

第四章 模擬實驗設計與結果

本章將以第三章所述之模擬實驗架構，並以「Repast」(<http://repast.sourceforge.net/index.html>) 為模擬環境實作的工具，以Java 為語言撰寫，來分析比較社會性書籤網站上各種不同的使用者閱讀策略。

第一節 模擬環境塑模工具——Repast

現行多重代理人模擬開發工具有由MIT Media Lab所開發的StarLogo、Santa Fe Institute所開發的Swarm、美國西北大學所開發的NetLogo等等。而Repast是其中較為普及，且兼具功能複雜性與開發容易性者(Railsback, Lytinen, & Jackson, 2006)。

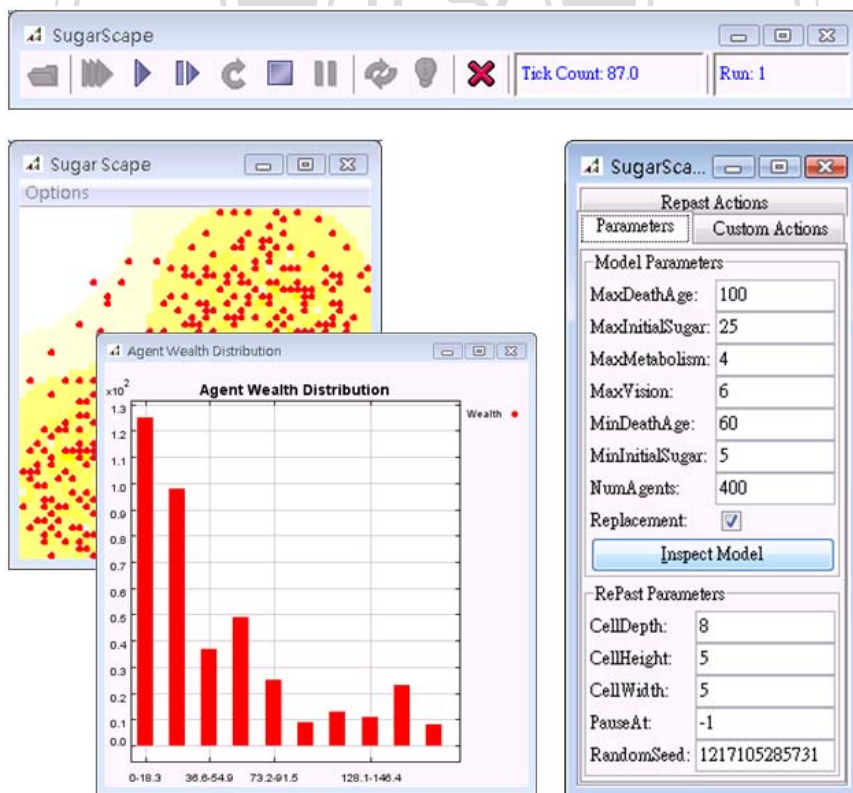


圖 4-1 Repast 運作畫面

資料來源：本研究整理

Repast (<http://repast.sourceforge.net/index.html>) 的全名為 The Recursive Porous Agent Simulation Toolkit，是一套免費、開放原始碼且跨平台的代理人模擬開發工具，由 David Sallach 等人在美國 Argonne 國家研究室 (Argonne National Laboratory) 及芝加哥大學的社會科學電腦運算研究中心 (Social Science Computing Research) 所開發出來的，現由非營利志工組織「ROAD」(Repast Organization for Architecture and Development) 負責管理與維護。最新的版本為 Repast Symphony，而本研究所使用的則是 Repast3 中支援 java 的 RepastJ，為 java 的程式庫 (library)。

Repast 擁有開發多重代理人模擬模型時所會用到的大量程式庫像是基因演算法、類神經網路、亂數統計、社會網路模型、二維模擬空間等等；它是物件導向的開發工具，並支援了包括 java、C#、Visual Basic .Net、Python 等多種程式語言。

一個典型的 Repast 模擬模型，在執行時會顯示如圖 4 - 1 的畫面。最上方的控制列，用來控制 Repast 中所內建的排程 (Schedule) 模組，其中類似收音機播放、暫停、停止等功能的圖示按鈕，讓模擬者可以控制模擬實驗的執行、停止、或一次只執行一個時間單位，上面亦會顯示目前的模擬時間。右方的變數設定區，可以讓模擬者透過程式碼來定義實驗中可供調整的變數，並以文字方塊來調整變數數值，以核取方塊來調整布林變數，以下拉選單從既定清單中選取變數。中間的幾個視窗，則依模擬者的需求，透過程式碼的撰寫來加入相關的圖表或代理人的運作空間，用來監控實驗進行中的各項數值變化。本章節中各個實驗的圖表，即是由 Repast 所提供的圖表模組所產生。

Repast 中的各個元件模組如圖 4 - 1 所示。Engine Module 為最核心的模組，負責控制模擬活動的進行，包括實驗的開始、暫停、時間控制、代理人及工作任務管理等等。其中又包含了 Engine Controller、Scheduler、Action 和 Agent

等幾個元件。Logging Module 負責的是紀錄模擬結果；Interactive Run Module 通常扮演了 Engine Module 及模擬實驗者之間的橋樑，模擬者可以透過這個互動模組所提供的視覺介面，與 Engine Module 來互動，操控模擬的進行，觀察模擬的結果。其中又包含了 Environment Visualizations、Probes、Graphs、Reports 等元件；相較於 Interactive Run Module，Batch Run 的模擬中使用者在一切變數設定好、模擬開始進行後，並不會再和模擬系統之間有所互動；最後，Adaptive Behaviors Module 可以讓模擬者使用一些調適性演算的功能，例如基因演算法、類神經網絡及其他人工智慧工具等等；而 Domains Module 則提供一些特殊領域的模擬時所需要的工具，例如社會網路、地理資訊系統、系統動力學、賽局理論等等。

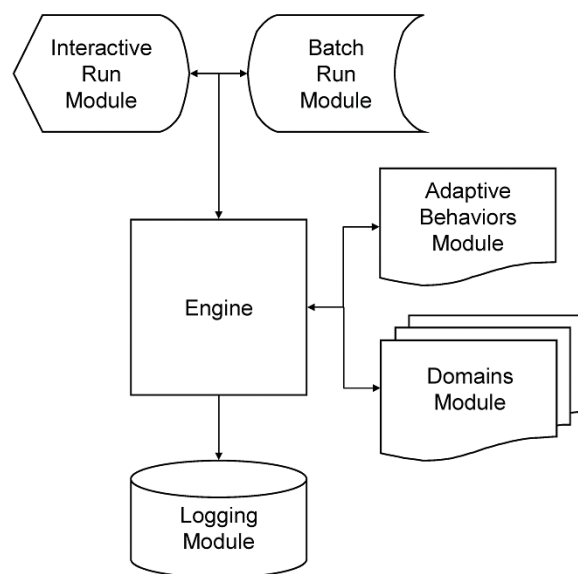


圖 4-2 Repast 元件模組

資料來源：NORTH, COLLIER, & VOS (2006)

第二節 實驗目的與實驗設計

社會性書籤網站的推薦系統，是藉由匯整使用者在閱讀內容時所同時給予的評價資訊，來對所有平台上的內容進行整理、排序，以讓好的內容能被篩選出來，給使用者更好的閱讀經驗。在第三章中曾經提到，使用者的閱讀行為同時具有消

費內容與協助篩選工作的雙重意義。因此，如何同時兼顧工作的分配與使用者的閱讀滿意度，是使用者閱讀行為設計時的重要課題。本研究的實驗設計，藉由調整使用者的閱讀行為，觀察各項指標的變化，來尋找使用者的閱讀行為策略的設計方向。

在 Lerman(2007a)的研究中，文章的曝光度受到在熱門排行中的序位、在時序排行 (upcoming stories queue) 中的序位、以及使用者的社交關係等因素的影響，越熱門、越新的文章曝光率越高。同時，在序位上，首頁的曝光率最高，有最大的機率被使用者看見，並隨著頁次往後而曝光率遞減。本研究不考慮社交關係，在閱讀策略的設計上，同時參酌實際網站運作情形，讓使用者同時從熱門排行清單及時序排行清單中不同的區段，選擇文章來閱讀。下圖以熱門排行清單為例，說明模擬實驗中的閱讀策略設計：



圖 4-3 文章閱讀策略設計

資料來源：本研究整理

每個清單均被切成四個段，每段的長度為實驗中可調整的變數，熱門排行清單中第一段的長度為 VL1，第二段為 VL2，第三段為 VL3，最後一段則為清單中扣除前三段後，至清單尾之所有文章；同樣的，時序排行清單中的前三段長度分別為 TL1~TL3。每個使用者的 agent，在實驗中每個時間單位，會分別從兩個清單中的每個段落，選擇數篇文章來閱讀，而所選取的數量亦為可調整的變數。在熱門排行清單中，每段所選取的文章數分別為 VLc1~VLc4，而時序排行清單中每段所選取的文章數則為 TLc1~TLc4。依照不同的變數組合，可以設定每個段

落中的文章被選取、閱讀的機率；但是同一段中每篇文章被選取閱讀的機率是一樣的。合理的變數設定應為，第一段的文章被讀取的機率最高，象徵首頁的頭條能獲得最多的曝光機會。依序往後，越後面的文章曝光率越低。實驗中的相關變數設置，如表 4 - 1 所示。

表 4-1 模擬實驗變數表

變數名稱	意義	變數名稱	意義
NumUsers	使用者 agent 的數量	Runs	進行時間單位
OriginalArticle	初始文章篇數	aging	過時文章清除基數
VL1	熱門排行段 1 長度	TL1	時序排行段 1 長度
VL2	熱門排行段 2 長度	TL2	時序排行段 2 長度
VL3	熱門排行段 3 長度	TL3	時序排行段 3 長度
VLc1	熱門排行段 1 選擇篇數	TLc1	時序排行段 1 選擇篇數
VLc2	熱門排行段 2 選擇篇數	TLc2	時序排行段 2 選擇篇數
VLc3	熱門排行段 3 選擇篇數	TLc3	時序排行段 3 選擇篇數
VLc4	熱門排行段 4 選擇篇數	TLc4	時序排行段 4 選擇篇數

