

植基於本體論的 RSS 個人化推薦系統的設計與實作

趙漢強¹ 廖宜恩¹ 高國峰^{1,2}

¹國立中興大學資訊科學系 ²私立修平技術學院資訊網路技術系

Email: s9356031@cs.nchu.edu.tw, ieliao@nchu.edu.tw, kfkao@mail.hit.edu.tw

摘要

RSS 是一種用來傳遞資訊(例如新聞)的 XML 文件格式,透過 RSS 的使用,供應資訊內容的人可以很容易地傳播資訊。網路上充斥多樣化的 RSS 新聞頻道,與其中的即時資訊,因此,提供個人化服務幫助使用者過濾新聞。是一個重要的課題。

本研究提出一個以使用者訂閱的本體論為基礎的個人化 RSS 推薦系統,運用新聞頻道的關係,以內容導向過濾(Content-Based Filtering)的方式和文件分類的技術,找出使用者感興趣的新聞,並提供興趣值的欄位,讓使用者能於 RSS 閱讀器上依照此欄位排序,同時配合 Sliding Window 與 Feedback 的機制,讓系統的推薦能依照使用者的需求變動。

關鍵字: 本體論、個人化、資訊過濾、文件分類

1. 簡介

資訊科技的發展,知識是人類進步的原動力,原始社會進化到文明昌盛、科技發達,都是依靠遺留下來的寶貴知識的順利傳承。現代社會快速的發展,也是因為日新月異的科技的累積與進步。資訊社會中,渴求快速取得系統化的資訊。

XML 格式的產生,使得資訊可以以半結構的方式儲存,之後的 RSS (Really Simple Syndication) 格式標準則制定了固定的 XML 格式供新聞使用。目前在各大入口網站,都有提供新聞的 RSS 文件的訂閱網址,除了新聞的提供外,目前流行的個人部落格 (blog) 也有提供 RSS 的訂閱網址。RSS 成為了現代資訊交流的一種媒介,使資訊能夠以結構化的方式寫入其中,能給更多使用者閱讀,知識因此而傳遞。

RSS 格式在目前已經成為了一種趨勢,因此,產生各種針對方便閱讀 RSS 提供訊息的方式,開發 RSS 閱讀器軟體讀取 RSS 訊息,依 RSS 的結構將訊息的內容呈現出來。

近年來,資訊的傳遞朝電子化的目標邁進,RSS 就是其中的一種媒介,RSS 已有部分的個人化概念,使用者透過訂閱不同的 RSS 來源網址,等同於自己做資訊主題的過濾。而 RSS 中包含的是即時性的資訊,以天計數的資訊量,使用者通常都在自己訂閱的 RSS 匯入器 (Feed),針對每一筆資訊自行過濾。

RSS 新聞的取得都是依賴各網站提供的訂閱 RSS 網址,使用者可以依照自己的需求選擇要訂閱的 RSS 新聞頻道,對於使用者而言依據使用者選擇的 RSS 資訊,提供個人化服務是很實用的服務。

我們的目標則是針對上述的需求,希望在現有的 RSS 閱讀器上,設計一基於本體論的 RSS 個人化推薦系統,將 RSS 的資訊過濾,以提供符合使用者興趣的 RSS 資訊。

本文結構如下:第二節為相關研究,介紹有關 RSS 個人化與推薦系統的相關技術,第三節為設計 RSS 個人化推薦系統的研究方法,第四節為 RSS 個人化推薦系統的架構,第五節為實驗與結果分析,第六節為結論與未來展望。

2. 相關研究

RSS (Really Simple Syndication) [6]是一種簡單化的 XML 技術,主要是應用在網站分享其新聞標題及提供資訊傳遞服務的整合用途。RSS 的資料格式具有簡易解析及標準的特性,很多入口網站的新聞皆以 RSS 格式發佈新聞,讓使用者可以從 RSS 的來源,方便的取得新聞。

