

MS3 First Draft of Semester Report

Group number:2 Group name: 聊天室2起來
Members:B0420132 林芹學, B04201043 楊宗穎

Abstract

當我們下定決心買一支股票前，除了關心最近股價的走勢，也會關心公司本身的狀況，像是財務狀況、相關新聞等等，前者常被稱為技術面，後者則為基本面。對於沒時間每天操盤的人想長期持有同一支股票，又或是想要每年領取股利的投資人，這時候對這些投資者一項最重要的訊息就是這家公司的財務結構，當這家公司財務結構越穩健，我們就覺得這家公司不會突然宣布破產，讓股票成為廢紙。而我們認為影響公司營運狀況及獲利情況的原因不只在於決策者的智慧，產業的因素也不容忽視；例如旅遊業的淡旺季影響了公司的每季營收、油價上升影響重工業的成本因素。就算不同的公司的CFO決策風格不同，也有可能因為產業趨勢而導致公司的營收或財務結構而變得相似，而我們的目標就是找出那些相似的點。如果能找出這些產業特有的財務結構與盈利曲線模式，第一我們就有機會用過去的資料去預測同一季的營收；第二，當投資者為初次接觸某產業的公司時，可以透過此產業的特徵(characteristic)，去判斷此家公司在此產業中是否屬於常規狀態，作為投資的一個標的。

1.Background

(i)Definition

(1)Characteristic:

(a)Ratio Characteristic

這個主要分析三大財務報表：資產負債表、綜合損益表、現金流量表。由於台灣有些產業是屬於寡斷市場，資產及現金的數目相距甚遠，所以我們將以各項內容在財務報表所佔的比率去嘗試做出這個產業的財務結構。

(b)Time Series Characteristic

此項主要用於分析每月盈餘及每月平均股價，我們假設產業如果有週期的話，那這個週期必定會反映在此兩者上頭，所以我們也將這兩者也列入可以作為產業特徵的因素。

(2)Oligopoly

此項主要用來分析某些產業雖然有一些公司，我們說這個產業為Oligopoly如果前三家盈餘最多的公司的加總佔了整體產業盈餘的77%。

(ii)Some Assumptions

(1)我們相信大部份的公司及會計師是誠實的，財務報表不會有超過50%的公司作假

(2)對寡佔產業而言，由於前幾大公司可以憑藉龐大的資產去調整財務結構，對產業分析可能造成誤差，所以遇到此情況會分全部一起比較和只分析後段公司資料再和前段公司比較。

(3)我們會忽略產業內可能會有的競爭狀況。拿汽車業做例子，各家汽車業都為互相競爭關係，但有可能因為台灣環保意識抬頭，車子需求減少，雖然有競爭關係，但各家的盈利都降低了。我們的目標是考慮大環境對產業所做的功，對於內部系統的互相作用我們一概忽略。

(4)在某一產業，不會有一家公司的個別特徵就能影響到整個產業。如果是多家公司有某項特徵，則要超過66%數量的公司都擁有，否則我們都假設不會影響到整個產業。

(iii) Programming language

(1) Matlab

(2) Python

*(3) Javascript(考慮中)

(iv) Data

公開資訊觀測站

(1) 資產負債表

(2) 綜合損益表

(3) 現金流量表

(4) 每月盈餘

台灣證卷交易所

(1) 每月平均交易價格

2. Methods and Algorithms

(i) Data access

在網路上的資料通常為一季一個表格，為了獲取各季資料並整合，我們首先使用python寫出一個可以獲取目標網站的內容的自動程式，目標是把近幾年來的財務資料匯集成一個檔案。

(ii) Data munging

整理成一個資料後，接下來的問題就是清理這些資料。我們主要分為下列兩種情況來討論：

(a) 同一家公司不同時間造成資料需要整理，以下面為例：

非流動負債合計	57,983,520	20.01	62,094,107	21.21	53,118,431	18.59
負債總計	127,563,378	44.02	130,726,151	44.66	127,507,218	44.62