資料來源：本研究整理

本研究在模擬實驗設計上，有以下幾個假設：

1. **使用者的理性推薦**：於本研究的實驗設計中，使用者針對特定文章推薦與否，完全由文章的品質來決定，當文章的品質設定為 5 時，即有五成的使用者會推薦該篇文章。在實際的系統運作上，使用者可能有許多非理性、非客觀的推薦行為，需要透過演算法或誘因設計來排除；而本研究則著重在使用者理性行為下，閱讀策略的設計，因此演算法的部分也相對簡單，僅採用使用者針對每篇文章一人一票，票票等值的方式。
2. **使用者之間無社交關係**：本實驗設計中，使用者僅從熱門排行清單及時序排行清單兩個清單中來閱讀文章，未有社交關係，也沒有透過社交關係來閱讀文章的行為。
3. **所有使用者的參與程度相同**：本實驗設計中，所有使用者對網站平台的投入程度是相同的，每個時間單位所閱讀的文章數相同，並且在閱讀文

章的同時均參與推薦。在實際的網站運作中，使用者會有不同程度的參與，有的使用者投入大部分的活動，有的使用者僅僅有最小的參與而無貢獻，例如僅閱讀而不推薦（即 free rider）。

為了充份了解模擬環境中的系統表現，本實驗亦設計下列的觀察指標：

1. **使用者閱讀滿意度**：使用者閱讀滿意度的計算方式為，將每個時間單位中，該名使用者所有選取閱讀的文章品質加總後，取其平均，即為該時間單位該使用者的閱讀滿意度。其中比較特別的是，如果使用者在平台中看到的文章不常更新，每次所讀到的文章有大部分是先前曾經讀過的，則雖然該文章仍屬佳作，但仍無法讓使用者在重覆閱讀中獲得滿足。因此，在本研究的實驗設計裡，如果該篇文章是使用者在先前曾經讀過的，則縱然該文章的品質很高，但重覆閱讀時依然只以品質 1 來計算，以表達使用者再次閱讀同一篇文章時所得到的滿意度極低。因此，必需有較高的文章週轉時間，讓文章有一定的更新頻率，才能將使用者滿意度提高。但是這是一項落後指標，必需參照其他的指標才能分析出背後造成使用者滿意度波動的原因。
2. **文章閱讀重覆率**：前面提到，當使用者重覆讀取到先前曾經讀過的文章時，其滿意度會僅用 1 來計算，因此文章閱讀的重覆率對使用者滿意度的影響甚大。文章重覆率越高，使用者的滿意度會越低。文章閱讀重覆率的計算方式為，將該時間單位使用者所選取閱讀的文章中，重覆閱讀的篇數除以該時間單位總閱讀的篇數，再轉換成百分比。
3. **首頁文章平均品質**：本研究將熱門排行清單中的前二十篇文章定義為首頁，首頁是群體智慧進行文章篩選的成果展現。群體的運作是否能有效的把好的文章篩選出來，可以由這個指標來觀察。其計算方式為，將

熱門排行清單中前二十篇文章的品質加以平均。

4. **首頁文章週轉時間**：首頁文章週轉時間指的是，被篩選至首頁的文章，平均會在首頁停留多少時間。如首頁文章的週轉時間太低，縱然首頁的文章平均品質很高，但是表示僅有少數文章一直停留在首頁上，其他高品質的文章沒有被賞識的機會。同時也表示，系統中的工作分配不佳，導致清單後方的文章無法得到有效的篩選。當首頁文章的週轉時間低時，使用者容易讀到重覆的文章，因此滿意度會降低。
5. **工作完成度**：每篇在平台上的文章，可以視為是一個待完成的工作，使用者一方面閱讀文章，一方面也是在鑑賞、評薦的過程中，來完成工作。一篇好的文章還必需要有一定的完成度，才足以被發掘出來，甚至排上首頁。太低的工作完成度，或工作完成度的不平均，都會使得好文章可能被埋沒。工作完成度的計算方式為，閱讀過該篇文章的人數，除以平台上所有使用者的數量，再轉換成百分比。
6. **頭條議題熱門度**：在第一章中曾經提到，社會性書籤網站同時具有傳統媒體的性質。在新聞媒體中，藉由頭條的安排，可以做議題的塑造，創造閱聽人、社會大眾之間共通的關注焦點。前述文章週轉時間短，雖然可以加速平台上文章的新陳代謝，但相對的首頁上每篇文章停留的時間都不長，可能較無法引起大多數使用者的共鳴與迴響。因此，從媒體運作的角度來看，頭條議題的熱門度是必需考量的指標。本研究的模擬實驗假設，一篇文章如果受到越多讀者的閱讀，會越容易形成共通的議題，因此以首頁（熱門排行清單中前二十篇文章）的平均工作完成度，做為衡量頭條議題熱門度的依據。

第三節 模擬程式架構

本研究的實驗設計利用 Repast 發展模擬環境，總共包含四個 java 的 class

檔：

1. **SMSModel.class** 這個 Class 檔繼承了 Repast 中的 SimModelImpl，是整個模擬實驗的核心，控制了模擬的進行。其中包含了 SimModelImpl 中必備的 setup、begin、buildModel、buildSchedule、buildGraphic 等 method，main 也在這個 class 之中。實驗中的變數設定、代理人的產生、時程控制，都由這個 class 來負責。此外，所有的使用者及文章，在產生後，都放置在這個 class 下的 ArrayList 中。
2. **UserAgent** 這個 class 代表的是模擬代理人，模擬實驗中所有的使用者代理人都由這個 class 所產生。這個 class 中除了紀錄使用者相關的屬性，更定義了使用者的閱讀行為。在下面實驗中談到閱讀策略的控制，都是透過在這個 class 中的 readerStep 這個 method 來定義。
3. **Article** 這個 class 代表的是文章，模擬實驗中所有的文章都由這個 class 所產生，其中包含了文章的各種屬性，如品質、所得到的推薦數、工作完成度等等。文章的推薦功能，也定義在這個 class 中。
4. **Graph** 本研究的模擬實驗產生大量的結果圖表，如折線圖、直方圖、散佈圖等等。所有圖表的初始化、相關設定及繪製動作，均定義在這個 class 當中。

模擬實驗的進行流程如下：

1. 執行 SBSModel 後，開始執行模擬實驗。首先，透過 SBSModel 完成初始值的設定，並且由 userAgent 及 Article 來產生所有初始的使用者及文章。同時，所有圖表的初始設定也在 Graph 產生後設定完成。
2. 模擬者按下開始按鈕後，Repast 的 Scheduler 模組開始模擬時間的跳動 (tick)。在每一個 tick 當中，每個 userAgent 的 readerStep 會被

執行一次，依其定義的閱讀策略，完成文章的閱讀。同時，每一個時間單位也會有若干篇新的文章產生。另一方面，每個時間單位均由 SBSModel 中的 report 這個 method，將該時間點的各個觀察指標值顯示出來。

3. 部分折線圖，經程式的設定，在 Scheduler 中註冊為每隔十個時間單位紀錄一次最近的值，值的抓取及繪製是由 Graph 中的 method 所定義。
4. 本研究中所有的實驗均進行 150 個時間單位。在經過 150 個時間單位後，模擬停止，並且由 Graph 將所有的散佈圖繪製出來，完成該次的實驗。

第四節 實驗與結果分析

【實驗 1】 熱門排行清單 vs. 時序排行清單 篩選效果比較

實驗 1 中，希望比較在閱讀策略設計時，於熱門排行清單與時序排行清單中選擇時所產生的篩選效果。因此，在本組實驗中調配三種變數的組合，以不同的閱讀策略分別進行模擬：實驗 1-2 中使用者完全透過時序排行清單來閱讀文章，表示只閱讀最新、較新的文章。實驗 1-3 完全透過熱門排行清單來閱讀，表示只閱讀最熱門或較熱門的文章。而實驗 1-1 則是從兩個清單中分別閱讀十篇。閱讀策略相關的變數設定，如表 4 - 2 所示。

在每次的實驗中，均安排 50 個使用者的 agent、有 300 篇的初始文章在時間點 0 的時候產生、並且每輪進行 150 個時間單位。清單的段落切割上，第一段為 20 篇，第二段為 70 篇，第三段為 100 篇。每次實驗會有不同的閱讀策略設計，但每輪每人均維持閱讀 20 篇文章，將 20 篇文章分配至熱門排列清單與時序排列清單的各個段落中，讓第一段被閱讀的機率最高，第二、三段依次遞減，最後一段的機率微乎其微，模擬使用者顯少點選至後面的頁面瀏覽。

表 4-2 實驗 1 變數設定

實驗 1-1 (最新、最熱門各讀一半)				實驗 1-2 (只讀最新)				實驗 1-3 (只讀最熱門)			
變數	數值	變數	數值	變數	數值	變數	數值	變數	數值	變數	數值
VL1	20	TL1	20	VL1	20	TL1	20	VL1	20	TL1	20
VL2	70	TL2	70	VL2	70	TL2	70	VL2	70	TL2	70
VL3	100	TL3	100	VL3	100	TL3	100	VL3	100	TL3	100
VLc1	3	TLc1	3	VLc1	0	TLc1	6	VLc1	6	TLc1	0
VLc2	3	TLc2	3	VLc2	0	TLc2	6	VLc2	6	TLc2	0
VLc3	2	TLc3	2	VLc3	0	TLc3	4	VLc3	4	TLc3	0
VLc4	2	TLc4	2	VLc4	0	TLc4	4	VLc4	4	TLc4	0

共同設定：NumUsers=50 OriginalArticle=300 Runs=150

資料來源：本研究整理

每組變數設定，均進行十次實驗，並取每次實驗於結束時間（t=150）時六個指標值，加以平均，如下表所示。每個指標在三組實驗中的最佳值以深色背景標示、次佳值以淺色背景標示：

表 4-3 實驗 1 觀察指標模擬結果

觀察指標	實驗 1-1 (各讀一半)	實驗 1-2 (只讀最新)	實驗 1-3 (只讀最熱門)
平均閱讀滿意度	2.74	4.08	1.18
平均閱讀重覆率	43.50 %	8.82 %	81.22 %
首頁文章平均週轉時間	57.6	31.8	66.8
首頁文章平均品質	8.77	8.21	8.38
頭條議題熱門度	99.96%	52.92%	100%
平均工作完成度	15.22 %	23.13 %	6.83 %