使用者可以透過 RSS 匯入器(RSS Feed),通常是一個網址,讓使用者取得一份合於語法的 XML 文件,得到有關於此新聞頻道的資訊。使用者取得 RSS 文件後,可以透過 RSS 閱讀器 (RSS Reader) 閱讀 RSS 文件中包含的資訊,與 RSS 聚合器(RSS Aggregator)來管理自己感興趣的 RSS 匯入器。目前大多數 RSS 閱讀器皆包含 RSS 聚合器的功能。RSS 閱讀器又分為兩種方式,一種是以瀏覽器的方式提供服務,如 Bloglines, NewsGator 等,另一種則是以軟體的方式提供服務,如 NewsDesk, RSSOwl 等。RSS 聚合器的服務讓使用者管理感興趣的 RSS 匯入器,被視為是 RSS 的個人化服務的一種。

除了 RSS 聚合器的個人化服務外,網路上還有 AmphetaRate[2]的個人化推薦服務。AmphetaRate 是在伺服器端的個人化 RSS 新聞推薦的服務,針對英文新聞所提供的服務為主,透過使用者評分,再以內容導向過濾的方式比對評分新聞內容找出相似的新聞,並透過協同式過濾的方式,讓有相同興趣的使用者分享訊息,最後提供包含有個人化 RSS 資訊的 RSS 匯入器給使用者。

3. 研究方法

本研究著重的是客戶端的個人化服務，針對使用者所選的新聞頻道，依照個人的興趣，呈現出新聞頻道中使用者感興趣的新聞。因此，如何運用使用者選擇的 RSS 匯入器之間的關係以達成較好的推薦是我們的主要目標，再依據每一新聞頻道過濾 RSS 匯入器中的新聞以提供 RSS 新聞的個人化推薦服務。

我們採用內容導向過濾 (Content-based Filtering) 的方式，分析使用者提供的資訊，於使用者模型中歸納出使用者的偏好，判定使用者可能感興趣項目，提供使用者在客戶端的服務。

研究方法是針對如何提供妥善的個人化推薦服務而設計的，運用本體論的概念找出使用者訂閱的新聞頻道的關係，再應用文件分類的技術建立出使用者興趣模型，過濾新聞頻道內的新聞找出使用者感興趣的新聞，更新使用者興趣模型能維持與使用者興趣相符的模型。最後，運用興趣值公式計算推薦新聞的興趣值，呈現給使用者，讓使用者能進行排序以達到更妥善的推薦服務。

3.1 本體論

本體論 (Ontology) 能表現出網路上許多知識這些的分類結構，本體論的定義是將概念明確的描述。RSS 是資訊傳遞的媒介，存在 RSS 中的資訊有經過分門別類，每個 RSS 新聞頻道都可視為一個概念，我們將本體論運用在新聞頻道的關係上，能更準確的推論出使用者的興趣。

我們由各大網站提供的 RSS 新聞頻道定義出 RSS 新聞頻道的本體論，如圖 1，概念之間的關係為父子類別的關係，使用者訂閱的新聞頻道可以對應在本體論上，產生個人本體論，運用新聞頻道間父子關係，使得推薦更為準確，更符合使用者興趣。

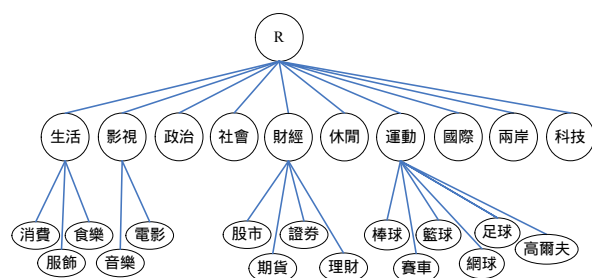


圖 1 新聞頻道本體論架構圖

由本體論架構圖 2，使用者感興趣的新聞頻道為 T_i 類別， $C_1 \sim C_m$ 為 T_i 的特殊化，則 T_i 有可能包含其下的 $C_1 \sim C_m$ 的新聞。若使用者對 C_1 較其他感興趣，我們就可以推論出，在 T_i 中若出現 C_1 類型的新聞，則使用者對這類的新聞會較其他子類別類型的新聞感興趣。