圖一：104年Q1台泥資產負債表

非流動負債合計	53,118,431	18.59	58,663,668	20.75	61,556,516	22.56
負債總額	127,507,218	44.62	129,062,162	45.66	127,469,535	46.72

圖二：103年Q1台泥資產負債表

負債總計				
負債總額	127,469,535	46.72	127,469,535	46

圖三：抓取完資料並合併的結果

圖一、二告訴我們“負債總額”與“負債總計”其實是同樣的意思，只是因為差了一個字，導致合併的時候會認為這兩者是不同的。為了避免這種情況，我們需要整理這些原本應為同一列的資料。

(b)不同家公司需要選擇資料，以下例說明：

投資活動之現金流量	投資活動之現金流量
取得備供出售金融資產	取得備供出售金融資產
對子公司之收購（扣除所取得之現金）	處分備供出售金融資產
取得不動產、廠房及設備	取得無活絡市場之債務工具投資
處分不動產、廠房及設備	取得採用權益法之投資
取得無形資產	取得不動產、廠房及設備
長期應收租賃款減少	處分不動產、廠房及設備
其他金融資產減少	存出保證金減少
其他非流動資產增加	取得無形資產
其他預付款項增加	取得投資性不動產
收取之利息	其他非流動資產減少
收取之股利	其他預付款項增加
投資活動之淨現金流入（流出）	投資活動之淨現金流入（流出）

圖四：104Q1台泥現金流量表

圖五：104Q1亞泥現金流量表

同樣是水泥產業的台泥與亞泥，在同一季的現金流量表也有很多不同之處，紅色框選處就是在這一家公司有這行資料且另外一家沒有，對於這種情形，還有考慮到我們是對整個產業做分析，我們決定只做超過66%同產業的不同公司共同出現的項目，其餘的我們將認為是屬於個別公司的特徵，並不影響整個產業的特徵。除此之外，我們還發現藍色框選出來的地方，雖然只差兩個字，但是他們的意義是一樣的，對於這種情況，我們需要靠人工去調整並重新統一這些標籤。

(iii) Basic visualization

對於清理好的資料，視覺化也會變得較簡單。此處我們選擇Matlab作為繪圖的工具。

(a)折線圖：每月營收、每月股價平均價格

(b)圓餅圖：三大報表

(iv)Observe trend and Average

首先是比例圖，我們選擇最簡單的全部平均。折線圖則選擇先標準化後，再取均線。

*(v)machine learning

既然我們想找出產業特徵，由於上述的平均似乎忽略了許多因素。所以我們想說透過機器學習中的，讓機器學習的演算法幫我們算出決定某家公司是屬於某個產業的重要因素為何。為了找出重要的因素(變量)，所以我們考慮決策樹(Decision Tree)或隨機森林(Random forest)演算法，來找出分界這些產業重要的因素為何。

Reference

資料出處：

- (1)公開資訊觀測站
- (2)台灣證卷交易所

參考文章：

- 1.梁亦鴻(無日期)。財務報表透視篇:「如何看懂財務報表內的秘密」。

106年5月1日，取自：[http://libsvr.sfi.org.tw/download/knowledge/102投資未來/1020501\(臺北信義社大\)-梁亦鴻副總.pdf](http://libsvr.sfi.org.tw/download/knowledge/102投資未來/1020501(臺北信義社大)-梁亦鴻副總.pdf)

- 2.作者不詳。增強財務分析的有效性有技巧。

106年5月1日，取自：<https://read01.com/Ex5Aya.html>

- 3.Turney, P.D., Cost-Sensitive Classification: Empirical Evaluation of a Hybrid Genetic Decision Tree Induction Algorithm. *Journal of Artificial Intelligence Research* 2:369-409 (1995).

參考書籍：

- 1.Joel Grus(2015).*Data Science from Scratch*. California: O'Reilly

- 2.Wes McKinney(2012). *Python for Data Analysis*. California: O'Reilly

- 3.張智星。Matlab 程式設計。106年5月1日，取自：<https://mirlab.org/jang/books/matlabProgramming4beginner/>