



國立臺灣科技大學
資訊工程系

碩士學位論文

基於極限梯度提升的新進玩家流失預測模型應用於手機
免費遊戲數據集

New Player Churn Prediction Model Based On Extreme
Gradient Boosting Is Applied to Mobile Free-to-Play Game
Dataset

研 究 生：柯名鴻

學 號：M10915Q05

指導教授：戴文凱博士

中華民國一一年六月三十日

第 1 章 緒論

1.1 研究背景與動機

對於許多遊戲商而言，準確預測玩家流失對於長期成功至關重要。近年來，手機遊戲商大多以免費遊戲為主，不再是以往的買斷制或月費制，在免費遊戲的模式之下，遊戲商之營收有非常顯著的成長，例如：「絕地要塞 2」(Team Fortress 2) 原本為買斷制型式遊戲，新收入的來源僅限於沒購買過遊戲的人，這樣的商業模式與他們的發展策略並不完全吻合，於是，在 2011 年改為免費遊戲，除了擴大玩家受眾，還透過遊戲不斷地更新保持玩家的興趣，最終，遊戲營收提高達 12 倍之多 [1]，但是，免費遊戲在定義是否為流失玩家上極為困難，一旦玩家想離開了隨時都能停止遊玩，沒有必要告訴遊戲商其決定，遊戲商也容易因此失去挽回玩家的機會。此外，在遊戲中，獲得一個新玩家的成本也比留住一個玩家要昂貴許多。

因此，精準地預測玩家流失，即使是微小地提升，也可能導致營收有顯著的提升。透過上述情境發想，若能透過機器學習訓練出一流失預測模型並運用於遊戲領域，利用其預測結果，並從中了解玩家流失的原因與動機，將可以交由運營人員作為後續挽留玩家的操作依據，以強化留存並進一步增加營收。

1.2 研究目標

隨著免費遊戲類型的遊戲客戶獲取成本不斷提高，留住玩家成為重要議題。另外，根據網站 Swrve 於 2014 年提供的報告指出，數十款的遊戲中，有 19.3% 的新玩家只玩一次特定遊戲，新玩家的次日留存率為 33.9%，而第 30 天的留存率只剩下 5.5% [2]，可以看出新進玩家的流失率極高。因此，若能運用巨量資料探勘框架，來協助預測新進玩家是否有機會流失，將可以針對可能流失玩家作為挽留的目標並進行操作。

本論文的研究目標為預測新進玩家是否流失與了解玩家流失的原因。我們將運用一巨量資料探勘框架，包含對於資料集之前處理，透過特徵工程建立大量特徵，並藉由資料分析方法來探索資料之特性，隨後採用機器學習之分類預測，來

預測出可能流失的新進玩家，最後依其機器學習預測結果，分析各個特徵中的突出性，來進行新進玩家流失原因的解釋說明。

1.3 研究方法概述

本論文將新進玩家創帳號後的天數切分為三個時期：(1) 觀察期：玩家創帳號後前幾天，會將此時期的玩家遊戲軌跡作為資料特徵來訓練模型；(2) 挽留期：於觀察期之後，作為給運營操作的時間；(3) 表現期：於挽留期之後，主要決定玩家是否流失，如果玩家在此時期未登入則視為流失玩家，反之視為非流失玩家。

此外，本文還運用一巨量資料探勘框架：此框架將由四大階段組成，(1) 資料前處理階段：首先從資料庫群中整合所有需要的資料，並過濾掉無價值玩家，將觀察期視為資料特徵探勘期，並對其進行特徵工程，將原始特徵透過加總、平均等統計手法來建立新特徵，也額外以天為單位計算來獲得更多特徵，最後著手準備目標值，以利後續分析及機器學習使用；(2) 資料分析階段：使用前階段產出之資料，透過統計圖表來觀察資料特性，進行探索性資料分析 (Exploratory Data Analysis)，藉由流失玩家與非流失玩家的資料分佈來檢查是否有不合適之資料特徵，並觀察資料特徵是否可以提供給學習模型較多之資訊；(3) 機器學習階段：首先將處理後的資料集分割為訓練集及測試集，隨後針對訓練集進行少數群樣本權重值放大以處理不平衡資料，並透過交叉驗證 (Cross Validation) 找出機器學習模型的最佳超參數以獲得最佳模型，其中學習模型選用決策樹 (Decision Tree)、隨機森林 (Random Forest) 及極限梯度提升 (Extreme Gradient Boosting)，最後藉由測試集來驗證評估最佳模型，產出預測結果；(4) 預測結果分析階段：使用前階段產出之預測結果進行資料特徵重要性分析，透過計算各資料特徵於各學習模型中之 Gini Importance，並搭配決策樹作為代理人模型，以利更加了解及解釋資料特徵與遊戲所提供之體驗綜合評估。