資料來源：本研究整理

從本次實驗結果的簡表中可以發現，實驗 1-2 在閱讀滿意度、閱讀重覆率、週轉時間、完成度等幾項指標中，表現均是三組當中最優的，但是在頭條議題熱門度方面，卻相當不理想。頭條議題熱門度最佳的為完全依熱門排行清單閱讀的實驗 1-3，但是實驗 1-3 的其他指標均敬陪末座；反而是實驗 1-1，綜合熱門及時序兩種閱讀策略，在頭條議題熱門度上和實驗 1-3 僅有極些微的差距，但是其

他指標大多遠勝過實驗 1-3。

更進一步從圖表來分析，本研究的模擬實驗中同時產生四種不同的圖表：《使用者滿意度暨文章平均品質折線圖》（如圖 4 - 4）顯示的是在模擬進行中每個時間點的閱讀滿意度與首頁文章品質。最上面三條折線分別代表的是首頁的前 10 篇、前 20 篇及前 30 篇文章的平均品質（本研究中的指標，是將前 20 篇設定為首頁），中間黑色實心三角形折線代表的是所有平台上文章的平均品質，而最下面三條代表的是使用者閱讀的滿意度，紅色實心圓點折線為所有使用者中滿意度最高者，青色空心方形折線為平均滿意度，而藍色打叉折線為滿意度最低者。

《文章產生時間 vs. 工作完成度分佈圖》（如圖 4 - 5）呈現的是文章產生時間與工作完成度之間的關係，橫軸代表的是文章產生的時間，而縱軸則是該篇文章的工作完成度，將平台上所有的文章依照這兩個數值，標示在相對應的座標位置中，同時以顏色表示不同的文章品質，這個圖主要想觀察在不同時間所產生的文章，是否能夠在平台上獲得同樣的曝光機會。《熱門排行清單序位 vs. 文工作完成度分佈圖》（如圖 4 - 6）同樣是反應文章的工作完成度，也以顏色來標示文章品質，但是該圖形以每篇文章在熱門排行清單的排序做為橫軸，藉以反應出經過書籤網站的群體智慧運作後，對文章篩選的效果，圖中呈現的是熱門排序清單中的前 1000 篇文章。而《文章產生時間 vs. 熱門排行清單序位分佈圖》（如圖 4 - 7）則呈現出熱門排行清單中的文章，分別是在哪些時間單位所產生，其趨勢為何，其縱軸是文章位於熱門排行清單中的序位取前 1000 篇，縱軸是該文章所產生的時間。

由圖 4 - 4 可以看到，熱門排行清單中不論是前 10 篇、前 20 篇或前 30 篇，均高於平均品質，顯示經由推薦機制所產生的熱門排行清單，確實能將品質較高的文章篩選出來。但是在平均滿意度的部分（青色空心正方形折線）卻有隨時間些微下降的趨勢。

