HIERARCHICAL MULTI-LABEL OF SHORT DOCUMENT CLASSIFICATION USING TERM EXPANSION AND LABEL POWERSET

ZAID FAROOQ SALIH

DISSERTATION SUBMITTED IN FULFILMENT FOR THE DEGREE OF
MASTER OF INFORMATION TECHNOLOGY (INFORMATION SCIENCE)

FACULTY OF INFORMATION SCIENCE AND TECHNOLOGY
UNIVERSITI KEBANGSAAN MALAYSIA
BANGI

2018

PENGELASAN DOKUMEN PENDEK BERHIERARKI PELBAGAI LABEL MENGGUNAKAN PENDEKATAN PENGEMBANGAN ISTILAH DAN LABEL POWERSET

ZAID FAROOQ SALIH

DISERTASI YANG DIKEMUKAKAN UNTUK MEMENUHI SEBAHAGIAN DARIPADA SYARAT MEMPEROLEH IJAZAH SARJANA TEKNOLOGI MAKLUMAT (SAINS MAKLUMAT)

FAKULTI TEKNOLOGI DAN SAINS MAKLUMAT
UNIVERSITI KEBANGSAAN MALAYSIA
BANGI

2018

DECLARATION

I hereby declare that the work in this thesis is my own except for quotations and summaries which have been duly acknowledged.

06 August 2020

ZAID FAROOQ SALIH
P84488

ACKNOWLEDGEMENT

First and foremost praise be to Almighty Allah for all his blessings for giving me patience and good health throughout the duration of this master research.

I take immense pleasure to express my sincere and deep senses of gratitude to my supervisor Dr. Sabrina Tiun for providing support, excellent supervision, encouragement, good teaching, brilliant ideas and motivation in order that I could complete this thesis.

To the spirit of my mother, may God have mercy on her, to my family my wife, my children, my father and all those who supported me during my studies. I would like to say thank you so much.

ABSTRACT

The process of classifying any text set aims to assign a predefined set of categories or classes for such documents in accordance to their contents. Greater challenges for text classification when the process needs to consider multi-label classes and hierarchical classification. This task aims to provide a number of appropriate classes for a single text document in a hierarchy structure. The task of hierarchical classification is getting more challenging when handling short text. Short text documents contain a limited number of words which make it highly ambiguous regarding the difficulty of extracting contextual information. Several approaches have been proposed for the task of hierarchical text classification. However, such approaches have used the One-against-all mechanism which seems to be inefficient for the short text classification. Therefore, this study aims to propose a combination of term expansion method and Label Powerset mechanism for the short hierarchical classification using the Support Vector Machine (SVM) classifier. The term expansion aims to handle the problem of ambiguity that lies behind the short text by providing semantic correspondences using WordNet dictionary. In addition, an appropriate feature extraction approach has been used with the term expansion method in order to identify the most important terms within the expanded text. Such method has been utilized by a modified version of TF-IDF, which is the Interesting Term Count (ITC). On the other hand, the Label Powerset mechanism will be utilized with the SVM classifier in order to handle the problem of hierarchical text classification. A discretization process has been applied in order to convert the hierarchy into a flat structure. To test the proposed method, a short text dataset of ACM has been used in the experiments which contains vast amount of titles and keywords related to publication articles. The evaluation has been conducted by comparing the term expansion and without applying the expansion. Experimental results have shown that the proposed method (with the term expansion) has achieved the best F-measure of 93%. This indicates the effectiveness of term expansion with Label Powerset approach in hierarchical multi-label of short document.

PENGELASAN DOKUMEN PENDEK BERHIERARKI PELBAGAI LABEL MENGGUNAKAN PENDEKATAN PENGEMBANGAN ISTILAH DAN LABEL POWERSET

ABSTRAK

Proses pengelasan mana-mana set teks adalah bertujuan untuk menetapkan set yang telah ditetapkan untuk kategori atau kelas bagi dokumen-dokumen tersebut mengikut kandungannya. Cabaran yang lebih besar dalam mengelaskan teks adalah apabila perlu mengambil kira kelas pelbagai label dengan struktur yang berhierarki. Tugas tersebut bertujuan untuk menyediakan beberapa label kelas yang sesuai dengan cara berhierarki. Tugas pengelasan hierarki ini semakin mencabar apabila perlu mengendalikan teks pendek. Dokumen teks pendek mengandungi bilangan perkataan yang terhad, yang menyebabkan semakin sukar untuk mengestrak maklumat mengikut konteks. Beberapa pendekatan telah dicadangkan pada permasalahan pengelasan berhierarki bagi dokumen pendek. Walau bagaimanapun, pendekatan-pendekatan tersebut menggunakan mekanisme *One-against-all* yang nampaknya tidak efisyen bagi pengelasan teks pendek. Oleh itu, tujuan kajian ini adalah untuk mencadangkan gabungan kaedah pengembangan istilah dan mekanisme *Label Powerset* untuk pengelasan hierarki doumen pendek menggunakan pengelas Sokongan Mesin Vector (SVM). Pengembangan istilah bertujuan untuk menangani masalah ketaksaan yang terkandung dalam teks pendek dengan memberikan hubungan semantik menggunakan kamus WordNet. Di samping itu, pendekatan pengestrakkan fitur yang sesuai telah digunakan bersama dengan metod pengembangan istilah dalam mengenal pasti istilah yang paling penting dalam teks yang dikembangkan. Pendekatan yang telah digunakan ialah versi TF-IDF yang telah diubahsuai, iaitu Kiraan Istilah Menarik (ITC). Manakala, mekanisme – *Label Powerset* telah digunakan dengan pengelas SVM dalam usaha untuk menangani masalah pengelasan teks berhierarki. Proses pendiskretan (*discretization*) telah digunakan untuk menukar struktur hierarki pada strukur rata. Untuk menguji metod kajian yang dicadangkan, set data teks ringkas ACM telah digunakan dalam eksperimen yang mengandungi sejumlah tajuk dan kata kunci yang berkaitan dengan artikel penerbitan. Penilaian telah dijalankan dengan membandingkan pengelasan yang menggunakan pengembangan istilah dan tanpa pengembangan istilah. Keputusan eksperimen telah menunjukkan bahawa kaedah yang dicadangkan (dengan pengembangan istilah) telah mencapai *ukuran-F* yang terbaik sebanyak 93%. Ini menunjukkan keberkesanannya pendekataan peengembangan istilah dan *Label Powerset* dalam pengelasan pelbagai label berhierarki bagi dokumen pendek.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 OVERVIEW