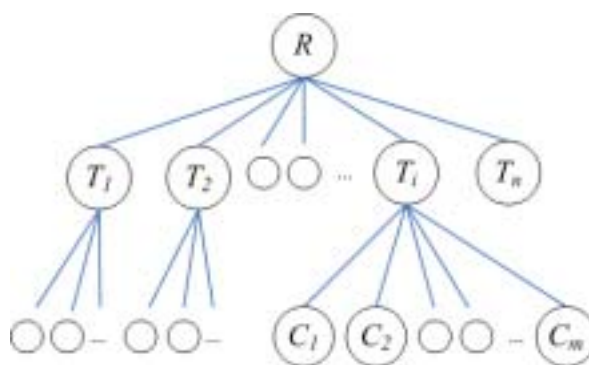


圖 2 本體論架構圖

我們以各新聞頻道閱讀新聞的次數來代表使用者對此新聞頻道的興趣，使用者對該新聞頻道的興趣與該新聞頻道及其下分支新聞頻道的閱讀數目成正比，對 T_i 新聞頻道的總閱讀數目的計算如公式 1 所示。

$$S_{T_i} = R_{T_i} + \sum_{j=1}^m R_{C_j}, \quad i = 1, 2, \dots, n \quad (1)$$

S_{T_i} : T_i 新聞頻道的總閱讀數目

R_{T_i} : T_i 新聞頻道的閱讀數目

R_{C_j} : T_i 的子新聞頻道 C_j 的閱讀數目

針對 T_i 新聞頻道的興趣，我們會由子類別 C_k 在 $C_1 \sim C_m$ 加總的閱讀數目中佔的比例推測出使用者在 T_i 新聞中出現子類別相關新聞的興趣度，各子類別的閱讀比例計算如公式 2。

$$P_{C_k} = S_{C_k} / \sum_{j=1}^m S_{C_j} \quad (2)$$

P_{C_k} : C_k 子類別的閱讀比例

S_{C_k} : C_k 的閱讀數目

S_{C_j} : C_j 的閱讀數目

圖 3 為計算子類別閱讀比例的實例，B 類別有 3 個子類別，S 為總閱讀數目，P 為閱讀比例，B 的子類別閱讀數目由左到右分別為 3、6、3，則閱讀比例依序為 $3/(3+6+3) = 0.25$ 、 $6/(3+6+3) = 0.50$ 與 $3/(3+6+3) = 0.25$ 。則我們可以由子類別的閱讀比例來針對 B 類別中有出現子類別的新聞進行推薦。

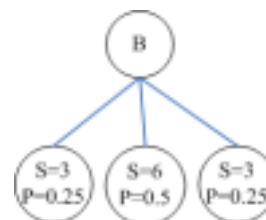


圖 3 各類別的閱讀比例

3.2 文件分類

依據各別新聞頻道中使用者閱讀的記錄，運用監督式文件分類的技術找出使用者興趣模型 (User

Profile)，以使用者感興趣與否做為文件的類別，將預測的文件歸類在這兩種類別之中，換句話說，運用文件分類技術在各別新聞頻道中找出使用者感興趣與不感興趣新聞之間的區別。

文件分類的處理步驟包含前置處理、特徵選擇與產生分類規則。

■ 前置處理

前置處理的主要目的是將使用者閱讀的記錄轉換為可以進行分析的關鍵詞的組合，將使用者每一筆閱讀的新聞內容視為一文件，針對每筆文件擷取關鍵詞，將文件轉換為關鍵詞的組合。

