set using cross- validation

Using K-NN algorithm

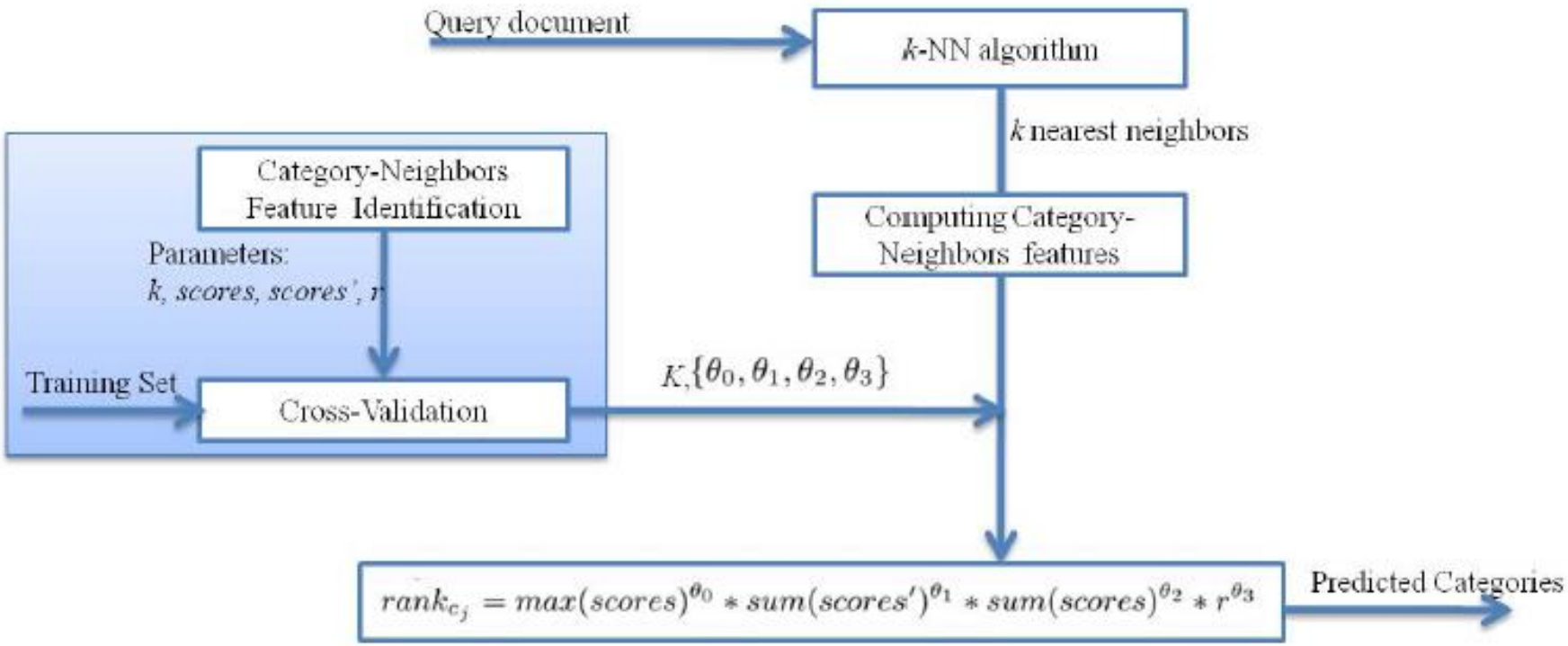


Image taken from short paper “An Optimized K-Nearest Neighbor Algorithm for Large Scale Hierarchical Text Classication” by Xiaogang Han et al.

Top-Down Approaches

- These approaches use a classifier at each internal node of the hierarchy
- An SVM classifier is more accurate while a Naïve Bayes classifier is more time–efficient.
- It is observed that accuracy difference between the SVM and the Naïve Bayes classifier decreases in the lower levels of the hierarchy
- A combination of both can be used – SVM for nodes in the higher levels and NB classifier for lower levels.

Top-Down Approaches

Method	Accuracy (%)	Tr. Time (hours)	Test Time (secs)
SVM-TD	35.58	35	20
SCS-τ, $\tau = 60$	35.19	22	12
SCS-τ, $\tau = 30$	34.68	12	5
AH-CS	**35.66**	35.25	**4**
NB-TD	22.22	0.25	0.5

Table 2. Trade-off between Prediction Accuracy in %, Total Training for entire dataset in hours, and Average Test Time per Instance in seconds

Time – Accuracy trade off

Image taken from *Adaptive Classifier Selection in Large-Scale Hierarchical Classification by* Ioannis Partalas et al. presented in ICONIP'12

Top-Down Approaches

The main disadvantage of this approach is error propagation

- A misclassification in the higher levels of the hierarchy will propagate to the deeper levels.