This chapter aims to provide the fundamental elements of the research in which the problem statement is being discussed. In addition, the research scope and objectives will be tackled. Finally, the methodology of conducting the research is being illustrated.

1.2 RESEARCH BACKGROUND

The last decade has witnessed a growth of the textual data over the internet. Such exponential growth has motivated many researchers to analyze this data in order to identify meaningful patterns or extracting valuable information (Simoes et al. 2009). One of the main analysis tasks is the classification using the supervised machine learning techniques. These techniques aim at allocating an appropriate class label for each text document (Zhu & Goldberg 2009). The class label can be described as a set of predefined categories such as 'right' and 'wrong', or 'low', 'medium' and 'high'. The key issue behind such task lies on the historical or examples data that have been given the exact class label, this data is being utilized for training purposes (Kotsiantis et al. 2007). In this manner, the classification techniques will be adapted to training on such historical data and identify correlations among the attributes.

Recently, several issues have arisen in the field of text classification. One of these issues is dealing with short text documents. Short text documents contain a limited number of words which make it highly ambiguous regarding the difficulty of extracting contextual information (Lu et al. 2015). Basically, dealing with short text is considered to be a challenging task in terms of the performance of the classification (Song et al.

2014).

Another issue is the hierarchical or multi-label text classification. Unlike the traditional classification task, the multi-label text classification aims to identify one or more classes for a particular document (Sucar et al. 2014). In this case, the prediction will not be depending on certain classes, but rather it would provide a probability for each document (Santos & Rodrigues 2009). The single document may classify as one or more classes. In this manner, this research aims to focus on the above-mentioned issues by examining a new method for handling multi-label text classification.

1.2.1 Short Text Classification

The last decade has witnessed a booming in Web 2.0 technology where numerous e-commerce and social media applications have been released (Brown 2010). One of the characteristics of these applications can be represented as a new type of text which is the short text. Short text can be found on the web as text messages, chat, logs, posts (related to Facebook) and tweets (related to Twitter) (Sriram et al. 2010). According to Song et al. (2014), the average length of the short text is no longer than 200 characters. In this manner, mining such forms of the short text has become a crucial task in order to facilitate the search and retrieval process. However, mining or classifying short text would encounter several obstacles. The main characteristics of short text can be described as follows:

- First, contrary to the traditional text, a short text is characterized by the noisy contained (Song et al. 2014). This noisy comes from the fact that these sorts of text have been formulated in a un-standard manner where dialects play an essential role in terms of increasing the noisy (Lu et al. 2015). Another aspect contributes toward increase the noisy is that the short text is being formed by a regular user who may do not have a grammatical capability to produce a well-organized text (Yin et al. 2015b). An example of the existence noisy can be represented as a typo in writing some words (e.g. writing ‘good’ as ‘god’). Besides the typos, grammatical mistakes and dialects could cause noisy too.

- Second, the short text contains a limited number of words which makes it tend to be highly ambiguous (Wang et al. 2014). This ambiguity comes from the difficulty of extracting contextual information from the text. For example, the sentence ‘a strange man was playing at midnight’, the word ‘play’ is referring to different meanings such as a ‘game’, ‘acting role’ and ‘fiddle an instrument’. In the traditional text classification, the ambiguity can be eliminated by determining the exact meaning of a particular word using the neighbouring words, for example, finding the word ‘guitar’ in the same sentence with ‘play’ would definitely clarify the meaning of the word ‘play’ as using an instrument (Zhang et al. 2008). In contrast, the short text usually has limited contextual information that could facilitate the process of extracting features. This makes the classification of short text is more complicated and a challenging task.

In fact, the process of classification using supervised machine learning techniques is significantly depending on two main aspects; semantic and statistics (Kotsiantis et al. 2007). Semantic aspect facilitates the process of classification by identifying the similar words in terms of the meaning (i.e. synonyms). In this manner, realizing that the two words ‘faculty’ and ‘college’ are synonyms and have the same impression, would lead to classifying the documents that contain both words into a single class (i.e. Education). In addition, the statistics aspect aims to utilize the frequency of terms in order to extract meaningful patterns, for instance, the frequency of the term ‘computer’ indicates that the document is related to computer science field.