■ 特徵選擇

前置處理後，再挑選出有代表性的關鍵詞來代表文件，我們以 TFIDF [7]的公式，如公式 3，計算每一文件中各關鍵詞的權重，以 TFIDF 權重大到小選擇出在文件中前 4 個重要的關鍵詞為特徵關鍵詞代表此文件。

$$W_{ij} = TF \times IDF = \frac{N_{ij}}{N_{all}} \times \log_2 \frac{N}{df_j} \quad (3)$$

W_{ij} ：關鍵詞 j 在文件 i 的權重值 (Weight)

N_{ij} ：單字 (詞) j 在文件 i 的出現次數

N_{all} ：文件 i 中所有具有意義的詞類數

N ：所有文件總數

df_j ：表單字 (詞) j 有出現過的文章總數

■ 產生分類規則

由於 SVM[3]是數值式的分類器，將文件必須將每一份文件轉換成為文字向量，文件向量是將選擇出的特徵關鍵詞比對份文件，以關鍵詞為屬性向量，可以把每一份文件轉換為一文字向量，屬性為特徵關鍵詞，而其中值為 1 代表文件存在該關鍵詞，0 表示文件不存在該關鍵詞，配合使用者感興趣與否的類別，以 SVM 產生分類模型 (Classification Model)，之後，透過預測文件中包含的特徵關鍵詞就可以預測新的文件是否為感興趣。

3.3 更新使用者興趣模型

我們使用了 Sliding Window 與 Feedback 機制，讓使用者興趣模型，能夠維持符合使用者興趣的狀態。

■ Sliding Window

RSS 格式的目的在於新資訊的傳遞上，而使用者對新資訊的興趣，應當和過舊的資料較無相關，本研究採用 Sliding Window 的方式，隨著時間過去，系統會隨著使用者最近的興趣，對使用者興趣模型做更動。使系統固定以最近一段時間內的使用者提供的資訊做為訓練資料，進行訓練，產生能針對使用者的興趣的使用者興趣模型。

本研究以重新訓練 (Retraining) 的方式更新使用者興趣模型在軟體結束之前會重新訓練，將使用者模型更新，以待下次軟體執行時使用，在每次訓

練時，選擇以最近 14 天內的使用者閱讀記錄為訓練資料。14 天一般能收集到數百筆的資訊，應當能反應出使用者的興趣。

■ Feedback

系統為了適度的調整使用者模型，讓使用者模型能夠滿足使用者的需要，所以，提供了回饋機制，讓使用者可以提供回饋給系統，系統就能從使用者提供的回饋對使用者興趣模型進行修正，以滿足使用者的需要。

使用者提供的回饋資訊分為二類，正回饋 (Positive Feedback) 和負回饋 (Negative Feedback)。正回饋，表示系統對項目的判斷滿足使用者的需求；負回饋，表示系統對項目的判斷不符合使用者的需求。

