

Audience 1.1 站台使用方法

RD2 黃彥鈞

大綱

- 更新說明
- 測試說明
- 使用說明
- 其他使用須知

更新說明

族群貼標任務調整內容:

1. 重構族群貼標任務方法，分離預測貼標功能為後端 API
2. 目前可以使用的貼標模型有關鍵字模型與正則表達模型
3. 微調前端網站功能，新增查看任務資訊頁面

資料整備任務與模型建立任務目前還未調整

使用者可以仍可透過現有服務更新規則，但無法訓練模型

測試說明1

- 站台網址: http://172.18.20.190:5000/audience/predicting_jobs/
- 麻煩同仁幫忙快速測試第二頁：2021 男女作者關鍵字(3) wh_tiktok 來源資料

任務列表

Show 10 entries Search:

#	任務名稱	任務描述	應用模型數	預測範圍數	建立者	建立時間	狀態
12	已婚_內文_正則表達式_2	ychuang created	1	4	ychuang	2021年12月3日 14:24	處理中
13	男女作者名稱	測試男女作者關鍵字模型貼標	2	3	ychuang	2021年12月7日 09:36	完成
14	2021男女作者關鍵字(1)	2021 年數據 男女作者關鍵字模型	2	5	ychuang	2021年12月13日 13:12	完成
15	2021男女作者關鍵字(2)	2021 年數據 男女作者關鍵字模型	2	7	ychuang	2021年12月14日 10:19	處理中
16	2021男女作者關鍵字(3)	2021 年數據 男女作者關鍵字模型	2	7	ychuang	2021年12月14日 13:23	處理中
17	2021男女作者關鍵字(4)	2021 年數據 男女作者關鍵字模型	2	5	ychuang	2021年12月14日 13:41	等待中
18	2021男女作者關鍵字(5)	2021 年數據 男女作者關鍵字模型	2	5	ychuang	2021年12月14日 13:49	等待中

Showing 11 to 17 of 17 entries

Previous 1 2 Next

測試說明2

新增預測資料範圍

Show 10 entries

Search:

#	資料範圍名稱	狀態	資料源	時間範圍	
102	wh_youtube	等待中	wh_youtube	2021年1月1日 ~ 2021年12月31日	操作 ▾
103	wh_twitter	完成	wh_twitter	2021年1月1日 ~ 2021年12月31日	操作 ▾
104	wh_tiktok	等待中	wh_tiktok	2021年1月1日 ~ 2021年12月31日	操作 ▾

11日 操作 ▾

月3 開始貼標

11日 取消貼標

11日 查看抽樣結果

月3 查看任務資訊

11日 修改

11日 刪除

測試說明3

- 開始貼標後約等 半分鐘 至 一分鐘，選擇以下：
 - 查看抽樣結果 是否能回傳如下一圖結果資料
 - 查看任務資訊 是否能回傳如下二圖任務資訊



任務ID	來源作者	標籤	爬取時間	來源ID	命中內容
6e0a9d925cbc11ecbf0000c2915d4f5	WH_F0127_醒醒吧女孩們	/female	2021-01-04T20:39:15	WH_F0127	醒醒吧女孩們
6e0a9d925cbc11ecbf0000c2915d4f5	WH_F0127_辰亦儒calvin	/male	2021-01-05T18:07:54	WH_F0127	辰亦儒calvin
6e0a9d925cbc11ecbf0000c2915d4f5	WH_F0127_Ken Worker	/male	2021-01-07T22:05:36	WH_F0127	Ken Worker
6e0a9d925cbc11ecbf0000c2915d4f5	WH_F0127_阿飛大叔	/male	2021-01-08T10:08:36	WH_F0127	阿飛大叔
6e0a9d925cbc11ecbf0000c2915d4f5	WH_F0127_醒醒吧女孩們	/female	2021-01-08T15:12:58	WH_F0127	醒醒吧女孩們
6e0a9d925cbc11ecbf0000c2915d4f5	WH_F0127_Ken Worker	/male	2021-01-20T14:28:41	WH_F0127	Ken Worker
6e0a9d925cbc11ecbf0000c2915d4f5	WH_F0127_阿飛大叔	/male	2021-01-21T18:46:56	WH_F0127	阿飛大叔
6e0a9d925cbc11ecbf0000c2915d4f5	WH_F0127_Ken Worker	/male	2021-01-22T08:48:27	WH_F0127	Ken Worker

