

Game Valuation

專題編號：107-CSIE-S017

執行期限：106年第1學期至107年第1學期

指導教授：王正豪

專題參與人員：104820018 段寶鈞

104820021 官茂原

104820027 黃千豪

一、摘要

此計畫的目的在於設計出一個可以蒐集並分析 Steam 平台上遊戲資料的系統，使用者可以將自己設計的遊戲介紹內容上傳到此系統，預測該遊戲的銷量。

首先我們使用 Python 撰寫網路爬蟲程式來獲取資料，後使用 MySQL 建立資料庫。接著再提取出資料特徵，並使用迴歸分析找出資料與銷量的關係。最後設計一網路平台供使用者輸入資料並獲取分析結果。在實驗過程中，我們會輸入已存在的遊戲進行測試，觀察測試結果後再不斷修正特徵的提取與迴歸分析以達到最高的準確率。

我們期望透過此預測系統，使用者得以調整其遊戲簡介、預覽圖或是售價來獲取更高的銷量。

關鍵詞：Steam、網路爬蟲、迴歸分析。

二、緣由與目的

許多人都會設計遊戲，但並不是每款遊戲都能受到顧客歡迎。因此我們希望能設計一個系統，它可以蒐集 Steam 遊戲平台上的資料並進行分析，找出一個遊戲介紹內容與銷售量之間的演算法。讓想將遊戲放到 Steam 上販售的人使用這套系統修正遊戲介紹內容與銷售量的預測。

三、研究範圍

Steam 平台上的遊戲。

Steam 是美國電子遊戲商維爾福於

2003年9月12日推出的數位發行平台，提供數位版權管理、多人遊戲、串流媒體和社群網路服務等功能。

四、使用技術方法

(一)網路爬蟲(web crawler)

通過網頁的鏈接地址來尋找網頁，從網站某一個頁面開始，讀取網頁的內容，找到在網頁中的其它鏈接地址，然後通過這些鏈接地址尋找下一個網頁，這樣一直迴圈下去，直到按照某種策略把該網站上所有的網頁都抓取完為止。

(二)迴歸分析(Regression Analysis)

建立因變數 Y (或稱依變數，反應變數) 與自變數 X (或稱獨變數，解釋變數) 之間關係的模型。簡單線性回歸使用一個自變量 X ，複迴歸使用超過一個自變量 $(X_1, X_2 \dots X_i)$ 。

目的在於了解兩個或多個變數間是否相關、相關方向與強度，並建立數學模型以便觀察特定變數來預測研究者感興趣的變數。更具體的來說，迴歸分析可以幫助人們了解在只有一個自變數變化時應變數的變化量。一般來說，通過迴歸分析我們可以由給出的自變數估計應變數的條件期望。

五、架構流程

(一)建立資料庫

將 Steam 遊戲平台上的資料經由網路爬蟲的方式取得資料，留取部分遊戲資料不輸入資料庫以便往後進行測試。

(二)資料分類

將資料中的各項特徵進行提取、分類並賦予其數值與加權。

(三)資料分析

將各項特徵進行迴歸分析。

(四)網頁建立

建立一個網站當作前端供客戶使用，客戶可在網站上輸入自己所開發的遊戲資料以便進行分析預測。

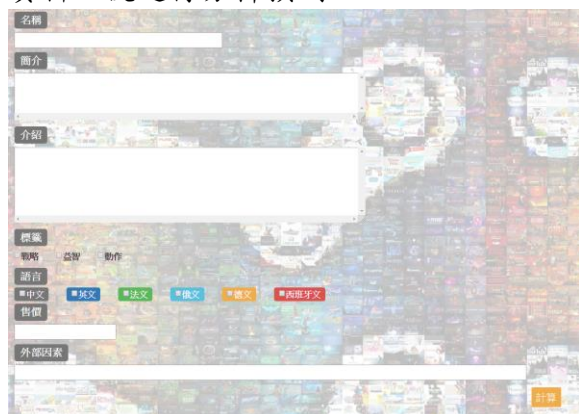


圖1. 客戶端介面

(五)實際測試

使用其他未輸入資料庫的 Steam 遊戲進行預測，觀察其準確率。

(六)修正預測結果

重複步驟(二)、(三)提高準確率。

六、工具說明

(一)爬資料

1. 語言：Python
2. 工具：Scrapy

(二)資料庫

1. 語言：SQL
2. 工具：MySQL

(三)網站架設(API)

1. 語言：Python
2. 工具：Flask

七、實驗結果

實驗結果得出的公式中價格和媒體宣傳影響較深，這兩者也是在遊戲擬訂後較好控制的因素。

八、結論

在使用者將自己的遊戲介紹相關內容(如簡介、分類、標籤、玩家人數、售價.....等等)上傳到我們設計的網站之後，它可以給出一銷量預測值。使用者獲得結果後，根據實驗結果，推薦可以修正其遊戲相關資料或增加外部媒體宣傳來獲取更大的利潤值。

參考文獻

- [1] Mark O'Neill, Elham Vaziripour, Justin Wu, and Daniel Zappala, "Condensing Steam: Distilling the Diversity of Gamer Behavior", *IMC '16 Proceedings of the 2016 Internet Measurement Conference*, pp. 81-95, 2016.
- [2] Scrapy 1.5.1 documentation (July 12, 2018). Retrieved from <https://doc.scrapy.org/en/latest/> (June 10, 2018)
- [3] scikit-learn: machine learning in Python — scikit-learn 0.20.0 documentation (September, 2018). Retrieved from <http://scikit-learn.org/stable/> (September 15, 2018)
- [4] Flask 1.0.2 documentation (May 2, 2018). Retrieved from <http://flask.pocoo.org/docs/1.0/> (September 20, 2018)
- [5] 唐新春，基于机器学习方法对销售预测的研究（2017年1月13日）。檢自 <http://www.infoq.com/cn/articles/a-study-on-sales-forecasting-based-on-machine-learning>（2018年8月18日）