使用者可針對系統推薦新聞來給予回饋，因此，每一則回饋資訊是由使用者判斷系統推薦的新聞或是系統不推薦的新聞產生。

由這些我們可以把回饋資訊，分成 4 種不同的情境處理。

使用者對系統推薦新聞判斷為正回饋，代表使用者滿意於系統的推薦。系統將和該筆回饋相關的使用者閱讀記錄修改為已閱讀，日後相關回饋的資料能以更高的分數進行推薦。

使用者對系統推薦新聞判斷為負回饋，代表系統推薦不符合使用者的需求。系統將和該筆回饋相關的使用者閱讀記錄修改為沒閱讀，以達成回饋。

使用者對系統不推薦新聞判斷為正回饋，代表系統不推薦的新聞符合使用者的需求。系統的推薦規則不需做更動。

使用者對系統不推薦新聞判斷為負回饋，代表系統誤判不推薦的新聞。系統會將和該筆回饋相關的使用者閱讀記錄修改為已閱讀，以達成回饋。

3.4 興趣值計算

興趣值的計算，我們基於以下的想法：使用者對新聞感興趣是因為新聞中出現使用者感興趣的關鍵詞。我們運用文件分類中特徵選擇的技術可以取得特徵關鍵詞。興趣值計算可由該推薦新聞中的特徵關鍵詞，配合以往的使用者閱讀記錄中該新聞頻道中該關鍵詞出現的分佈，推測出來。關鍵詞的分佈依據使用者記錄中各關鍵詞出現的新聞閱讀與否的比例來決定，運用這些相關度量，進行運算以找出推薦新聞在該新聞頻道的興趣度。

我們考慮的關鍵詞的相關量度包括正頻率、負頻率、正文件頻率與負文件頻率，以圖 4 說明之，假設使用者透過訂閱得到 4 份文件，每份文件皆包含某個關鍵詞，使用者閱讀文件 1 與 2，文件 1 與文件 2 中該關鍵詞分別出現 3 次、2 次，所以正頻率為 3+2，使用者閱讀了 2 份文件，因此，正文件頻率為 2。相反，使用者未曾閱讀文件 3 與 4，所

以負頻率為 1+2，負文件頻率為 2。關鍵詞興趣度如公式 4 由關鍵詞相關度量計算得到。

	關鍵詞出現次數	使用者閱讀?
文件1	3	1
文件2	2	1
文件3	1	0
文件4	2	0

圖 4 關鍵詞相關度量示意圖

$$it = \frac{Cp}{Cp + Cn} \times \frac{Dp}{Dp + Dn} \quad (4)$$

it : 關鍵詞興趣度

Cp : 關鍵詞的正頻率

Cn : 關鍵詞的負頻率

Dp : 關鍵詞的正文件頻率

Dn : 關鍵詞的負文件頻率

我們依該筆 RSS 新聞中出現的關鍵詞興趣度的平均，就是該筆內容的興趣度，如公式 5 所示，若該筆新聞，是由子新聞頻道推薦所得，則乘上該子新聞頻道的閱讀比例，若否，則為 1。

$$I = \frac{1}{N} \sum_{i=1}^N P \times it_i \times tf_i \quad (5)$$

I : 新聞興趣度

N : 推薦新聞有 N 個特徵關鍵詞

P : 子新聞頻道閱讀比例

it_i : 關鍵詞興趣度

tf_i : 關鍵詞在新聞的出現次數

由於大多數人的興趣會依時間流進而變化，因此，我們需要一個隨時間變化的方式來計算出興趣值。本研究將 14 天收集的使用者閱讀資訊放置在 7 個槽(Slot)中。演算法如下：

1. 取得與最近一次記錄相距天數 k 。
2. 往後位移 Slot 內的資訊 $k/2$ 次。
3. 若 1st Slot 中記錄了兩天的資訊且 $k=1$ ，則將 Slot 內的資訊往後位移 1 次。
4. 記錄使用者興趣資訊到 1st Slot 內。

透過這 7 個槽記錄關鍵詞對於新聞頻道的相關度量。我們運用這 7 個槽的時間順序的特性，以第一個槽為第一區段，第二到三槽為第二區段，第四到七槽則為第三區段，採半衰期方式遞減權重，來計算出該新聞的總興趣度 (SI)，如公式 6 所示。

$$SI = \sum_{x=1}^3 T_x \times I_x \quad (6)$$

SI : 新聞的總興趣度

I_x : 第 x 個區段的記錄計算出的新聞興趣度

T_x : $2^{(3-x)}$ ，第 x 區段的加權值，採半衰期遞減。

4. 系統模組

系統架構如圖 5，區分為四個核心模組：應用介面 (Application Interface) 記錄模組 (Recording Module) 訓練模組 (Training Module) 和預測模組 (Prediction Module)，系統以這四個核心模組整合實現研究方法。