1.2.2 Multi-Label Classification

The process of classifying any text set aims to assign a predefined set of categories or classes for such documents in accordance to their contents (Zhang & Zhou 2014). Basically, the training on such predefined classes plays an essential role in terms of classifying new, test or unseen documents. This is the core intention of supervised machine learning classification techniques. In fact, this can represent the traditional or classical classification problem.

Recently, a new challenge has been caught by several researchers which are the multi-label or hierarchical text classification (Ahmed et al. 2015). This task aims to provide a number of appropriate classes for a single text document in a hierarchy manner. According to Tsoumakas & Katakis (2006), multi-label classification methods have been divided into two main categories; *problem transformation methods* and *algorithm adaption methods*.

The first category aims to handle the multi-label classification as an algorithm independent where the categorization is being divided into multiple tasks of single-label classification. While the second category concentrates on manipulating the classification techniques such as Naive Bayes (NB), Support Vector Machine (SVM) and others. Such manipulation aims to extend the workflow of the algorithm itself in order to be able to handle multi-label text classification. In the traditional text classification, the predicting of the documents is being accommodated based on the right or wrong (i.e. 0 or 1). Whereas, in the multi-label text classification, the predicting will be divided into right, wrong and partially right.

Another version of WEKA (Hall et al. 2009) data mining software has been released which is a Multi-Label Extension MEKA. Such new version contains various types of classification methods that are intended to serve the multi-label classification problem. For example, the Multi-label K-Nearest Neighbor (MT-KNN), the Multi-label Support Vector Machine (MT-SVM) and the Multimodal Naïve Bayes (MNB).

1.3 SIGNIFICANCE OF THE RESEARCH

Since most of the applications of Web 2.0 nowadays are employing the short text, it is a vital task to analyse such text in order to provide sophisticated search and retrieval process (Bobicev & Sokolova 2008; Phan et al. 2008). Providing an accurate search and retrieval has a significant impact on multiple fields such as prediction, information extraction, sentiment analysis and question answering (Kiritchenko et al. 2014). The key characteristic behind the short text analysis lies on the challenging task of classifying such text. Challenges come from different aspects such as the ambiguity, sparsity and the hard process of extracting contextual information (Song et al. 2014). This is due to the limited number of words contained in the short text. In this manner, several approaches have been proposed to overcome this issue. Some of these approaches are relying on semi-supervised learning techniques, other depend on statistical approaches such as Latent Semantic Analysis (LSA) and Latent Dirichlet Allocation (LDA), and the rest are relying on external knowledge sources. Hence, it is a vital research study to identify the best approach for handling short text classification.

1.4 RESEARCH PROBLEM

Basically, to deal with short text classification problem, several studies have proposed numerous approaches to overcome the sparsity lies on short text (Lu et al. 2015; Mahajan & Sharmistha 2015; Yin et al. 2015b; Zhang & Zhong 2016). Generally, the key characteristic behind classifying any text lies on the frequency analysis of the words in which the classifier can attain the weight of each word. However, the traditional approaches that have been widely used for the regular text classification such as bag-of-words, term frequency, and n-gram method seem to be insufficient when dealing with short text classification. This is due to the too limited information that would be extracted from such short text which leads to increase both ambiguity and sparsity.

To overcome this issue, the text expansion has been proposed to solve the problem of sparsity and ambiguity of short text. This can be done by retrieving semantic correspondences for each word in the short text which leads to increase the contextual information. Usually, an external source is being used for such expansion mechanism such as dictionary or lexicon. One of the common dictionaries that have been used for this purpose is the WordNet (Miller 1995) which contains variety of semantic correspondences for English words. WordNet is a large dictionary that has the semantic correspondences for each word (e.g. synonyms, hyponyms and others). Hence, using such external knowledge has the ability to expand the representation of the short text which leads to better understanding of the contextual information. However, this approach suffers from the overload problem. Obviously, each word could have different meaning or sense. In this manner, retrieving the semantic senses of a particular word would bring numerous irrelevant senses. Therefore, an approach need to be used where only the correct or exact senses of specific words are used as features representation for the document (Banerjee & Pedersen 2002; Vasilescu et al. 2004; Zouaghi et al. 2012).

Recently, the short text classification has been addressed with the multi-label text classification. The traditional text classification aims to categorize a text document into a specific class label. The expansion of text information has contributed to coming up a new kind of classification which is the multi-label text classification. There is some text document that belongs to multiple class label (Santos & Rodrigues 2009). In this manner, the classification process should take into the account this hierarchy of categorization. One of the common approaches used in the hierarchical multi-label text classification is the One-Against-All. This method aims to examine each data instance in terms of belonging in accordance to each class label separately. Nonetheless, this approach suffers of the certainty where the decision of belonging is being conducted in a binary manner (i.e. either ‘0’ not belong or ‘1’ belong). Therefore, an approach which can resolve the mentioned drawback need to be chosen instead.

1.5 RESEARCH OBJECTIVE

The main goal of this research is to examine the problem of short text toward hierarchical multi-label classification. In order to accomplish the main goal multiple objectives are being adjusted which can be stated as follows:

i. To propose an expansion method using Lesk algorithm and ITC feature extraction approach for solving the problem of short text.
ii. To propose Label Powerset mechanism for the short hierarchical classification using the Support Vector Machine (SVM) classifier.
iii. To validate the combination of term expansion method and label powerset mechanism.