Deep Classification

- Two stage algorithm
- Stage 1 – Search Phase
  - Document based strategy
    - Compare training documents and test documents
  - Category based strategy
    - Construct a category vector on the basis of pre-classified documents and compare test document vector and category vector

Deep Classification

- Stage 2 – Classification Phase
  - Any of the classifications
    - Flat Classification
    - Top-Down Strategy
    - Ancestor-assistant Strategy

Ancestor Assistant Strategy

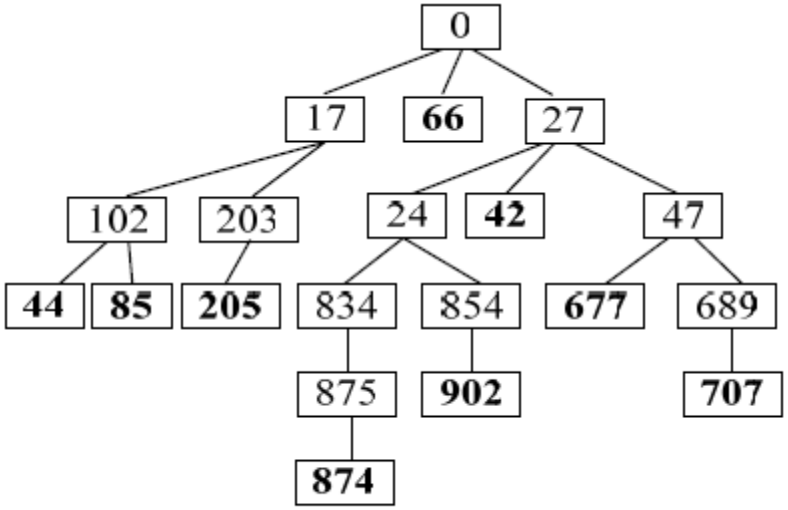


Figure 2. Pruned Hierarchy

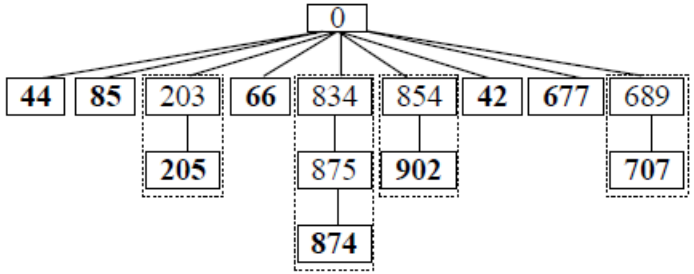


Figure 4. Ancestor-Assistant Strategy

- Pruned hierarchy transformed
- Leaf nodes use training data of those ancestors which are not common to other leaf nodes

Image taken from “Deep Classification in Large Scale Text Hierarchies” by Gui-Rong Xue et al

Evaluation

- Our idea is inspired from a similar problem hosted on Kaggle platform.
- We will also want to compare all the methods mentioned on accuracy and computational efficiency.

Dataset

- Wikipedia
  - Classification is multi – class and multi – label.
  - There are about 2.3 million documents and 300,000 categories.

References

- Deep classification in large scale text hierarchies-Gui Rong Xue, Dikan Xing, Qiang Yang, Yong yu
- An Optimized K-Nearest Neighbor Algorithm for Large Scale Hierarchical Text Classification-Xiaogang Han, Junfa Liu, Zhiqi Shen, Chunyao Miao
- A k-NN Method for Large Scale Hierarchical Text Classificationat LSHTC3-Xiaogang Han, Shaohua Li, Zhiqi Shen
- Enhanced K-Nearest Neighbour Algorithm for Large-scale Hierarchical Multi-label Classification- Xiao-Lin Wang, Hai Zhao and Bao-Ling Lu.

References

- The ECIR 2010 Large Scale Hierarchical Classification Workshop- A.Kosmopoulos, E.Gaussier, G.Paliouras, S.Aseervatham.
- Adaptive Classifier Selection in Large-Scale Hierarchical Classification- Ioannis Partalas, Rohit Babbar, Eric Gaussier, Cecile Amblard.
- On Empirical Tradeoffs in Large Scale Hierarchical Classification - Rohit Babbar, Ioannis Partalas, Eric Gaussier, Cecile Amblard.
- A Review on Multi-Label learning Algorithms- Min-Ling Zhang and Zhi-Hua Zhou.

Questions

Thank You