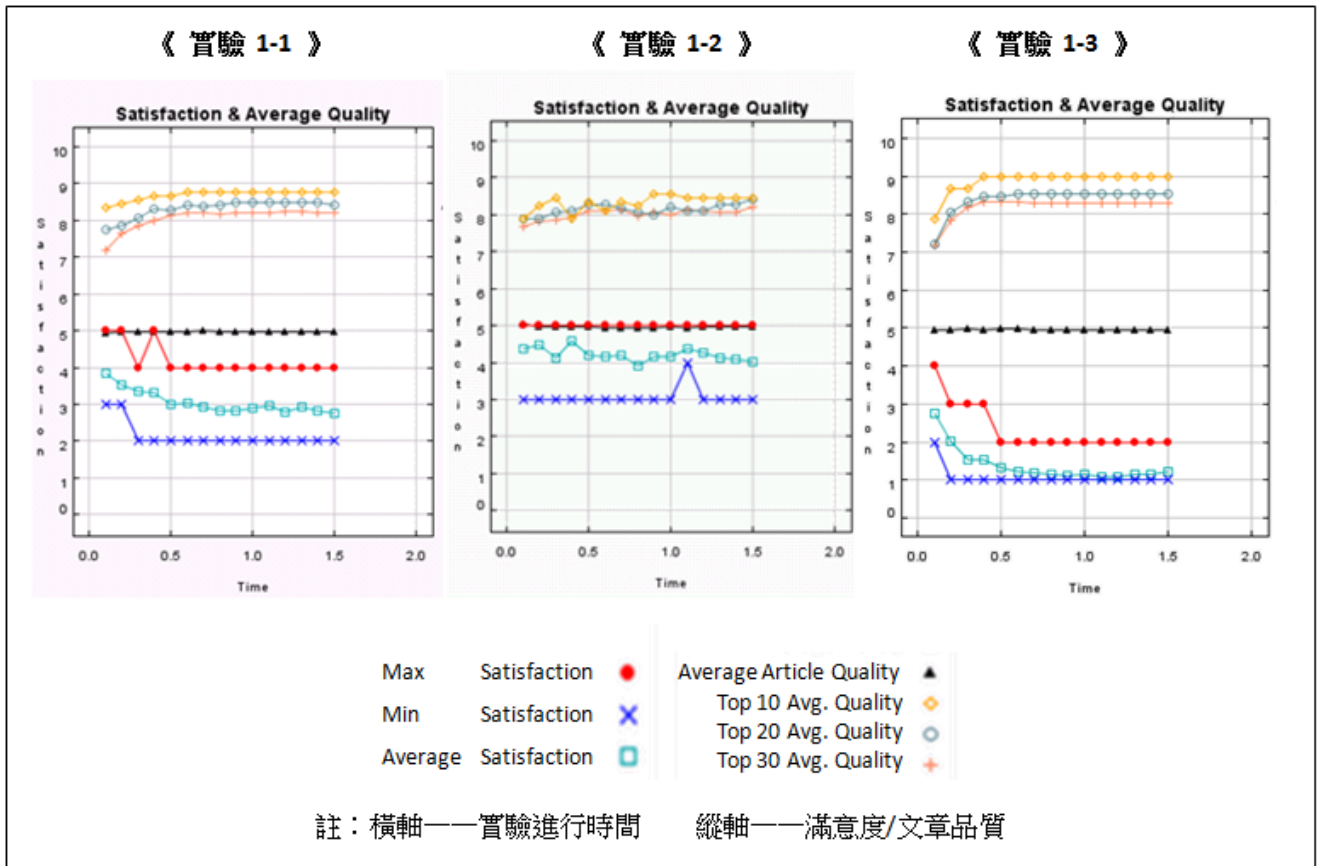


圖 4-4 實驗 1 使用者滿意度暨文章平均品質折線圖
資料來源：本研究整理

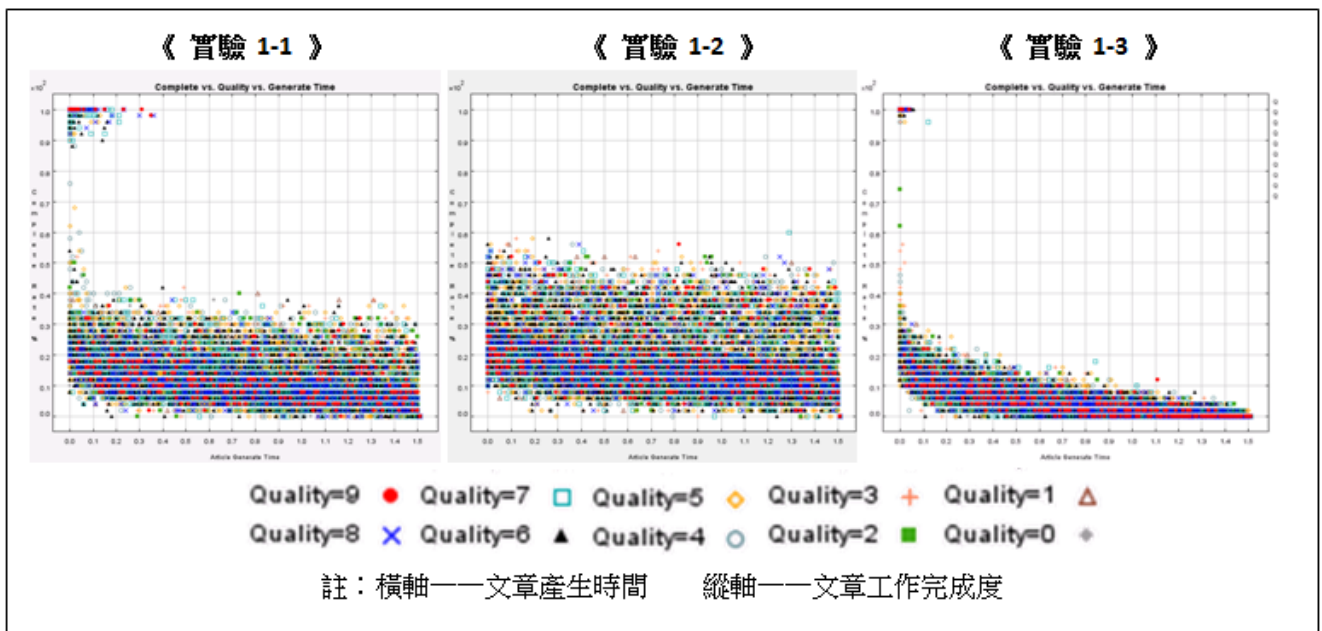


圖 4-5 實驗 1 文章產生時間 vs.工作完成度分佈圖
資料來源：本研究整理

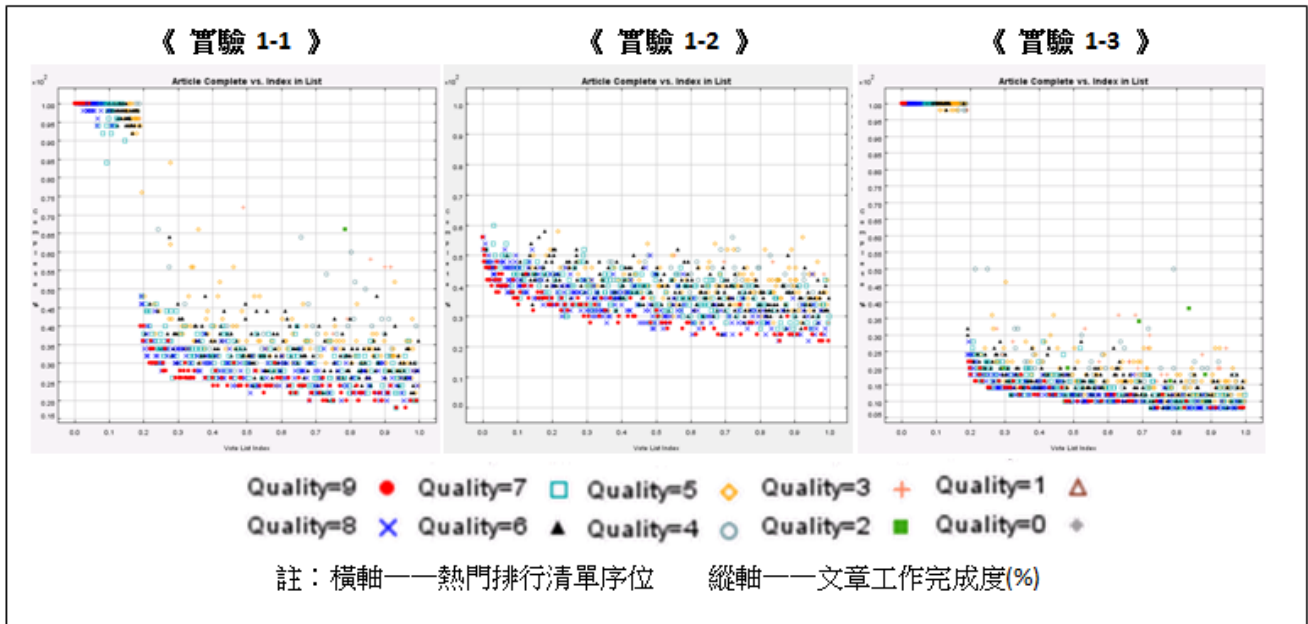


圖 4-6 實驗 1 熱門排行清單序位 vs.文工作完成度分佈圖
 資料來源：本研究整理

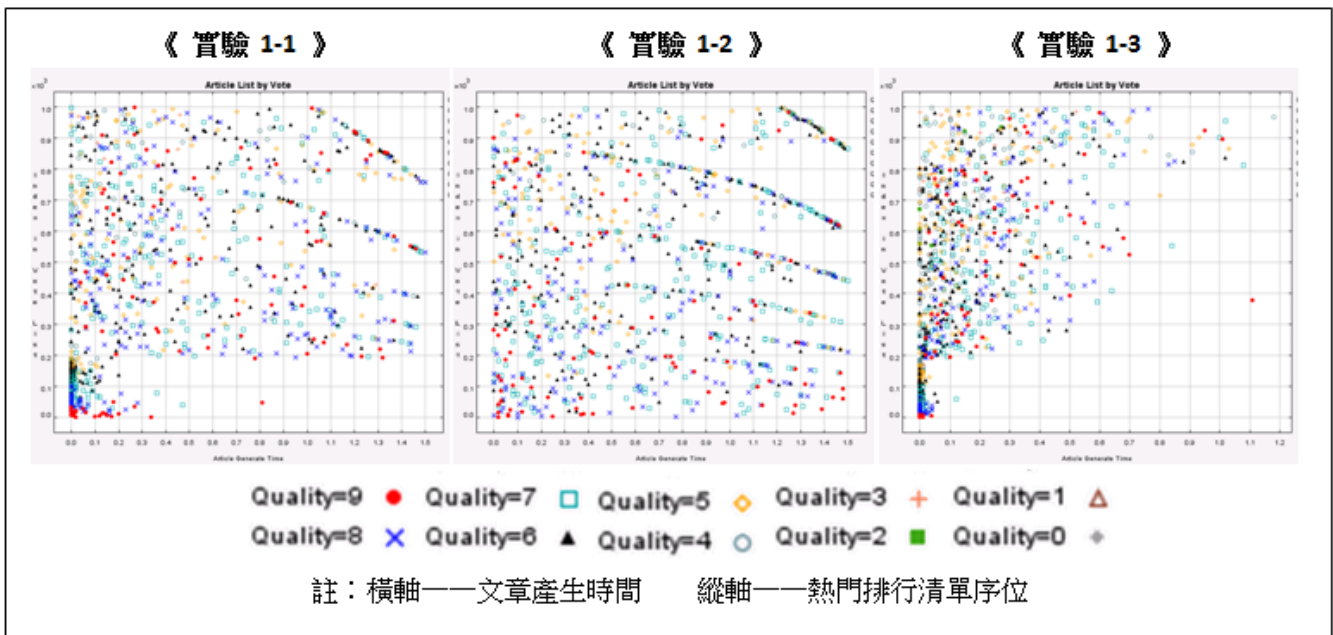


圖 4-7 實驗 1 文章產生時間 vs. 熱門排行清單序位分佈圖
 資料來源：本研究整理

觀察圖 4 - 5 可以發現，實驗 1-2 與實驗 1-3 的表現剛好是兩個極端，實驗 1-2 僅透過時序排行清單閱讀，各個時間點所產生的文章其工作完成度的分佈很平均，但是並沒有工作完成度特別突出者，這也使得實驗 1-2 的頭條議題熱門度偏低；而在實驗 1-3 中僅透過熱門排行清單閱讀時，越晚（越近）產生的文章完成度越低，但是有個很特別的現象：圖的左上方有一個文章的叢集，代表這一區極少數的文章所獲得的工作完成度特別突出，其餘的文章工作完成度都偏低，且這些被高度閱讀、完成工作的文章，高度集中在實驗剛開始的時間點。

圖 4 - 6 更清楚的呈現了實驗 1-3 在熱門排行清單前端的完成度特高（圖面左上方），而其餘文章的完成度則極低。反應在圖 4 - 7 中，可以發現實驗 1-3 中熱門排行清單前端兩百篇左右的文章，均集中在前五個時間單位所產生，顯示了一開始產生的少數文章一直盤踞在首頁，後方的文章得不到充份的曝光機會，這也充份的解釋了為何實驗 1-3 的頭條議題熱門度極高，但週轉時間長、閱讀重覆率極高，導致閱讀滿意度低落。

實驗 1-2 在圖 4 - 6 中的完成度雖高，但比較看不出篩選的效果。相較之下，實驗 1-3 在圖 4 - 6 中左上角的部分，可以很明顯的看到文章依據品質由高到低排列，依次為紅、藍、青、黑、黃。在圖 4 - 7 中也看到同樣的篩選效果，實驗 1-3 的文章排列則比較雜亂。實驗 1-1 因為融合了兩種不同的閱讀策略，所以在指標的表現上大部分居中，在圖表的呈現上也同時擁有雙方的特色。不過首頁的平均文章品質是三組實驗當中最好的。

綜合來說，實驗 1-2 雖然使用者的滿意度是最高的，但是因為使用者完全依時序來閱讀文章，推薦的功能完全沒有在閱讀策略上發揮作用，這也使得其議題熱門度遠低於其他兩組實驗，首頁文章的平均品質也是三組中最低的，並不符合社會性書籤網站藉由群體智慧來對文章進行篩選排序，並產生熱門議題的媒體功能。實驗 1-3 的表現和實驗 1-2 正好相反，有高的首頁品質和議題熱門度，但新

產生的文章完全沒有脫穎而出的機會。相較之下，融合兩者精神的實驗 1-1，會是比較好的方案。

由這個實驗可以分別展現出閱讀熱門文章和閱讀最新文章兩種閱讀策略對於平台運作、文章篩選結果的影響。讓使用者閱讀經篩選後、較熱門的文章，有助於達到群體智慧中的正回饋，讓較佳的文章能脫穎而出，且創造出熱門議題；而若使用者能夠即時的讀到最新的文章，則能加強工作分配的平均性，提升工作完成度，增加文章更新率。理想的閱讀策略設計，應該引導使用者同時閱讀最新及最熱門的文章。這也呼應了第三章中所探討的群體智慧產生要素，必需兼具正回饋與獨立性。熱門排行清單的閱讀是正回饋的安排，讓使用者直接取用前人篩選後的成果，跟著前人的腳步走，而時序排列清單的閱讀則讓使用者能夠發揮自我的判斷。

不過實驗 1-1 中工作分配不平均、集中於少數舊文章，導致重覆閱讀率大於四成（亦即每次閱讀的文章，十篇中有四篇是曾經看過的），週轉時間長達 57.6（在 150 個時間單位的閱讀行為當中，平均每等待將近 60 個時間單位才有一篇新文章出現）。

【實驗 2】 過時文章退場處理之效果

為解決上述重覆率過高、週轉時間太長的問題，實驗 2 嘗試導入文章「老化」（aging）、退場的處理，亦即每篇文章只要在平台上經過一段特定時間，即從熱門排行清單中清除，讓新產生的文章有機會被使用者閱讀到。本實驗在閱讀策略的各項變數設計，比照實驗 1-1，但是實驗 2-1 到 2-3 採用不同的老化時間，分別為 25、50 及 75 個時間單位，代表每篇文章一旦產生了，經過 25、50 及 75 個時間後，就從熱門排行清單中清除。實驗結果，將與實驗 1-1（也就是在相同的閱讀策略下，退場時間為 0 者）來比較。

表 4-4 實驗 2 變數設定

共同變數 (閱讀策略相關)				aging
變數	數值	變數	數值	
VL1	20	TL1	20	實驗 2-1 : 25
VL2	70	TL2	70	實驗 2-2 : 50
VL3	100	TL3	100	實驗 2-3 : 70
VLc1	3	TLC1	3	其他共通變數
VLc2	3	TLC2	3	NumUsers=50
VLc3	2	TLC3	2	OriginalArticle=300
VLc4	2	TLC4	2	Runs=150

資料來源：本研究整理

同樣將每組變數設定進行十次實驗，並取每次實驗於結束時間 (t=150) 時六個指標值，加以平均，如表 4-5 實驗 2 觀察指標模擬結果所示。每個指標依四組實驗各別表現，以背景的深淺表示。表現最佳者為最深色的背景，次佳者背景為次深，最差者背景為空白。

表 4-5 實驗 2 觀察指標模擬結果

觀察指標	實驗 2-1 (Aging=25)	實驗 2-2 (Aging=50)	實驗 2-3 (Aging=75)	實驗 1-1 (Aging=0)
平均閱讀滿意度	3.80	3.48	3.24	2.74
平均閱讀重覆率	26.70 %	31.66 %	35.84 %	43.51 %
首頁文章平均週轉時間	7.1	14.0	21.2	57.6
首頁文章平均品質	7.92	8.2	8.3	8.77
頭條議題熱門度	80.17 %	92.78 %	97.02 %	99.96 %
平均工作完成度	12.66 %	12.90 %	15.56 %	15.22 %