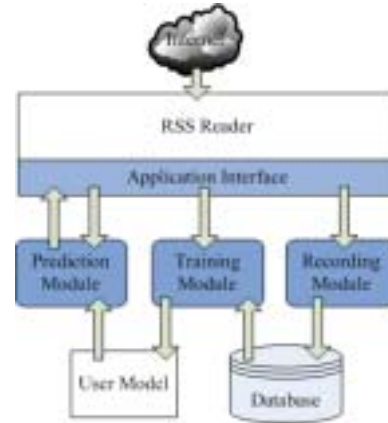


圖 5 系統模組示意圖

■ 應用介面

應用介面是做為軟體和後端模組連結的中間層，透過修改 RSS 閱讀器的內容，使其能夠導向使用者的閱讀資訊和後端的模組相溝通，並能夠將預測的資訊呈現於軟體上。

■ 記錄模組

記錄模組的功能是儲存使用者瀏覽 RSS 新聞的相關記錄，記錄模組的記錄下的資訊包括使用者所點選的 RSS 新聞頻道的名稱及 RSS 新聞頻道中的新聞內容，與使用者閱讀了哪些新聞，以及使用者提供的回饋 (Feedback)，取得和使用者興趣相關的資訊供訓練模組分析。

■ 訓練模組

訓練模組運用在研究方法中所提的本體論、文件分類與更新使用者模型的技術。訓練模組的主要目的是產生和使用者興趣相符的使用者模型。

首先，將記錄下來的新聞標題，比對我們定義的 RSS 本體論中的概念 (關鍵詞)，可以建立出使用者的個人本體論，在每次訓練時，將超過 14 天以上的使用者記錄刪去，以實現 Sliding Window，再將回饋反應到訓練資料中，再進行文件分類，產生每一個新聞頻道的使用者興趣模型，訓練模組的演算法如圖 6 所示。

- 1 將訂閱的新聞頻道標題與本體論中的概念 (關鍵詞) 對應產生個人本體論
- 2 For 每一新聞頻道
- 3 刪除超過 14 天之使用者記錄
- 4 將回饋修改到訓練資料中
- 5 文件分類:前置處理
- 7 文件分類:特徵選擇
- 8 文件分類:產生 SVM 分類模型
- 10 EndFor

圖 6 訓練模組演算法

文件分類步驟中的前置處理，分析使用者記錄，並記錄下關鍵詞相關度量，作為興趣值計算的依據，與各新聞頻道使用者感興趣新聞的標題關鍵詞，以做為父類別新聞頻道中的新聞判斷是否為子類別新聞頻道的新聞的依據。

■ 預測模組

預測模組的功能是提供新聞的預測資訊，運用訓練模組的個人本體論，與各新聞頻道的分類模型，預測未來出現的新聞是否感興趣，若為感興趣則計算興趣值在介面上以欄位的方式呈現給使用者，達到推薦使用者新聞的目的，演算法如圖 7 所示。

```
1 If 預測新聞的新聞頻道有子新聞頻道且
  預測新聞有包含子新聞頻道之關鍵詞{
3   使用該子新聞頻道分類模型預測
4   If 預測結果為感興趣{
5     由子新聞頻道之關鍵詞相關度量計算
      出新聞興趣度。
6   Else
7     新聞為不感興趣。
8   EndIf
9   If 新聞為不感興趣
10    由新聞來源新聞頻道分類模型預測
11    If 預測結果為感興趣{
12      由新聞來源新聞頻道之關鍵詞相關度
          量計算出新聞興趣度。
13    Else
14      新聞為不感興趣。
15  EndIf
```