在方法驗證上，本論文將藉由混淆矩陣 (Confusion Matrix) 所延伸之接收者操作特徵曲線 (Receiver Operating Characteristic Curve) [3] 與精確召回曲線 (Precision-Recall Curve) [4] 來協助驗證學習模型之優劣，並同時利用 Weighted F_β - Score [5] 來選出最佳模型與最佳參數解，隨後計算特徵重要性 (Feature Importance) 於各資料特徵中，以了解到何者於學習模型中貢獻了最多的資訊量，以利學習模型進行訓練與分類。

1.4 研究貢獻

本論文之研究貢獻為：

1. 提出一新進玩家觀察期，排除舊有玩家，以利學習模型著重於新進玩家的資訊上。
2. 提出一新進玩家挽留期，預留時間以方便運營做挽留操作。
3. 提出一新進玩家表現期，配合新進玩家觀察期，以利學習模型預測玩家是否流失。
4. 透過資料特徵工程，以統計手法建立更多資料特徵。
5. 於資料集中進行資料特徵之探勘，藉由不同種類與面向之方式，挑選出適合用來呈現流失玩家的資料。
6. 整理出適合於不平衡資料集中的評估值方式，將對於學習模型之預測結果提供合理的評估，進而進行比較。
7. 整理出資料特徵重要性之計算，以利分析資料特徵的突出性與其貢獻的資訊量。

1.5 本論文之章節結構

第 2 章 文獻探討

本章節針對免費遊戲興起介紹，並探討關於資料前處理、學習模型選擇、資料不平衡處理以及其評估方式的相關文獻。

2.1 免費遊戲興起

免費遊戲，是一種玩家無需支付任何費用，即可遊玩該遊戲之大部分內容，與付費型遊戲（買斷制或月費制）形成對比。在免費遊戲中，遊戲商可以藉由遊戲內購買或遊戲內置入廣告等方式來賺取營收 [6]。近幾年內，遊戲商皆轉以開發免費遊戲為主，因其類型所帶來之營收，已遠大於付費型遊戲 [7]。另外，免費遊戲還能夠有效的讓玩家流失量降低，透過其無需支付任何費用就能遊玩遊戲的特性，使玩家進入遊戲的門檻大為降低 [8]。

雖然免費遊戲能讓玩家流失量降低，但相對的，對於玩家是否流失變得極難定義，因為玩家可以在沒有任何通知的情況下停止遊戲，沒有明確的流失事件，例如玩家取消訂閱。儘管玩家流失模型在商業領域已經存在了幾十年，但隨著機器學習方法近年來的進步，它們的複雜性和準確性也都在提升，因此，即使在高維數據上，也可以應用極限梯度提升等機器學習來創建非常準確的模型 [9]。

2.2 資料前處理

在將巨量資料應用於機器學習前，資料的前處理也是極為重要，相較於在學習模型上進行深入研究與改進，透過資料特徵之轉化及選擇顯得更為重要且有效 [7]。

首先將對資料進行清理，只收集有價值之資料，例如：Tamassia 等人只收集遊玩時間超過給定門檻之玩家 [10]、Periáñez 等人只收集消費金額遠高於一般玩家者 [11] 或 Runge 等人只取付費玩家中前 10 % 者 [12]，上述之清理方式皆只著重於具有高資訊量的資料，而不將無價值的資料放入機器學習中。

接著，將玩家的資料以天數分割為三個時期，在本文將其依序定義為觀察

期、挽留期與表現期。將觀察期的玩家行為軌跡作為機器學習的訓練資料，表現期則用來決定玩家是否流失，然而，在玩家離開的前一天預測流失是沒有意義的，因為已經來不及在那個時候挽留玩家了，因此在觀察期與表現期之間加入挽留期，從而實施防止流失的策略，即新的促銷活動、內容更新等，以留住潛在流失者 [7]。