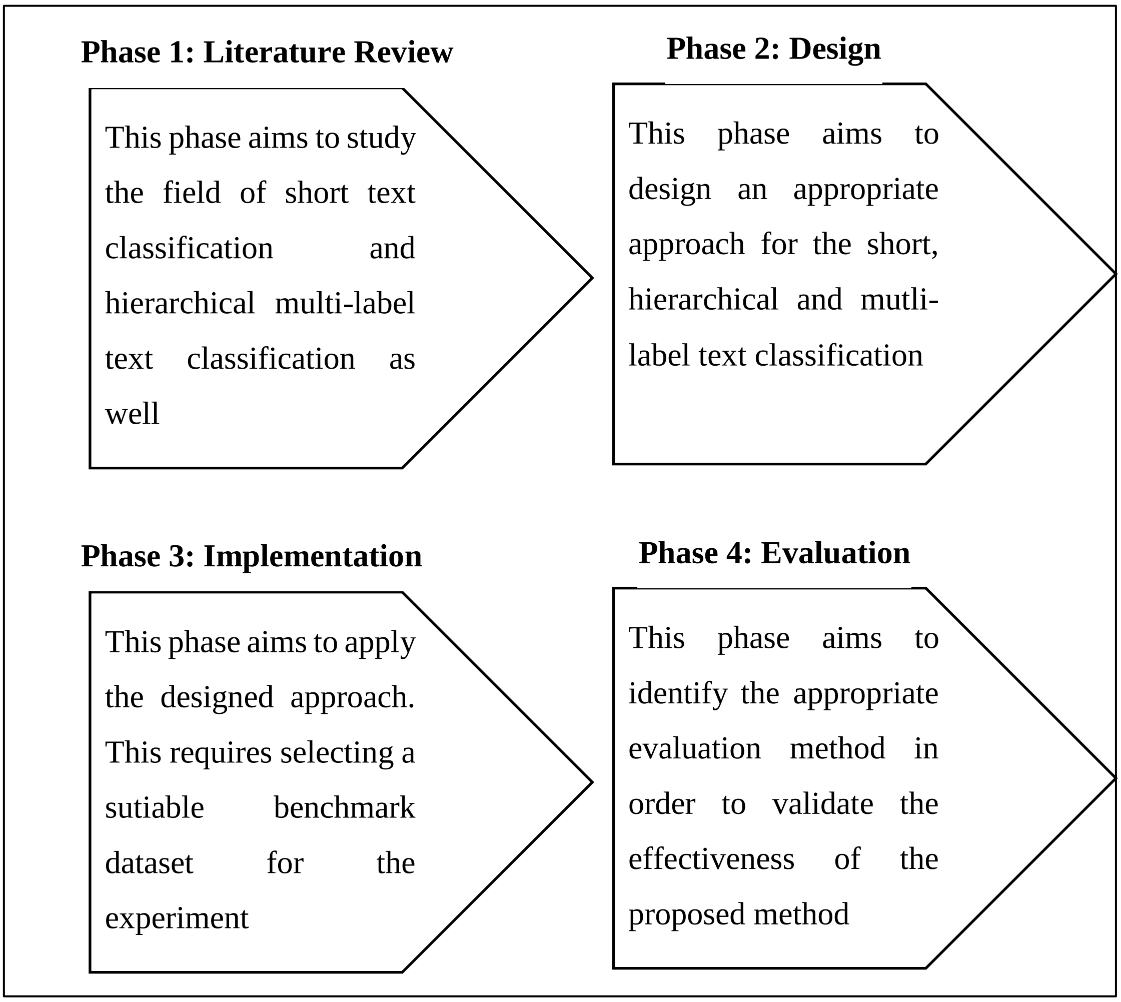
1.6 SCOPE OF THE RESEARCH

This study aims to propose a combination of expansion method and a Label Powerset approach for classifying multi-label short text classification. For this purpose, a collection of a short text document with multiple class labels in a hierarchy manner will be brought from (Santos & Rodrigues 2009) who adopted the ACM publication dataset that has been introduced in 2008. Such data consists of short documents that have numerous class labels.

Since the text is short, the expansion method will aim at expanding the context of such text. Using an open-dictionary of WordNet, each word will get multiple correspondences therefore, Lesk algorithm has been used in this study in order to assign the appropriate semantic correspondences. In addition, a modified version of the TF-IDF which is ITC will be used for the purpose of feature extraction. Finally, the Label Powerset hierarchical classification mechanism will be utilized using Support Vector Machine (SVM) classifier. Note that, 2-level of hierarchy has been used in this study

1.7 RESEARCH METHODOLOGY

This research has been conducted within four phases including literature review, design, implementation and evaluation. Literature review phase aims to survey different area of researches including short text classification, multi-label text classification, and hierarchical text classification. Such investigation aims to identify the existing limitations and gaps. Design phase aims to design an appropriate approach that would have the ability to handle short, multi-label and hierarchical text classification. Implementation phase aims to carry out the designed method in order to validate its obtained results. To do so, it is necessary to identify a benchmark dataset that can fit the experiments of classification. Finally, evaluation phase aims to assess the effectiveness of the proposed method. This can be performed by identifying an appropriate evaluation method. Figure 1.1 depicts the research methodology’s phases.