資料來源：本研究整理

由表 4-5 可以發現，在實驗 1 中所出現，閱讀重覆率太高、週轉時間太長的問題，在經過文章的過時退場處理後，的確得到了改善，其改善的程度，以 aging 的時間最短的實驗 2-1 為最。但是在頭條議題熱門度、首頁文章平均品質、工作完成度等幾項指標的表現，則恰好相反，加入退場處理後各項表現均較差，其中實驗 2-1 中，當所有文章最多只能在平台上停留 25 個時間單位之下，頭條議題熱門度下降了將近兩成。

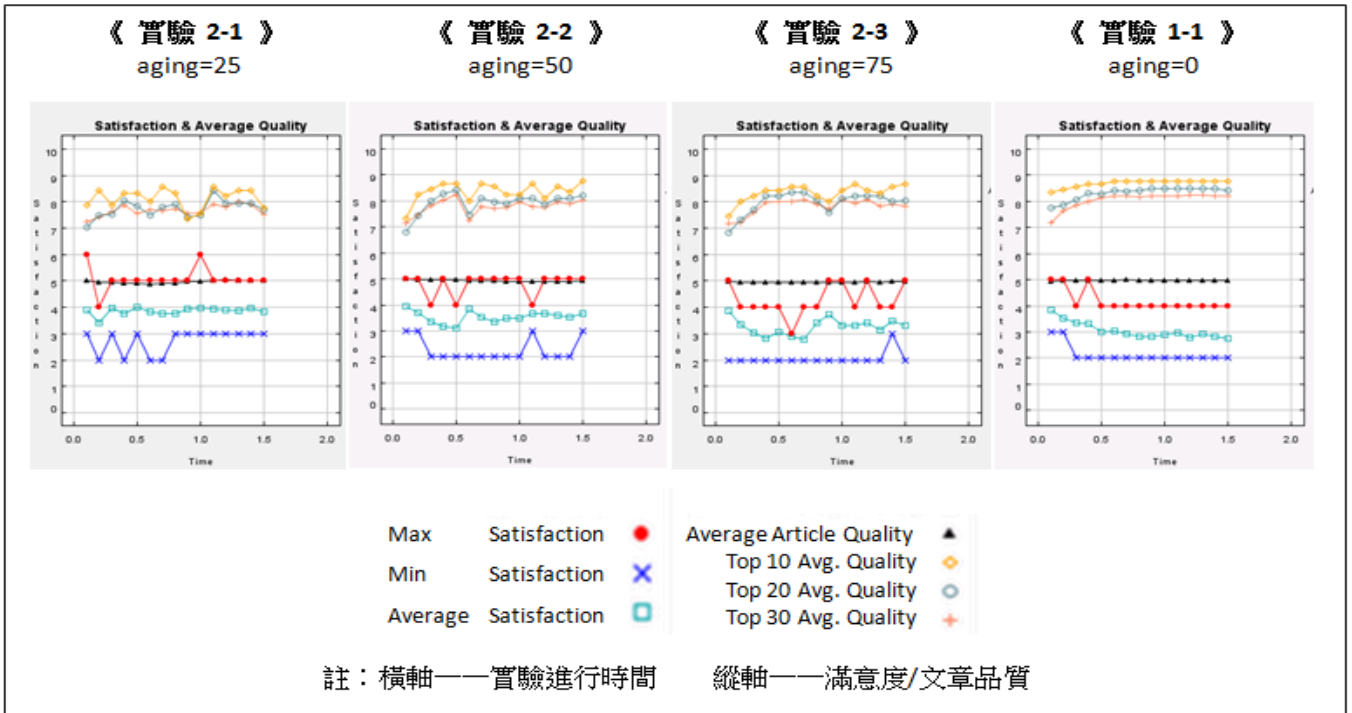


圖 4-8 實驗 2 使用者滿意度暨文章平均品質折線圖
資料來源：本研究整理

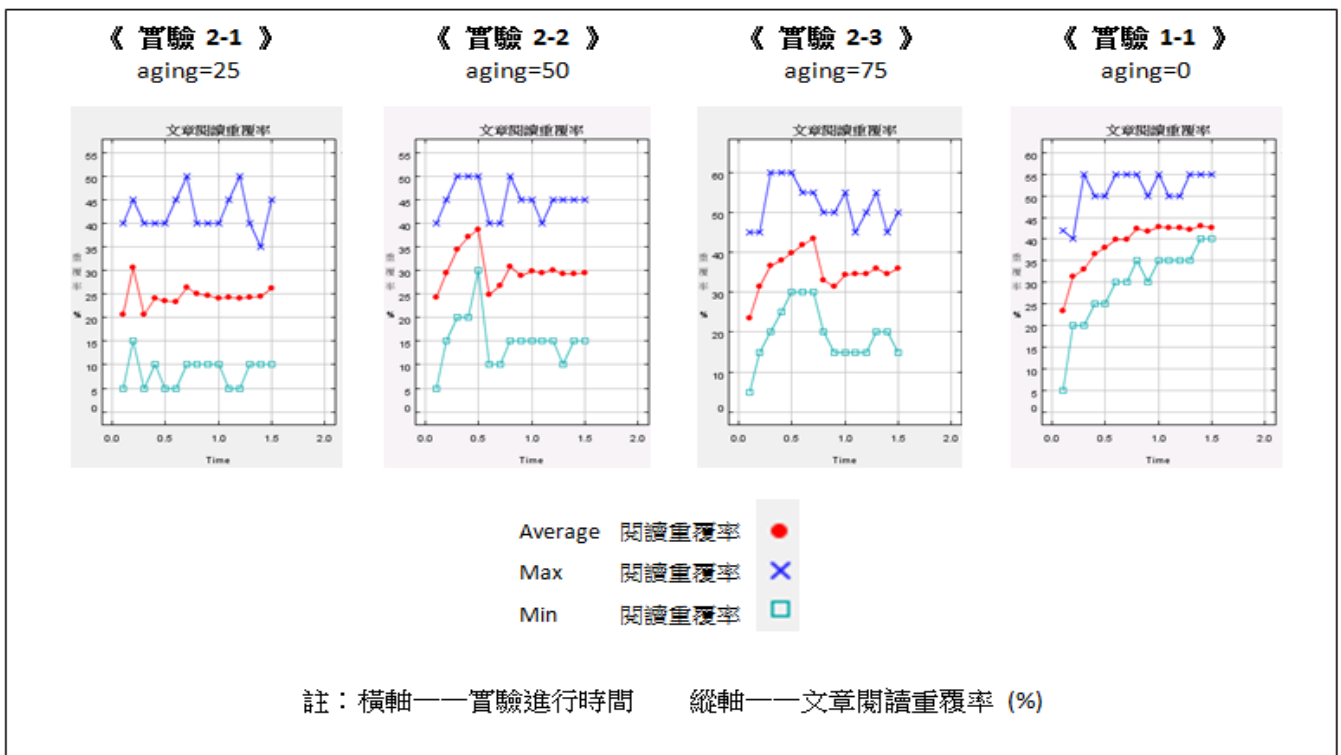


圖 4-9 實驗 2 文章閱讀重覆率折線圖
資料來源：本研究整理

同樣的由幾個圖表來詳細檢視各個實驗的模擬系統運作狀況：在圖 4 - 8 中顯示，各個實驗中的平均滿意度（青色方形空心）折線，相對於實驗 1-1 的一路下滑，實驗 2-1 至 2-3 在經過所設定的 aging 時間（25、50 及 75）後，均止跌拉高，並在些微的震盪中逐漸達到平衡。這也可以從圖 4 - 9 中看出端倪：實驗 1-1 中的閱讀重覆率隨著時間的經過，不斷的攀升，但實驗 2-1 至 2-3 則是在經過 aging 時間點後，因為部分文章的退場，使用者讀到重覆文章的機會降低了，使得閱讀重覆率陡降。正因為閱讀重覆率下降了，所以滿意度得以提升。但是相對的，圖 4 - 8 中在經過 aging 時間點時，則因為時間點 0 時所產生的高品質文章退場了，導致各實驗中的首頁文章品質均突降，不過隨後亦在震盪中逐漸回復。