而針對資料特徵之探勘，Sifa 等人 [13] 將探勘玩家基本資料及玩家行為，再將其進行轉化，例如：取平均值與偏差值於玩家遊玩時間、將玩家國籍分類等。Xie 等人 [14] 使用遊戲內的事件發生頻率來預測玩家對遊戲的參與度。Lee 等人 [15] 將探勘玩家行為、玩家購買商品數量、玩家遊戲內交易及玩家遊戲內社交，主要針對玩家每日於遊戲內的行為軌跡。Gregory [9] 將玩家行為透過相對時間及絕對時間進行轉化，以此提升資料集的特徵數量。Hadiji 等人 [16] 將探勘玩家消費商品數量、玩家遊玩天數等等。

2.3 學習模型選擇

在機器學習中，於分類預測的應用上樹狀結構之學習模型最為主流且有效，因其建樹之方式，可以清楚的解釋該筆樣本之預測路徑，進而針對各資料特徵進行重要性的計算與分析 [7]。另外，在樹狀結構之學習模型中，主流以裝袋算法 (Bagging) 及提升方法 (Boosting) 兩種建樹想法為準：

- 裝袋算法：為從訓練資料集中，隨機取樣並訓練成多份分類器，而每次訓練資料取出後放回，再抽取，最後之預測結果將由多個分類器投票選出，採多數決，且各分類器間的權重關係皆為相等 [17]，如圖 2.1，例如：隨機森林 [18] 即為裝袋算法 + 決策樹 [19]。
- 提升方法：為從訓練資料集中，每次訓練使用相同資料，而第 n 個分類器於訓練時，將針對第 $n-1$ 個分類器分類錯誤的資料增大其權重值，以修正分錯的資訊，希望將分錯的資料減少，預測結果將由多個分類器投票選出，各分類器間的權重關係不同，錯誤率越低的分類器，擁有越高的權重 [20]，如圖 2.2，例如：極限梯度提升 [21] 即為提升方法 + 決策樹。

Chen 與 Guestrin [21] 實作出高效率的梯度提升 (Gradient Boosting)，稱其為極限梯度提升，除了使用提升方法建樹外，還針對錯誤修正的步驟，引入梯度下

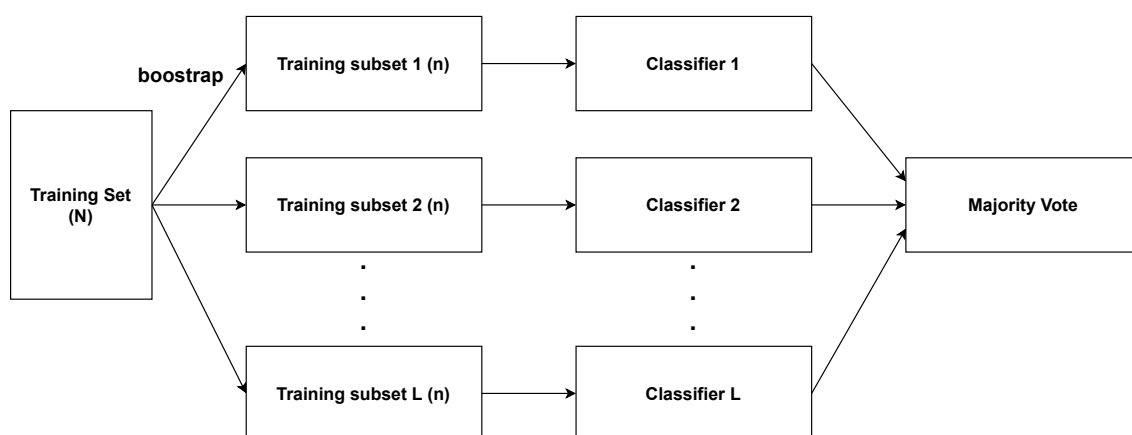


圖 2.1: 裝袋算法方式建樹示意圖 (此圖出自 [22])