Phase 1: Literature Review

This phase aims to study the field of short text classification and hierarchical multi-label text classification as well

Phase 2: Design

This phase aims to design an appropriate approach for the short, hierarchical and mutli-label text classification

Phase 3: Implementation

This phase aims to apply the designed approach. This requires selecting a sutiable benchmark dataset for the experiment

Phase 4: Evaluation

This phase aims to identify the appropriate evaluation method in order to validate the effectiveness of the proposed method

Figure 1.1 Research methodology

1.8 THESIS ORGANIZATION

This research has been structurally organized into five chapters that are stated as follows:

Chapter I provides the main elements of the research including research background, problem statement, research objectives, research scope, and research methodology.

Chapter II provides an investigation for the literature review regarding different research areas including short text classification, multi-label text classification and hierarchical text classification. In addition, this chapter focuses on the approaches that have been proposed in the literature to handle the latter tasks. This can be represented by reviewing the state of the art classification methods and identifying the existing limitations and gaps.

Chapter III discusses the research method in which the objectives are being applied. The objectives are being depicted by carrying out an appropriate classification method for the hierarchical text classification. To discuss the experiment, this chapter illustrates the dataset that has been used to apply the proposed method.

Chapter IV discusses the results obtained by the proposed method in which the experiment settings are being tackled. In addition, the evaluation method that will be used to validate the performance of the proposed method are being discussed too in this chapter.

Chapter V provides the final conclusion and the opportunities that can be suggested for future researches.

CHAPTER II

LITERATURE REVIEW

2.1 INTRODUCTION

This chapter aims to present the literature review behind this study. This can be represented by highlighting the issue of short text classification. In addition, this chapter aims to describe the process of multi-label, hierarchical classification. Finally, a critical analysis of the related work is being provided.

2.2 MACHINE LEARNING TECHNIQUES

The exponential growth of data nowadays has posed an essential demand to analyse such data in order to extract meaningful knowledge. The basic information of an employee such as his name, tasks address, salary, and contact number can be used to determine the productivity of such employee, annual income and even his political affiliation. This can be performed using sophisticated analysis techniques such as the machine learning. Machine learning techniques aim at accommodating extensive analysis on raw data in order to extract knowledge (Witten & Frank 2005). Such knowledge can be represented in a classification form, clustering form or even in rules form. Basically, there are three paradigms included in the machine learning techniques; supervised learning, semi-supervised learning and unsupervised learning. These paradigms can be illustrated in the following sub-sections.

2.2.1 Supervised Learning

This paradigm aims to exploit historical data for the training purposes in order to classify new data (Kotsiantis et al. 2007). Historical data refers to previous data that has been annotated with class label with each instance as shown in Table 2.1.

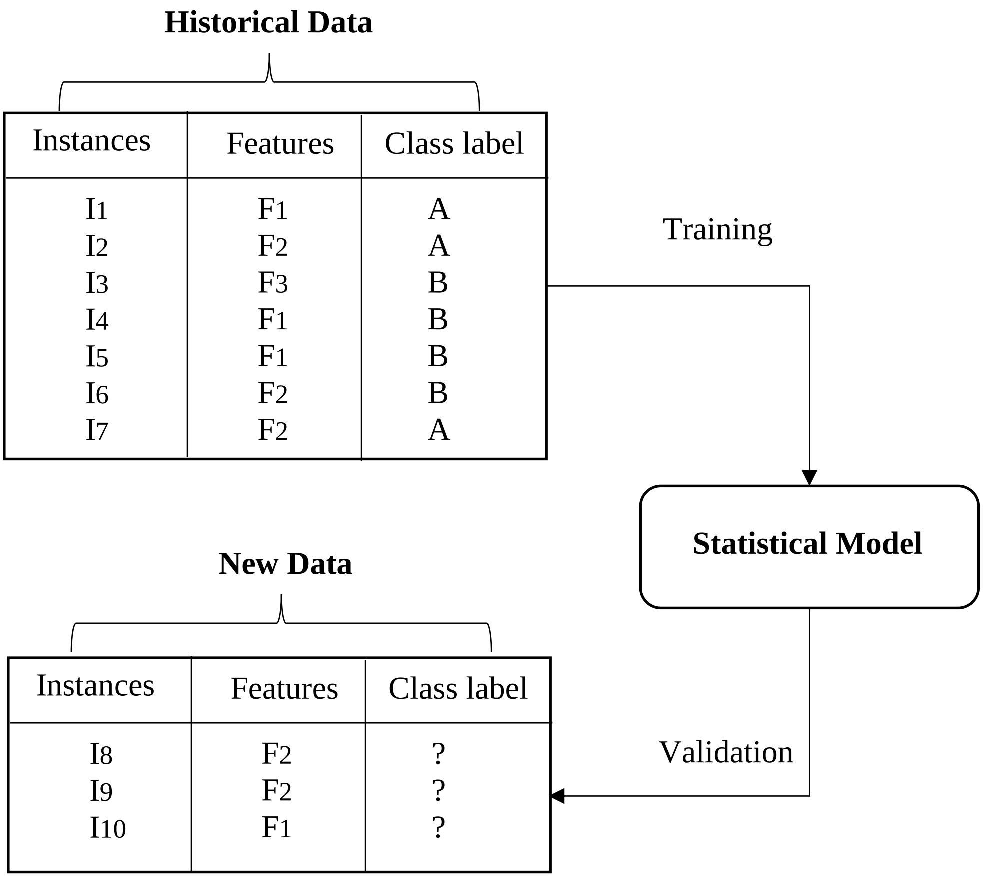
Table 2.1 Sample of historical data

Instance	**Class label**
I_1	A
I_2	B
I_3	C

The instances might be real values, nominal, numeric or text. Whereas, the class label represents the equivalent category for each instance. For example, the instance can be a numeric value that reflects an employee’s salary and the class label represents if he is qualified to apply for loan or not. Another example is a real value instance that reflects the climate temperature and the class label represents whether it would rain or not.