圖 7 預測模組演算法

運用這些預測的結果，呈現於介面上。預測模組的目的是要自動的挑選使用者感興趣的新聞，讓使用者不需花大量心力去找出自己欲閱讀的新聞在進行閱讀，並且節省使用者的時間。

5. 實驗與結果分析

對於前面所提出的研究方法與架構，進行驗證，並提出實驗數據與實驗數據的分析，以確定我們的方法的成效如何。

5.1 系統環境

我們透過修改 RSSOwl 驗證我們提出的方法，RSSOwl 是 RSS 閱讀器且有 RSS 聚合器的功能，是由 Java 語言開發的軟體，以 IBM 開發的 SWT (Standard Widget Toolkit) 元件為基礎開發圖形介面；選用了 Mysql 儲存系統模組所需要的資訊，透過 JDBC 存取；文件分類我們以 LIBSVM[5]的函式

庫產生的分類模型提供服務；使用 CKIP[1]，由中研院開發的中文斷詞系統，用以分析出中文文章中，內含文字的詞性，以擷取出文章的關鍵詞，我們擷取下名詞、英文的部分，以供進行分析。

5.2 系統畫面

我們修改 RSSOwl 使其能儲存下使用者閱讀的資訊，包含新聞頻道資訊與新聞內容和回饋資訊，處理這些資訊，以提供推薦服務。我們將 RSSOwl 的新聞顯示方式修改為如圖 8 所示，於新聞列表提供一個欄位表示預測的資訊，顯示由預測模組計算出的興趣值，並提供使用者排序的功能。

若該筆新聞為推薦新聞，則前面會有興趣值，代表該筆新聞興趣的資訊；若沒有可供預測的關鍵詞，會以“?”來顯示；“-”則是代表該筆新聞為不推薦。



圖 8 系統顯示推薦的畫面

回饋資訊，以使用者在該推薦新聞上點擊右鍵會出現 Provide Feedback 的選單，選擇“Positive Feedback”或“Negative Feedback”，回饋給系統，如圖 9。

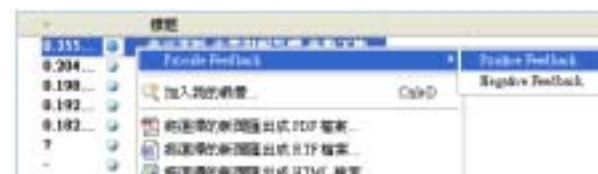


圖 9 使用者提供回饋介面

5.3 實驗

我們以使用者對系統推薦的新聞的滿意度做為實驗的評分標準，讓使用者對推薦新聞評分，有五個等級，分別為“非常滿意”、“滿意”、“普通”、“不滿意”、“非常不滿意”，分別以分數 0.9, 0.7, 0.5, 0.3, 0.1 表示。由這些分數對系統推薦的效果進行評估。

藉由記錄下的使用者對推薦系統評分資訊，以三種標準來評估推薦系統的滿意度，第一種為計算

使用者對推薦系統的平均滿意度、第二種採用 Breese 與 Kadie[4]所提出的排序分數度量值(Rank Score Measure)。第三種為系統誤判率。

■ 平均滿意度

平均滿意度為先計算出每位使用者的滿意度平均值為其個人滿意度，再計算所有使用者個人滿意度的平均值，即為平均滿意度。

■ 排序分數度量值

排序分數度量值是計算使用者對系統推薦的排序的滿意程度，計算方式如公式 7。如果使用者喜歡的新聞都在推薦新聞中比較前面的位置，而不喜歡的在排序比較後面的位置，則 R_x 可以得到較大的值。反之，那 R_x 的值就會很低。若系統排序完全依照使用者滿意度排序，則可以得到最大的值 R_x^{max} 。排序分數度量值需要正規化，如公式 8。