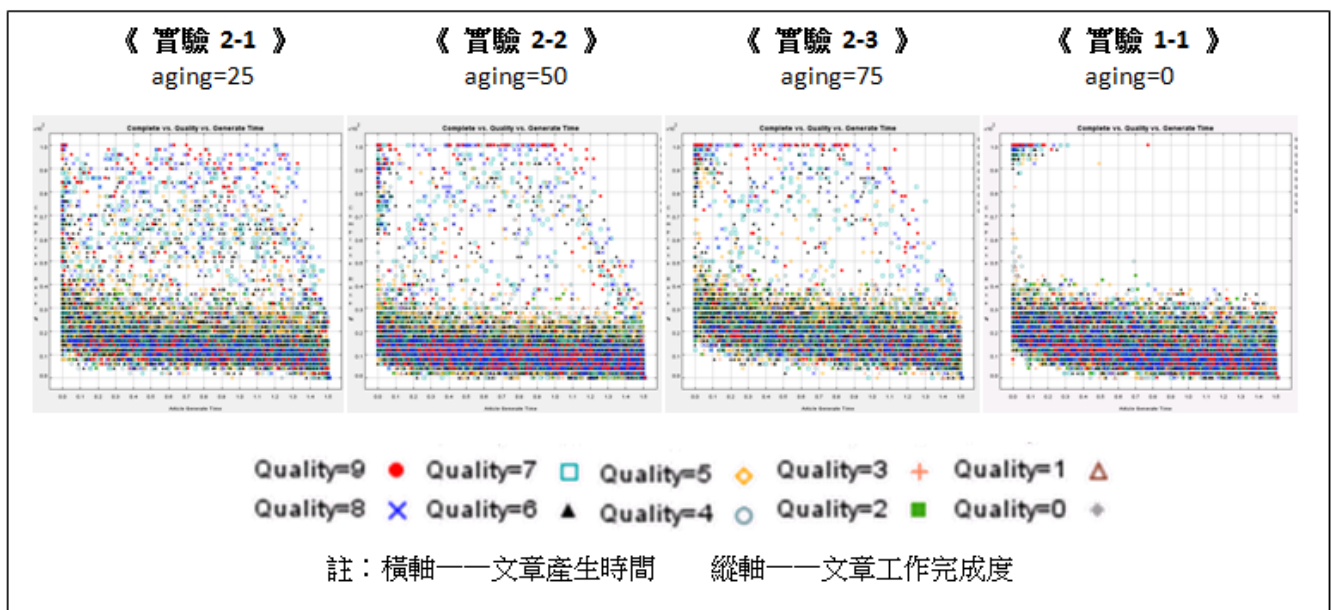


圖 4 - 10 實驗 2 文章產生時間 vs.工作完成度分佈圖

資料來源：本研究整理

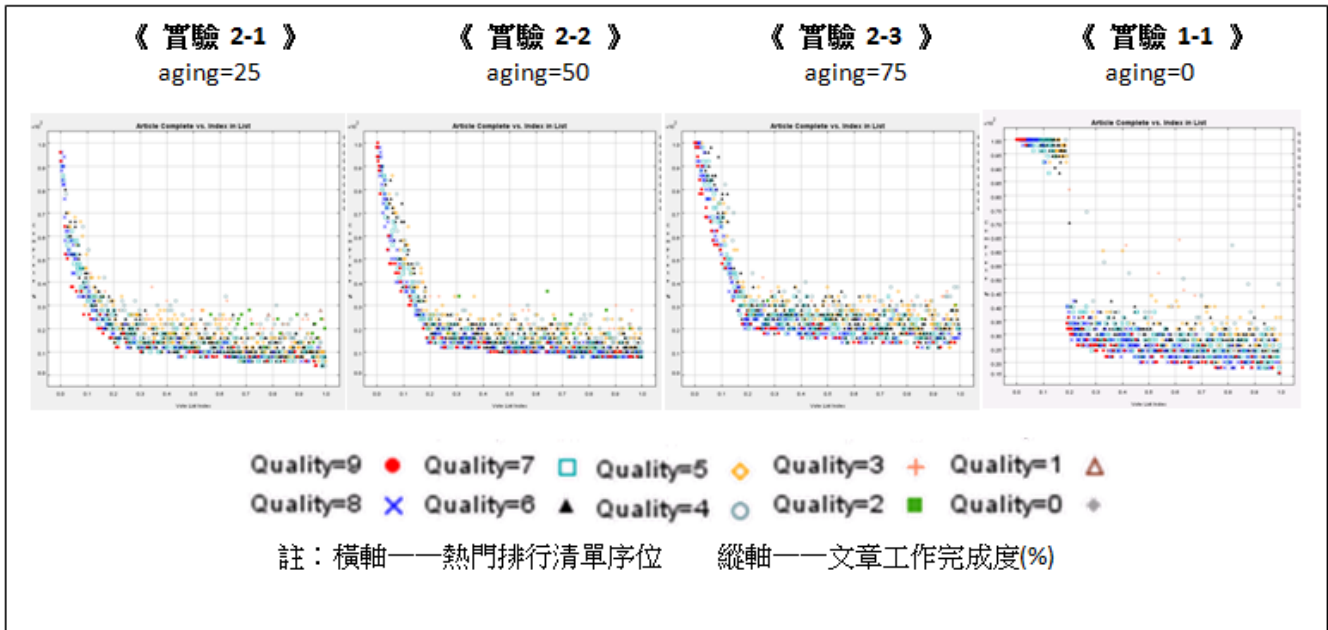


圖 4-11 實驗 2 熱門排行清單序位 vs. 文工作完成度分佈圖

資料來源：本研究整理

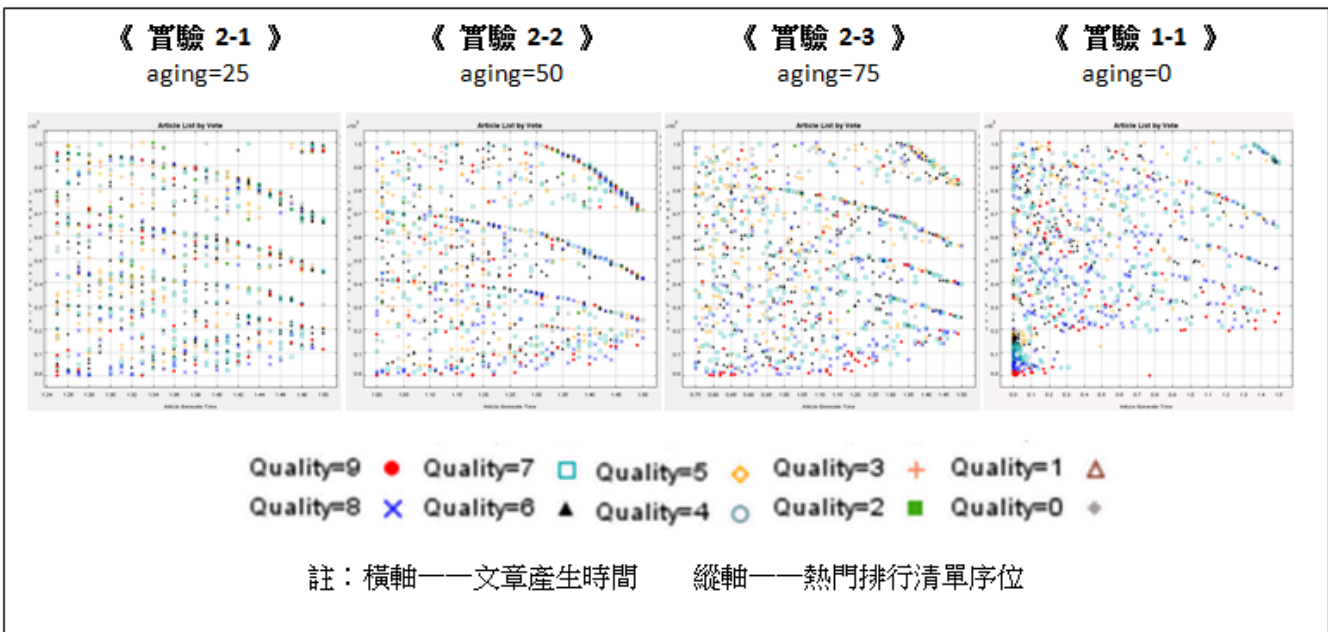


圖 4-12 實驗 1 文章產生時間 vs. 熱門排行清單序位分佈圖

資料來源：本研究整理

圖 4 - 10 中可以觀察到，原來在實驗 1-1 中左上方群集現象，在實驗 2-1 至 2-3 中均獲得改善，而且 aging 的時間越短，工作的分配越平均。不過實驗 2-2 與 2-3 仍保有實驗 1-1 中的特性，橫軸每個時間點所產生的文章，均有小部分品質不錯者（紅色或藍色）得到最高 100%的工作完成度，顯示每個時間點所產生的文章均有機會獲得青睞，這也反應在圖 4 - 12 中，頭條位置包含了各個時間點所產生的文章，亦即各個時間點所產生的文章均有機會被推舉上頭條，顯示較好、較平均的篩選效果。同樣的，在圖 4 - 11 中也可以看到原先在實驗 1-1 中工作分配極度極中的狀況大大得到改善。不過值得注意的是，實驗 2-1 與 2-2 雖然在平均工作完成度上分別為 12.66%、12.90%，遠低於對照組實驗 1-1 的 15.22%，不過在右下方低完成度群的部分，卻是高於實驗 1-1，表示其平均工作完成度的下降係因工作被均分，導致無超高完成度群的產生，但是其實在整體的工作完成度上，反而是比實驗 1-1 來得好。

綜合上述，實驗 2-1 因為每篇文章可以停留、被閱讀的時間過短，雖然在重覆率、週轉時間上較佳，但是篩選效果差，首頁文章品質不高，議題熱門度更是遠遠低於其他幾組。相較之下，實驗 2-2 及 2-3 雖然頭條議題熱門度低於實驗 1-1，但是仍得到九成以上使用者的迴響，首頁文章品質也達到 8 以上，而且使用者閱讀滿意度亦高於實驗 1-1 許多。由此可見，適度的加入文章的過時退場處理，對於整體系統的表現是有幫助的。

另一方面，綜合實驗 1 與 2 可以發現，閱讀滿意度、閱讀重覆率、首頁文章週轉時間這三項指標可以歸為同一個群組，均和文章的更新率有關；而首頁文章平均品質、頭條議題熱門度則歸於同一個群組，反應出系統的篩選效果。這兩個群組之間是魚與熊掌不可兼得的交換關係，希望篩選效果較好、工作完成度較佳時，就會犧牲掉文章的更新頻率。因此，在實際的網站運作上，必需在這兩種不同的目標之間加以權衡。

【實驗 3】 抑制 (Inhibition) 與健全度 (robust)

在第二章關於群體智慧的探討中曾經提過，正回饋可以幫助較佳的方案從各種不同的可能性中被篩選出來，而負回饋或抑制 (Inhibition) 的機制，則可以降低錯誤決策所產生的影響，使得系統更加穩定。實驗 1 與實驗 2 中的正回饋產生方式是讓使用者在閱讀文章時決定是否推薦，並且透過熱門排行清單的閱讀，讓較受歡迎的文章得到更多的曝光機會，但是並沒有抑制或負回饋的設計。本實驗將導入抑制的概念，來觀察整體系統的表現是否有所影響。

在實驗 1 與實驗 2 中，使用者讀完一篇文章後，如果覺得這篇文章是值得推薦的，便會給予推薦，使該篇文章的推薦數增加一票；如果覺得不值得推薦，則該文章的推薦數不增加，亦不減少。在實驗 3 中，以「負推薦」的方式來進行「抑制」：亦即，如果該篇文章使用者覺得不值得推薦，則不但不增加其推薦數，反而將其原有的推薦數減少一票。不過負推薦的機制只針對熱門排行清單。其理念為，可能有部分文章因為少數使用者偏執、不客觀的判斷，加上正回饋的機制，而被篩選出來，佔有熱門排行清單中較前面的位置，但實際上其品質並沒有受到其他使用者的肯定。因此，透過對熱門排行清單的負推薦處理，來平衡不恰當的正回饋結果。

實驗 3-1 採用實驗 1 與實驗 2 中較佳的變數組合，亦即在熱門排行清單及時序排行清單中各選取 10 篇來閱讀，並且以 50 個時間單位做為文章的退場時間。不同的是，加入熱門排行清單中的「負推薦」機制。實驗結果，並與實驗 2-2，也就是相同變數設定下，無負推薦機制者來比較。相關變數設定見表 4 - 6。