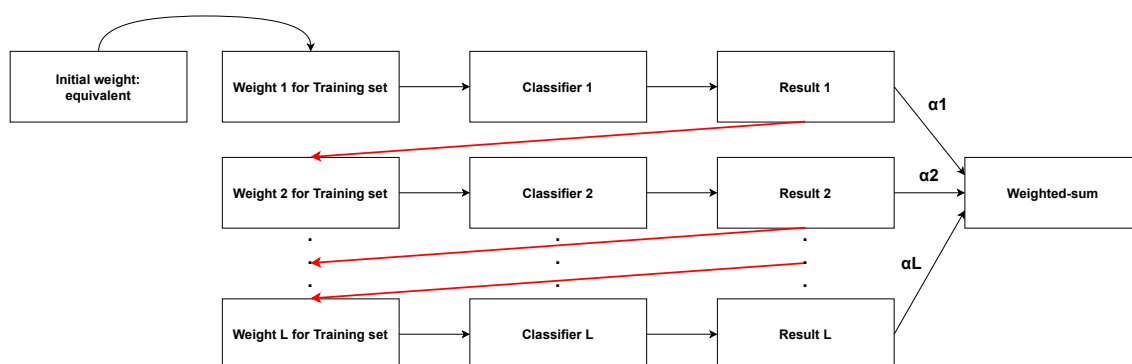


圖 2.2: 提升方法方式建樹示意圖 (此圖出自 [22])

降法 (Gradient Descent) 的概念，加速了學習模型的收斂速度，使其修正錯誤的能力更加精準，大幅減少訓練的時間成本。近年來透過極限梯度提升來訓練的研究越來越多，且其預測能力皆有不錯的表現 [9] [23] [24] [25]，明顯優於裝袋算法建樹方式的其他學習模型。

從上述得知，選用樹狀結構之學習模型將有助於預測分類問題，且其中使用極限梯度提升之成效最佳。因此，為求本論文之預測新進玩家流失能夠達到預期，將採用決策樹、隨機森林與極限梯度提升來驗證樹狀結構之優勢以及極限梯度提升之最佳表現。

2.4 資料不平衡處理及其評估方式

在遊戲領域進行機器學習訓練時，往往會遭受到資料不平衡的影響；例如：於預測是否付費上，非付費玩家會遠多於付費玩家，導致付費玩家資料過少 [13]。於預測是否流失上，流失玩家會遠多於非流失玩家，導致非流失玩家資料過少 [15]，前述研究都採以針對資料集進行處理的方式解決資料不平衡，例如：合成少類別過抽樣技術 (Synthetic Minority Over-sampling Technique)，於少數群添加假資料，使得少數群之樣本數與多數群相等 [26]。而本論文不只預測玩家是否會流失，還需分析其原因，如在資料集中填入假資料，將會使得分析失準，無法得到有效的資訊，所以我們將採用在機器學習訓練時，放大少數群之樣本權重值，使得學習模型更加著重於少數群的資訊，如同提升方法建樹時，藉由權重值的不同，修正分類錯誤的資訊 [20]。

在評估資料不平衡資料集時，如果單純計算學習模型之 Precision、Recall 或 F - Score [27]，將導致多數群之評估結果壓過少數群之評估結果，使得最終評估失真，無法有效驗證學習模型之成效。因此，在評估不平衡資料時，Sifa 等人額外運用幾何平均數 (Geometric Mean) [28] 來評估學習模型之成效 [13]。藉由上述的概念，本論文將採用 Weighted F_{β} - Score 來評估不平衡資料，使得少數群之評估不被多數群所壓過，使用樣本間的數量權重差來計算多數群與少數群的 F_{β} - Score，希望能夠合理的評估學習模型間的表現。

参 考 文 献

- [1] P. Miller, “Gdc 2012: How valve made team fortress 2 free-to-play,” *Gamasutra. Haettu*, vol. 7, p. 2012, 2012.
- [2] Swrve, “The april 2014 new players report.” <https://www.swrve.com/resources/weblog/the-april-2014-new-players-report>, 2014.
- [3] T. Fawcett, “An introduction to roc analysis,” *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.
- [4] D. Powers, “Evaluation: From precision, recall and f-factor to roc, informedness, markedness and correlation,” *Mach. Learn. Technol.*, vol. 2, 01 2008.
- [5] C. Goutte and É. Gaussier, “A probabilistic interpretation of precision, recall and f-score, with implication for evaluation,” in *ECIR*, 2005.
- [6] Wikipedia contributors, “Free-to-play — Wikipedia, the free encyclopedia.” <https://en.wikipedia.org/w/index.php?title=Free-to-play&oldid=965292994>, 2020. [Online; accessed 10-July-2020].
- [7] E. Lee, Y. Jang, D.-M. Yoon, J. Jeon, S.-i. Yang, S.-K. Lee, D.-W. Kim, P. P. Chen, A. Guitart, P. Bertens, □. Periañez, F. Hadiji, M. Müller, Y. Joo, j. Lee, I. Hwang, and K.-J. Kim, “Game data mining competition on churn prediction and survival analysis using commercial game log data,” *IEEE Transactions on Games*, vol. 11, no. 3, pp. 215–226, 2018.
- [8] R. Flunger, A. Mladenow, and C. Strauss, “Game analytics on free to play,” in *Big Data Innovations and Applications* (M. Younas, I. Awan, and S. Benbernou, eds.), (Cham), pp. 133–141, Springer International Publishing, 2019.
- [9] B. Gregory, “Predicting customer churn: Extreme gradient boosting with temporal data,” *arXiv preprint arXiv: 1802.03396*, 2018.
- [10] M. Tamassia, W. Raffe, R. Sifa, A. Drachen, F. Zambetta, and M. Hitchens, “Predicting player churn in destiny: A hidden markov models approach to predicting player departure in a major online game,” in *2016 IEEE Conference on Computational Intelligence and Games (CIG)*, pp. 1–8, IEEE, 2016.