In this vein, the historical data is a data that has been previously annotated with the class label for each instance. Hence, the supervised machine learning technique would exploit such historical data in order to build a statistical model in the training phase. Such model aims to identify specific cases that probably or more likely associated with specific class label (Witten & Frank 2005).

After the training finishes and the statistical model is being built, the supervised learning technique can be validated by giving it a portion of new data without the class label. Based on the information of the statistical model, the supervised learning technique will classify such new data into its class label. Figure 2.1 depicts the procedures of supervised learning technique.



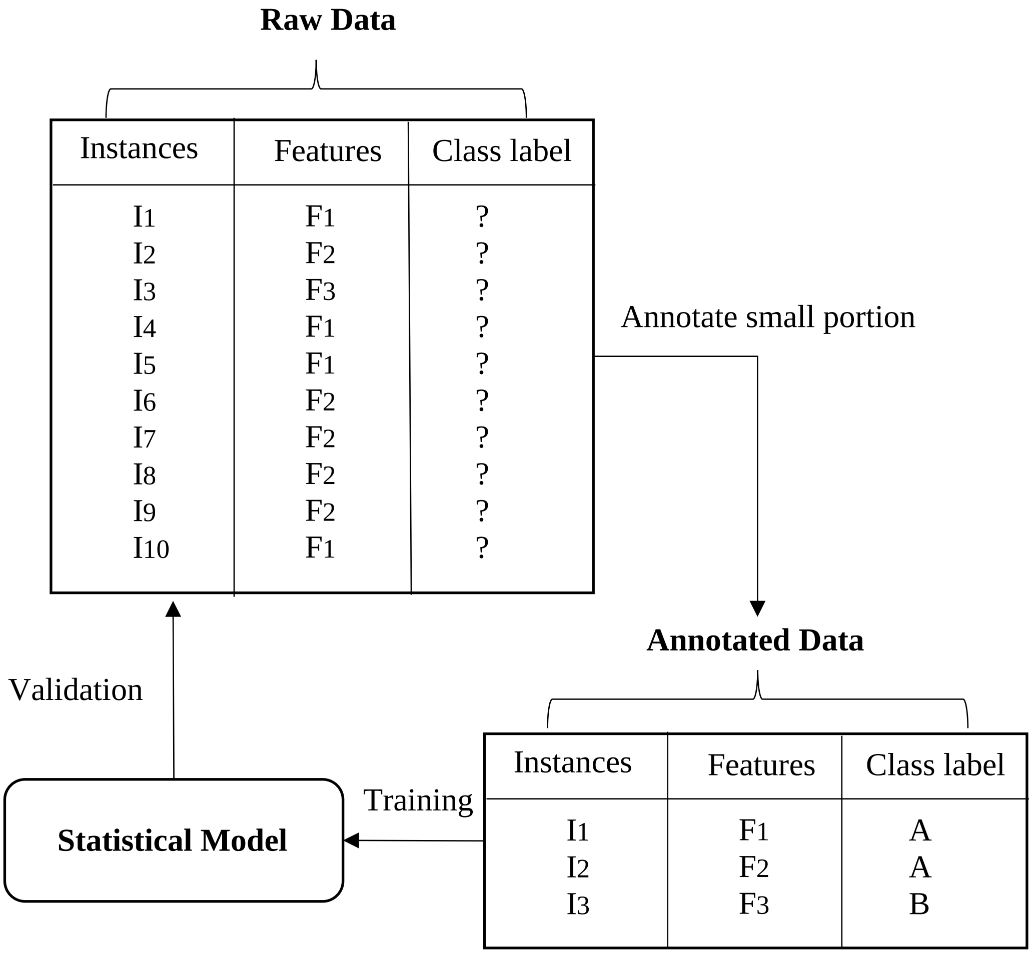
Instances	Features	Class label
I1	F1	A
I2	F2	A
I3	F3	B
I4	F1	B
I5	F1	B
I6	F2	B
I7	F2	A

Instances	Features	Class label
I8	F2	?
I9	F2	?
I10	F1	?

Figure 2.1 Procedures of supervised learning

2.2.2 Semi-Supervised Learning

Similar to the supervised learning, this paradigm is mainly depending on the training mechanism. However, this paradigm is related to the cases when acquiring an annotated data is a difficult task. In fact, providing an annotated data is not a trivial task where in some cases the historical data is confidential or contains sensitive data such as the medical records. In this manner, the semi-supervised learning techniques aim at processing the raw data and annotate a relatively small portion of the data in order to be used for the testing phase (Zhu & Goldberg 2009). In other word, the semi-supervised learning will depend on a small portion of data for the training purposes. Figure 2.2 shows the procedures of semi-supervised learning.