表 4 - 7 呈現本實驗的各項指標結果，並與實驗 2-2 無負推薦機制設計之結果來比較，可以發現，有了抑制的負回饋後，在屬於篩選效果的首頁文章平均品質、頭條議題熱門度等兩項指標皆得到改善，而在文章的更新率方面，包括閱讀重覆率、週轉時間的表現則些微下降，不過幅度不大。值得注意的是，閱讀滿

意度在文章閱讀重覆率上升的情況下，卻逆勢有微幅的成長，可見篩選效果的確有提升，使用者真正藉由閱讀較高品質的文章，得到較高的滿意度。

表 4-6 實驗 3-1 變數設定

閱讀策略相關變數				其他變數	
變數	數值	變數	數值	NumUsers=50 OriginalArticle=300 Runs=150 Aging=50	
VL1	20	TL1	20		
VL2	70	TL2	70		
VL3	100	TL3	100		
VLc1	3	TLc1	3		
VLc2	3	TLc2	3		
VLc3	2	TLc3	2		
VLc4	2	TLc4	2		

資料來源：本研究整理

表 4-7 實驗 3-1 觀察指標模擬結果

觀察指標	實驗 2-2 (無負推薦)	實驗 3-1 (加入負推薦機制)
平均閱讀滿意度	3.48	3.5
平均閱讀重覆率	31.66 %	31.96
首頁文章平均週轉時間	14.0	15.6
首頁文章平均品質	8.2	8.43
頭條議題熱門度	92.78 %	93.18
平均工作完成度	12.90 %	11.90%

資料來源：本研究整理

除了指標的表現較佳外，圖 4-13 中顯示的滿意度與重覆率，皆呈現一個現象，即有負推薦機制設計的實驗 3-1 組，無論在滿意度、首頁頭條平均品質、平均閱讀重覆率上，都比較穩定，震盪較無負推薦機制的實驗 2-2 來得小。在實驗 2 中曾經分析過，滿意度、重覆率及首頁平均品質在經過文章退場的時間點時（本組實驗為第 50 個時間單位），會有立即的反應，首頁品質會下降、滿意度會上升、重覆率會陡降。本實驗納入抑制機制後，雖然曲線在經過時間點 50 的時候仍開始有波動，不過相對於實驗 2-2 來說是比較平滑的。這也顯示了，有負回饋機制，能使群體智慧系統更加的穩定。

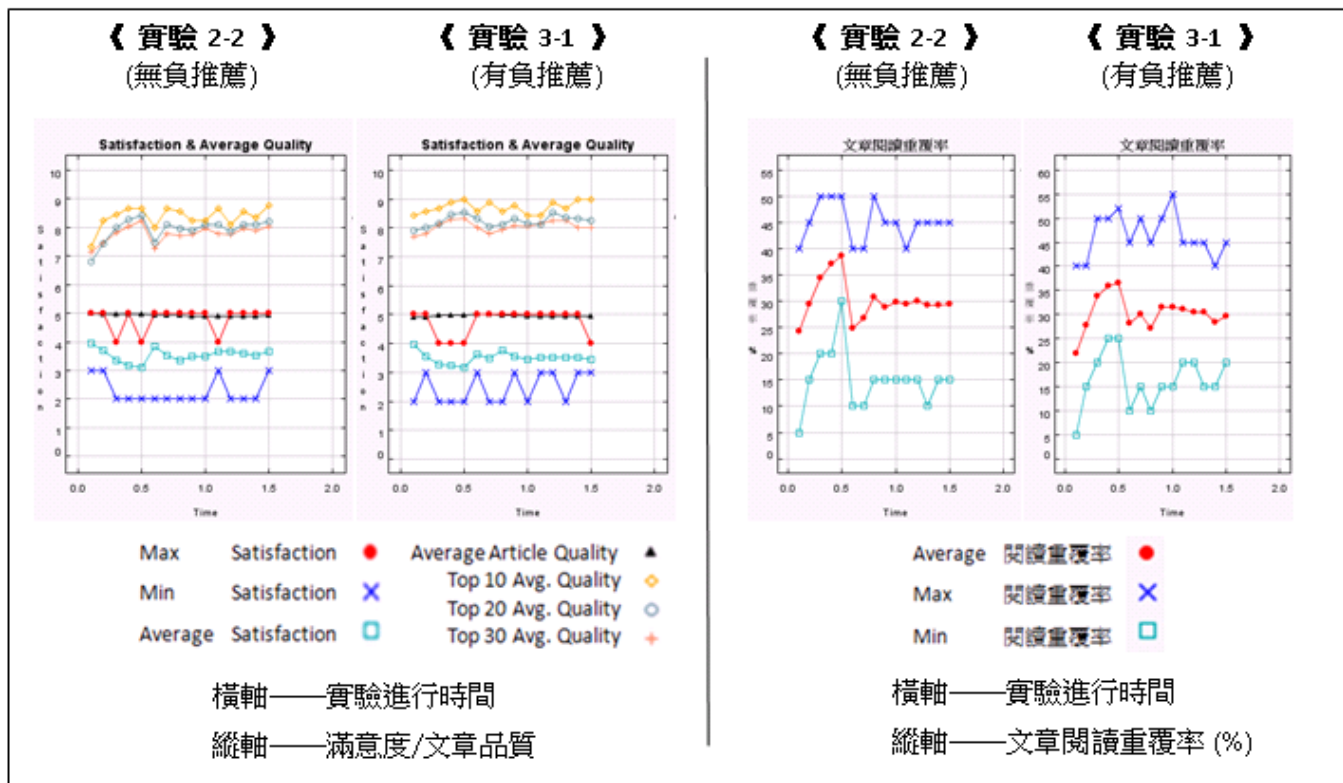


圖 4-13 實驗 3-1 使用者滿意度暨文章平均品質折線圖(左)、文章閱讀重覆率折線圖(右)

資料來源：本研究整理

為了更進一步觀察抑制的負回饋機制對系統所帶來的錯誤修復能力，在下面的實驗中導入了「非理性」的使用者，讓這些使用者做出非理性的推薦行為，觀察對系統造成的影響，以及負推薦的抑制效果是否能夠有效的修復錯誤的推薦。在本研究的設計中，理性的使用者會依據文章的品質來決定推薦與否，並設定當文章品質為 N 時，會有 N 成的比例的使用者會決定推薦這篇文章；而接下來的實驗，會讓「非理性」的使用者故意唱反調，當文章品質為 N 時，反而有 N 成的非理性使用者選擇不推薦這篇文章，藉以在平台中加入錯誤的推薦資料。在六個實驗中，分別安排了一成、三成及五成的非理性使用者，並且再分別以無負推薦機制及有負推薦機制，分別進行實驗。

表 4-8 實驗 3-2 ~ 3-7 變數設定

	實驗 3-2	實驗 3-3	實驗 3-4	實驗 3-5	實驗 3-6	實驗 3-7
非理性使用者比例	一成		三成		五成	
是否有負推薦機制	X	O	X	O	X	O
其他變數：						
NumUsers=50 OriginalArticle=300 Runs=150 Aging=50						
VL1=20 VL2=70 VL3=100 VLc1=3 VLc2=3 VLc3=2 VLc4=2						
TL1=20 TL2=70 TL3=100 TLc1=3 TLc2=3 TLc3=2 TLc4=2						

資料來源：本研究整理

表圖 4 - 10 的模擬實驗結果，將同一個非理性使用者比例的組別中，表現較佳者以背景標註。可以看到兩個趨勢：第一，隨著非理性使用者的增加，平台上的充斥越多的不合理推薦，則在各個指標中除了工作完成度及閱讀重覆率之外，其他指標的表現均隨之下降，亦即系統的整體表現是越差的；第二，有負推薦機制的組別，表現均較無負推薦機制者佳，這與實驗 3-1 的結果是相符合的，顯示抑制的作用，的確能增加群體智慧系統中的錯誤修復能力。另一方面，在圖 4 - 14 中，做縱向比較時可以明顯的發現，平台上越多的非理性推薦者，會導致首頁品質的波動越大，並且使得使用者滿意度些微下降，但在重覆率的部分則沒有太大的影響。而橫向的比較，則可以得出和實驗 3-1 相同的結論，亦即抑制可以使整體系統的表現較為穩定，減緩波動的幅度。

表 4-9 實驗 3-2 ~ 3-7 觀察指標模擬結果

實驗編號	3-2	3-3	3-4	3-5	3-6	3-7
非理性使用者比例	一成		三成		五成	
是否有負推薦機制	X	O	X	O	X	O
平均閱讀滿意度	3.48	3.57	3.32	3.46	3.21	3.24
平均閱讀重覆率	31.29%	30.97%	32.20%	31.23%	31.13%	32.02%
首頁文章平均週轉時間	14.4	16.8	13.6	15.4	13.2	14.2
首頁文章平均品質	7.98	8.29	7.27	7.66	5.08	6.5
頭條議題熱門度	92.96%	93.08%	92.04%	92.06%	90.04%	93.32%
平均工作完成度	11.97%	12.28%	12.40%	12.72%	12.40%	13.10%