$$R_x = \sum_j \frac{\max(v_{x,j} - d, 0)}{2^{(j-1)(\alpha-1)}} \quad (7)$$

R_x ：使用者 x 對新聞排序的滿意度

j ：推薦新聞排序的位置

$V_{x,j}$ ：是使用者 x 對推薦新聞 j 的滿意度

d ：為使用者滿意度的平均值。

α ：推薦數量，其作用為此公式中的半衰期。

$$R = 100 \times \frac{\sum_x R_x}{\sum_x R_x^{max}} \quad (8)$$

■ 系統誤判率

我們詢問使用者提供系統預測錯誤的比例，將系統誤判分類的數目統計出來，以計算出在預測新聞中，系統誤判分類的機率。

我們將此推薦系統給 10 人使用，以蕃薯藤的運動新聞網的頻道讓使用者訂閱，並包含運動的新聞頻道來進行測試。透過記錄下使用者對推薦新聞的滿意度，與系統的誤判數目，來評估系統。使用者共使用 4 天，一天的新聞量約為 150 筆左右。取得這 3 天間系統推薦的結果，如表 1 所示。

表 1 使用者回覆系統推薦的分析結果 (一)

	平均滿意度	排序分數度量值	平均誤判率
第二天	0.7575	76.0340	8.3%
第三天	0.7615	84.7431	7.3%
第四天	0.7565	85.2349	6.7%

表 2 呈現出在父類別與子類別之間的實驗結果的差異，運動頻道為父類別，其下的子類別頻道可能是足球、籃球或賽車等。

表 2 使用者回覆系統推薦的分析結果 (二)

	平均滿意度	排序分數度量值	平均誤判率
運動	0.7290	72.4452	11.7%
子類別	0.7732	87.7833	5.3%

5.4 實驗結果分析

實驗結果可以觀察到平均滿意度在 0.7 以上，代表推薦系統可以達到滿意的程度。

實驗結果可以觀察到在父類別頻道的誤判率高於子類別頻道的誤判率，且父類別頻道的滿意度低於子類別頻道的滿意度，這是因為，運動頻道的新聞除了自己本身的訓練模型產生的推薦外，還有子類別產生的推薦。對使用者來說，子類別的興趣很高，相反，運動頻道內的新聞很分散，若不考慮子類別的新聞在內，使用者對運動頻道閱讀量是較少的，因此，父類別（運動）的推薦結果主要受到子類別是否推薦而決定，如此，父類別的誤判率，必然高於子類別的誤判率。

6. 結論與未來展望

本論文研究著重在客戶端提供 RSS 個人化推薦系統，讓使用者能以軟體提供的友善的介面去瀏覽新聞，且以本體論的方式利用使用者訂閱新聞頻道的階層關係，產生推薦，並在各別新聞頻道以文件分類的技術與分類器來進行分類，推論出使用者的興趣，是一結合本體論的概念以達成符合使用者多重興趣的 RSS 個人化推薦系統，實驗的結果顯示本推薦系統的平均滿意度為滿意。

本研究的效果基於關鍵詞的擷取，未來可以考慮找尋更合適的特徵詞擷取方式，以提供更準確的推薦服務，另外，未來可以設計針對英文的個人化推薦系統。

參考文獻

- [1] 中文詞知識庫小組, “中文句結構樹建立原則” 網址: <http://godel.iis.sinica.edu.tw/CKIP>
- [2] AmphetaRate Overview <http://amphetarate.newsfairy.com/index.php>
- [3] Boser, B.E., Guyon I.M., and Vapnik V.N., “A Training Algorithm for Optimal Margin Classifiers,” *Proc. Fifth Ann. Workshop Computing Learning Theory*, 1995, pp. 144-152.
- [4] Breese, J.S., Heckerman, D., and Kadie, C., “Empirical Analysis of Predictive Algorithms for Collaborative Filtering.” *Fourteenth Annual Conference on Uncertainty in Artificial Intelligence(UAI-98)*, 1998, pp. 43-52.
- [5] Chang, C.C., and Lin C.J., LIBSVM, available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [6] RSS 2.0 Specification, available at <http://blogs.law.harvard.edu/tech/rss>
- [7] Sparck Jones, K., “A statistical interpretation of term specificity and its application in retrieval,” *In Journal of Documentation*, 28(1), 1972, pp. 11-21.