Instances	Features	Class label
I1	F1	?
I2	F2	?
I3	F3	?
I4	F1	?
I5	F1	?
I6	F2	?
I7	F2	?
I8	F2	?
I9	F2	?
I10	F1	?

Instances	Features	Class label
I1	F1	A
I2	F2	A
I3	F3	B

Figure 2.2 Semi-supervised learning procedures

2.2.3 Unsupervised Learning

Unlike the latter two previous learning paradigm, this paradigm does not depend on a training model. Instead, this paradigm utilizes specific functions to identify the category of each instance (Xu & Wunsch 2005). These functions could be distance function if the instances are either numeric or real values, or it could be a string-based similarity function if the instances are text. In this vein, the similarity or distance between the instances will be computed in order to make clusters. The similar instances will be joined in the same cluster. Hence, the data will be divided into clusters of similar data. Figure 2.3 shows the procedures of the unsupervised learning technique.

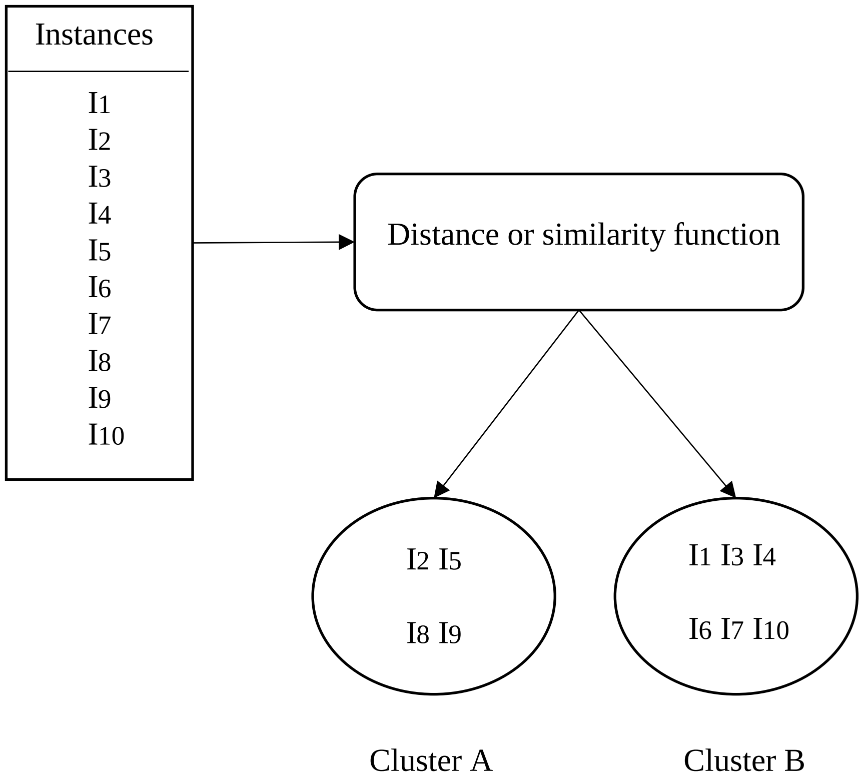


Figure 2.3 Unsupervised learning procedures

As shown in Figure 2.3, the unsupervised learning aims to utilize a distance or similarity function that will identify the similarity between the instances themselves. Hence, the raw data will be processed as input to the distance or similarity function. Then, the similar instances will be joined in one cluster (Berkhin 2006).

2.3 SUPERVISED CLASSIFICATION PARADIGMS

A wide range of applications in data mining and machine learning research community has addressed the classification task where the classification paradigm is being flat. Flat classification indicates the traditional problem of categorization in which the instances are being associated with constant number of class labels (whether binary or multiple classes). However, with the variety of real-life problems, there are several tasks from different domains have posed the use of hierarchical classification (Sun & Lim 2001). Hierarchical classification aims to deal with multiple class labels that are being arranged in a hierarchy manner. For example, a text document with a title of 'Recursive Neural Network' would be classified into a main class label which is 'Artificial Neural Network' as shown in Figure 2.4. Such class label contains two sub-class labels as 'Traditional Neural Network' and 'Deep Neural Network'. In this vein, the text document is also associated with the sub-class 'Deep Neural Network' because the Recursive Neural Network is considered to be a Deep Learning technique.