資料來源：本研究整理

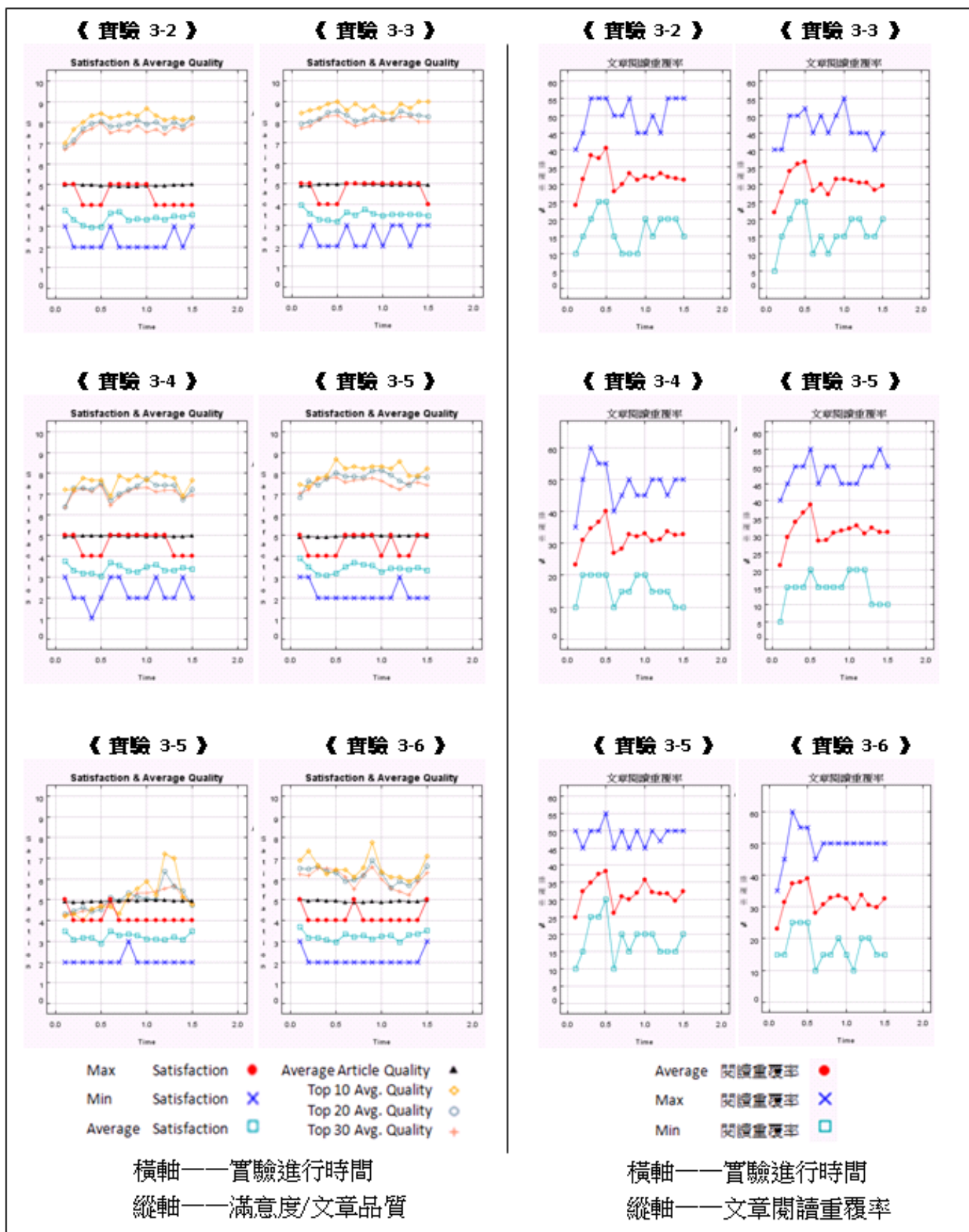


圖 4-14 實驗 3-2~3-7 使用者滿意度暨文章平均品質折線圖(左)、文章閱讀重覆率折線圖(右)

資料來源：本研究整理

【實驗 4】 熱門排行清單中各段落之閱讀效果。

在前面的實驗當中，熱門清單的四個段落每段大致上取相同的篇數（第一至第四段分別為三篇、三篇、兩篇、兩篇，總數保持在十篇）。本實驗試圖了解著重不同的段落時所會造成不同的結果。因此實驗 4-1 到 4-4，每次讓用者僅在特定一個段落中閱讀，將十篇集中在同一段中，如表 4 - 10 中陰影處所示。

表 4-10 實驗 4 變數設定

實驗 4-1 (只讀第四段)				實驗 4-2 (只讀第三段)			
變數	數值	變數	數值	變數	數值	變數	數值
VL1	20	TL1	20	VL1	20	TL1	20
VL2	70	TL2	70	VL2	70	TL2	70
VL3	100	TL3	100	VL3	100	TL3	100
VLc1	0	TLc1	3	VLc1	0	TLc1	0
VLc2	0	TLc2	3	VLc2	0	TLc2	0
VLc3	0	TLc3	2	VLc3	10	TLc3	0
VLc4	10	TLc4	2	VLc4	0	TLc4	0
實驗 4-3 (只讀第二段)				實驗 4-4 (只讀第一段)			
變數	數值	變數	數值	變數	數值	變數	數值
VL1	20	TL1	20	VL1	20	TL1	20
VL2	70	TL2	70	VL2	70	TL2	70
VL3	100	TL3	100	VL3	100	TL3	100
VLc1	0	TLc1	3	VLc1	10	TLc1	0
VLc2	10	TLc2	3	VLc2	0	TLc2	0
VLc3	0	TLc3	2	VLc3	0	TLc3	0
VLc4	0	TLc4	2	VLc4	0	TLc4	0

資料來源：本研究整理

表 4 - 11 的實驗結果中並與相同變數組合，但在各個段落中平均選取的實驗 2-2 做比較。表中並以長條背景標示，每欄中數值最大者以最長的長條標註。紅色者其值原則上越大者較佳，而藍色者則是較小較佳。從這個表中可以再次看到先前所說，在篩選效果與文章更新率之間的交換關係 (trade off)。越重視前段最熱門的文章區段，則篩選效果越佳，但更新率則越差，需要比較長的週轉

表 4-11 實驗 4 觀察指標模擬結果

	熱門度	首頁品質	週轉時間	完成度	滿意度	重覆率	註
實驗 4-1	25.80%	7.45	10	12.05%	4.12	10.20%	讀第四段
實驗 4-2	81.40%	7.55	5	11.35%	3.08	39.80%	讀第三段
實驗 4-3	91.90%	8.45	15	10.68%	2.98	45.10%	讀第二段
實驗 4-4	98.20%	7.7	32	10.17%	2.54	50.44%	讀第一段
實驗 2-2	92.78%	8.2	14	12.90%	3.48	31.66%	對照組

資料來源：本研究整理

時間且文章閱讀重覆率也偏高。相對的，越重視清單中的後段，則篩選效果越差，但文章更新度很高，同時整體工作完成度也比較高。比較特別的是實驗 4-3 僅讀第二段時之首頁品質是所有組別中最高的，推測其原因是，第二段中的文章品質相較於第一段並不遜色，但是在較低的曝光機會下無法受到更好的重視。實驗 4-3 中臨門一腳給予這些文章表現的機會，並適時淘汰在首頁中並非真的有頂尖的品質，但是在正回饋下占據頭條不放的文章，使得首頁的平均品質能顯著的提升。實驗 2-2 的對照組，平均在各段之間選取文章閱讀，其表現也在篩選效果和更新頻率之間取得了平衡，而且在滿意度、議題熱門度與完成度的表現，都是五組實驗中表現第二佳的，滿意度僅次於實驗 4-1，但實驗 4-1 的篩選效果是五組當中最差的；熱門度最佳的實驗 4-4，在重覆率、滿意度、週轉時間及完成度則均敬陪末座。由此可見，各段落之間平均分配是較佳的閱讀策略，偏重任一段都會造成結果的偏差。

第五節 實驗結果討論與網站運作建議

在第三章的圖 3-3 中曾經提到社會性書籤網站的運作架構，可以分為使用者端、介面端和演算法端。本章的實驗一與實驗四是關於使用者端的閱讀策略，在熱門排行清單及時序排行清單中的不同段落中做不同的選擇所會造成的影響；實驗二是關於系統背後的演算法端，加入過時文章的退場機制後，可以對整體運作表現有所幫助；實驗三也是偏向使用者端的行為設計，以增加負推薦功能

的方式讓使用者的行為能對系統形成抑制與負回饋，增加系統的穩定度。除此之外，在多次的實驗當中，也發現在模擬系統的運作中，篩選效果和文章更新率是兩個相對的目標，必需在兩者之間權衡。

綜合以上四個實驗，對於社會性書籤網站，仍至於其他運用到類似推薦機制的 web 2.0 網站的運作，有以下的討論與建議：

1. 在圖 3 - 6 中曾提到，控制使用者行為的手段可以透過公開演算法原則、誘因設計及介面設計等三個方式。但是控制的原則、方向、策略為何，則可由上述的實驗結果得到啟示。篩選效果與文章更新率這兩個相對的目標，通常必需取其平衡，以得到較佳的系統表現；當然，網站經營者也可以依據實際情況來決定所要偏重的運作目標。這可以藉由讓使用者同時閱讀最新及最熱門的文章、並且在熱門文章中均勻閱讀來達成。搭配控制的手段，可以在介面的設計上去調配不同文章區段的版面位置，或著以積點獎勵的方式誘導使用者均衡的閱讀各種文章。例如，依實驗一的結論，在首頁頭版中增加最新文章顯示篇數，以降低文章的週轉時間及閱讀重覆率；或依照實驗四的結論，當使用者閱讀次要熱門的文章時，可得到較多的積分獎勵，以藉此增次首頁被篩選出來的文章品質。
2. 重覆率是影響本研究模擬實驗中使用者滿意度相當重要的因素。當使用者閱讀到曾經讀過的文章時，並沒有辦法得到滿足，但是受歡迎的熱門文章勢必會在首頁停留一段時間，造成重覆閱讀。如果在實際網站的經營中加強文章的衍生討論，則熱門話題會有較多使用者投入討論，在該文章持續停留在首頁頭條的時間裡，雖然使用者會重覆讀到該篇文章，但每次可能會增加新的討論與迴響，這可以提升重覆閱讀時的使用者滿意度。
3. 在實驗三中，加入負推薦的功能後可以有效的修復錯誤，並使整體表現較於穩定。不過在實際網站運作中，還必需考慮到不理性的負推薦者惡意的負推

薦。模擬系統中大多假設使用者的行為是理性客觀的，但是在實際網站運作時則必需觀察、考慮各種非理性的使用者行為，並同樣的藉由公開演算法原則、使用者介面設計及誘因（懲罰）設計來讓使用者行為趨向理性。

4. 承上，在實驗三中的假設為，與大多數使用者的意見相左者，即為不理性的使用者。然而，有時候意見不同可能只是偏好、興趣的不同，不見得是惡意行為。這樣的一套推薦機制透過群體的投票，目標在於讓運作結果符合多數人的偏好。但如果使用者的興趣、意見過度分歧，導致有小眾的使用者無法得到滿足，則必需加強分眾或個人化的處理。
5. 本模擬實驗中的文章品質為預先設定，並以文章品質來計算使用者的閱讀滿意度、首頁文章平均品質等指標。在真實的網站運作中，文章的品質是難以客觀衡量的，但是文章的週轉時間可以得知，完成度也可以藉由 log 資料中使用者的點閱情形來計算。因為篩選效果和文章更新率成相對的關係，因此在實際網站運作中可以藉由週轉時間、熱門度等指標，來輾轉推估使用者的滿意度，做為經營的參考指標。