- [11] Á. Periañez, A. Saas, A. Guitart, and C. Magne, “Churn prediction in mobile social games: Towards a complete assessment using survival ensembles,” in *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 564–573, IEEE, 2016.
- [12] J. Runge, P. Gao, F. Garcin, and B. Faltings, “Churn prediction for high-value players in casual social games,” in *2014 IEEE conference on Computational Intelligence and Games*, pp. 1–8, IEEE, 2014.
- [13] R. Sifa, F. Hadiji, J. Runge, A. Drachen, K. Kersting, and C. Bauckhage, “Predicting purchase decisions in mobile free-to-play games,” in *Eleventh Artificial Intelligence and Interactive Digital Entertainment Conference*, 2015.
- [14] H. Xie, S. Devlin, D. Kudenko, and P. Cowling, “Predicting player disengagement and first purchase with event-frequency based data representation,” in *2015 IEEE conference on Computational Intelligence and Games*, pp. 230–237, IEEE, 2015.
- [15] S.-K. Lee, S.-J. Hong, S.-I. Yang, and H. Lee, “Predicting churn in mobile free-to-play games,” in *2016 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 1046–1048, IEEE, 2016.
- [16] F. Hadiji, R. Sifa, A. Drachen, C. Thureau, K. Kersting, and C. Bauckhage, “Predicting player churn in the wild,” in *2014 IEEE Conference on Computational Intelligence and Games*, pp. 1–8, 2014.
- [17] L. Breiman, “Bagging predictors,” *Machine learning*, vol. 24, no. 2, pp. 123–140, 1996.
- [18] L. Breiman, “Random forests,” *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [19] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and regression trees*. CRC press, 1984.
- [20] Y. Freund, R. Schapire, and N. Abe, “A short introduction to boosting,” *Journal-Japanese Society For Artificial Intelligence*, vol. 14, no. 771-780, p. 1612, 1999.
- [21] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, 2016.

- [22] T. Huang, “機器學習: Ensemble learning 之 bagging、boosting 和 adaboost.” <https://medium.com/@chih.sheng.huang821/%E6%A9%9F%E5%99%A8%E5%AD%B8%E7%BF%92-ensemble-learning%E4%B9%8Bbagging-boosting%E5%92%8Cadaboost-af031229ebc3>. [Online; accessed 10-July-2020].
- [23] A. Martínez, C. Schmuck, S. Pereverzyev Jr, C. Pirker, and M. Haltmeier, “A machine learning framework for customer purchase prediction in the non-contractual setting,” *European Journal of Operational Research*, vol. 281, no. 3, pp. 588–596, 2020.
- [24] A. Semenov, P. Romov, S. Korolev, D. Yashkov, and K. Neklyudov, “Performance of machine learning algorithms in predicting game outcome from drafts in dota 2,” in *International Conference on Analysis of Images, Social Networks and Texts*, pp. 26–37, Springer, 2016.
- [25] A. Janusz, T. Tajmayer, and M. Świechowski, “Helping ai to play hearthstone: Aaia’17 data mining challenge,” in *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pp. 121–125, IEEE, 2017.
- [26] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “Smote: synthetic minority over-sampling technique,” *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.
- [27] N. Chinchor and B. M. Sundheim, “Muc-5 evaluation metrics,” in *Fifth Message Understanding Conference (MUC-5): Proceedings of a Conference Held in Baltimore, Maryland, August 25-27, 1993*, 1993.
- [28] M. Kubat, R. Holte, and S. Matwin, “Learning when negative examples abound,” in *European Conference on Machine Learning*, pp. 146–153, Springer, 1997.