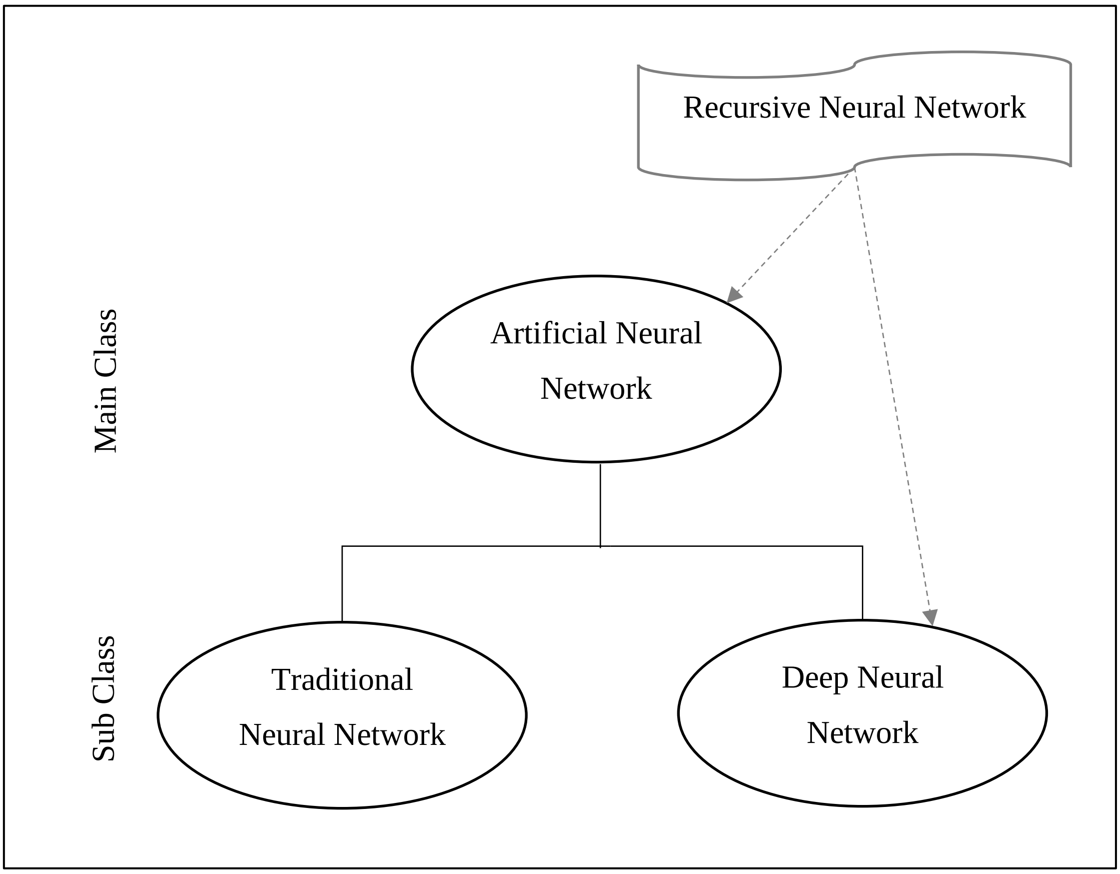


Figure 2.4 Example of hierarchical classification

Some domains such as the pattern recognition has used the hierarchical classification in different way where the classes are not being identified earlier. This means that the hierarchy of the classes can be articulated from the similarity between multiple classes. For example, in the process of recognizing the letters, the lowercase of 'q' can be a sub-class of the uppercase 'Q' (Koerich & Kalva 2005). However, some authors have claimed that although there is a common paradigm between such task and the hierarchical classification however, the main important thing in the hierarchical classification is the existence of predefined classes that are arranged in hierarchy

(Holden & Freitas 2008).

Another important factor that is necessary to be exist in the hierarchical classification, is the relation taxonomy where we can apply the 'is-a' relation between the leaf-node and the parent-node (Silla Jr & Freitas 2011). For example, in Figure 2.4, it is correct to say that 'Deep Neural Network' has a relation of 'is-a' with the parent node 'Neural Network'. In some field such as the biomedical domain, it is difficult to apply such factor in which some proteins could inherit specific characteristics from the parent node, meanwhile, it has different functions so it is incorrect to apply the 'is-a' relation (Holden & Freitas 2009).

In order to elaborate more on the hierarchical classification, let us discuss different issues including the binary classification, multiple classes classification, multi-label classification. These tasks will be discussed in the following sub-sections.

2.3.1 Binary Classification

First, the binary classification is the process of assigning a particular instance into a predefined pair of classes either 0 or 1, yes or no, and so on (Zhu & Goldberg 2009). Table 2.2 depicts an example of the binary classification.

Table 2.2 Binary classification

Instances	Class label
Impact of mobility on transaction management	Network
Time-series prediction with applications to traffic and moving objects databases	Not Network
Recovery guarantees in mobile systems	Network
Performance analysis of the link layer protocol for UWB impulse radio networks	Network

As shown in Table 2.2, the instances have been classified into either 0 or 1 this is why it has been called as binary classification.

2.3.2 Multiple-Classes Classification

Sometimes the classification may contain more than two class labels. Here the multiple classes classification can come up. It aims to classify the instances into a predefined set of class labels (more than two) (Zhu & Goldberg 2009) . Table 2.3 depicts an example of such classification.

Table 2.3 Multiple classes classification

Instances	Class label
Impact of mobility on transaction management	Wireless
Time-series prediction with applications to traffic and moving objects databases	Database
Recovery guarantees in mobile systems	Wireless
Performance analysis of the link layer protocol for UWB impulse radio networks	Protocols

As shown in Table 2.3, the instances have been classified into three distinct classes as red, blue and green. Note that, sometimes the number of classes would be big such as 50 or more. However, the key characteristic behind the multiple-classes classification can be represented as the instance is being associated with a single class from the multiple ones. This is why some researchers have called this classification as flat-classification.

2.3.3 Multi-label Classification

This paradigm of classification aims to associate the instances with one or more predefined set of classes. The key distinguish between the multiple-classes and multi-label classification is that multi-label might assign the single instance to multiple class label at the same time (Tsoumakas & Katakis 2006). Table 2.4 depicts an example of