任務ID	來源	貼標進度	任務狀態	開始時間	應用模型	預測對象	貼標率	檢查點	錯誤訊息
6e0a9d925cbc11ecbf0000c2915d4f5	wh_twitter	SUCCESS	finish	2021-12-14T17:01:31	keyword_model	author	20.2	None	None

測試說明4

- 右上角操作 > 查看任務資訊 確認是否出現多筆 任務資訊



2021男女作者關鍵字(3)

處理中

ychuang created this job at 2021年12月14日 13:23

2021 年數據 男女作者關鍵字模型

取消貼標任務 操作 ▾

應用模型列表

新增模型

Show 10 ▾ entries

Search:

#	↑↓	模型名稱	↑↓	模型類型	↑↓	優先度	↑↓	
24		男性作者名稱		關鍵字規則		0		操作 ▾
25		女性作者名稱		關鍵字規則		0		操作 ▾

預測資料範圍列表

新增預測資料範圍

Show 10 ▾ entries

Search:

#	↑↓	資料範圍名稱	↑↓	狀態	↑↓	資料源	↑↓	時間範圍	↑↓	
102		wh_youtube		等待中		wh_youtube		2021年1月1日 ~ 2021年12月31日		操作 ▾

測試說明5

- 若以上功能均無問題，代表站台貼標功能可以順利進行
- 其他設計如取消貼標，用途在於執行錯誤或重複任務，或是抽樣結果有問題要提前終止
- 任務**操作**建議在預測資料範圍列表進行，右上角的**操作** (開始貼標與取消貼標) 會影響該任務下全部的子任務。

2021男女作者關鍵字(3)

取消貼標任務 操作 ▾

處理中

ychuang created this job at 2021年12月14日 13:23

2021 年數據 男女作者關鍵字模型

應用模型列表

新增模型

Show 10 entries

Search:

#	↑↓	模型名稱	↑↓	模型類型	↑↓	優先度	↑↓	↑↓
24		男性作者名稱		關鍵字規則		0		操作 ▾
25		女性作者名稱		關鍵字規則		0		操作 ▾

預測資料範圍列表

新增預測資料範圍

Show 10 entries

Search:

#	↑↓	資料範圍名稱	↑↓	狀態	↑↓	資料源	↑↓	時間範圍	↑↓	↑↓
102		wh_youtube		等待中		wh_youtube		2021年1月1日 ~ 2021年12月31日		操作 ▾

使用說明1 族群貼標任務首頁

AUDIENCE
TOOLKITS

說明書

資料準備任務

模型建立任務

族群貼標任務

族群貼標任務

任務列表

Show 10 entries

Search:

#	任務名稱	任務描述	應用模型數	預測範圍數	建立者	建立時間	狀態
12	已婚_內文_正則表達式_2	ychuang created	1	4	ychuang	2021年12月3日 14:24	處理中
13	男女作者名稱	測試男女作者關鍵字模打貼標	2	3	ychuang	2021年12月7日 09:36	完成
14	2021男女作者關鍵字(1)	2021 年數據 男女作者關鍵字模型	2	5	ychuang	2021年12月13日 13:12	錯誤

Showing 11 to 13 of 13 entries

Previous 1 2 Next

建立新任務

族群貼標流程

設定好欲此用的模型列表，與目標資料範圍，系統就會排定貼標任務。

狀態說明

處理中：有任務未完成

完成：所有任務完成

錯誤：其中有任務錯誤

進入任務頁面

應用模型數
使用多少模型，如：
男性作者姓名 與 女性作者姓名 模型

預測範圍數
預測來源數量，如：預測 wh_bbs_01、
wh_bbs_02 共兩個資料庫來源

使用說明2 應用模型

2021男女作者關鍵字(1)

ychuang created this job at 2021年12月13日 13:12

2021 年數據 男女作者關鍵字模型

應用模型列表

新增模型

Show 10 entries Search:

#	↑↓ 模型名稱	↑↓ 模型類型	↑↓ 優先度	↑↓
20	男性作者名稱	關鍵字規則	0	操作
21	女性作者名稱	關鍵字規則	0	操作

Showing 1 to 2 of 2 entries

Previous 1 Next

修改或刪除模型

優先度不要調整，之後會進一步優化

1. 現階段應用模型功能僅限新增規則模型與關鍵字模型
2. 同任務請新增相同類型模型
 - 關鍵字
 - 正則表達
3. 優先度現階段請不做調整

使用說明3 來源資料

預測資料範圍列表

新增預測資料範圍

Show 10 entries

Search:

#	↑↓ 資料範圍名稱	↑↓ 狀態	↑↓ 資料源	↑↓ 時間範圍	↑↓
90	wh_bbs_01	處理中	wh_bbs_01	2021年1月1日 ~ 2021年12月31日	操作 ▾
91	wh_bbs_02	處理中	wh_bbs_02	2021年1月1日 ~ 2021年12月31日	操作 ▾
92	wh_forum_01	錯誤	wh_forum_01	2021年1月1日 ~ 2021年12月31日	操作 ▾
93	wh_forum_02	等待中	wh_forum_02	2021年1月1日 ~ 2021年12月31日	操作 ▾

1. 新增預測資料可以自定義時間範圍與目標資料庫

2. 狀態

- 處理中：任務還在執行
- 錯誤：任務執行發生錯誤
- 完成：任務執行完成
- 中斷：使用者取消貼標
- 等待中：任務尚未執行

使用說明4 來源資料-操作

查看抽樣結果

任務ID	來源作者	標籤	爬取時間
0e0ee3265bd711eca43c04ea56825bad	WH_F0001_AaronWang	/male	2021-01-01T00:00:00
0e0ee3265bd711eca43c04ea56825bad	WH_F0001_nekkibasara	/female	2021-01-01T00:00:00
0e0ee3265bd711eca43c04ea56825bad	WH_F0001_Allen0315	/male	2021-01-01T00:00:00
0e0ee3265bd711eca43c04ea56825bad	WH_F0001_doverdover	/female	2021-01-01T00:00:00
0e0ee3265bd711eca43c04ea56825bad	WH_F0001_boyambrose	/male	2021-01-01T00:00:00

查看任務狀態

任務ID	來源	貼標進度	任務狀態	開始時間	應用模型	預測對象	貼標率	檢查點	錯誤訊息
0e0ee3265bd711eca43c04ea56825bad	wh_bbs_01	SUCCESS	finish	2021-12-13T13:40:12	keyword_model	author	21.64	None	None

1. 開始貼標：執行任務
2. 取消貼標：使用者中斷任務，該任務無法再次被執行，再次選開始貼標會建立新的任務
3. 任務修改或刪除建議開始貼標前調整，若要中斷任務請取消貼標再刪除，避免資源浪費。
4. 開始貼標後可針對單筆來源資料查看抽樣結果與查看任務狀態

使用說明5 任務頁面-操作



1. 左上開始貼標、取消貼標任務、重新貼標任務，可以令任務執行所有來源資料貼標任務
2. 透過修改和刪除編輯任務資訊
3. 查看任務資訊可以列出所有來源資料任務資訊，下一頁有範例

使用說明6 任務頁面-操作

任務ID	來源	貼標進度	任務狀態	開始時間	應用模型	預測對象	貼標率	檢查點	錯誤訊息
0e0ee3265bd711eca43c04ea56825bad	wh_bbs_01	SUCCESS	finish	2021-12-13T13:40:12	keyword_model	author	21.64	None	None
d083d1b35bdd11ec96da04ea56825bad	wh_bbs_02	SUCCESS	finish	2021-12-13T14:27:58	keyword_model	author	21.6	None	None
daa6751d5bf811ec844f04ea56825bad	wh_forum_01	SUCCESS	finish	2021-12-13T17:41:32	keyword_model	author	20.81	None	None
dbe61ad35bf811ec85c104ea56825bad	wh_forum_02	SUCCESS	finish	2021-12-13T17:41:34	keyword_model	author	17.38	None	None
22536d2d5c7c11ec96ba04ea56825bad	wh_plurk	PENDING	None	2021-12-14T09:21:16	keyword_model	author	None	None	None

1. 貼標率：若執行失敗或未結束為 **None**
貼標率 = 結果資料表 (有被貼到標籤) 之不重複來源作者數 / 來源資料庫不重複來源作者數
2. 檢查點：預設為 **None**，若任務失敗 (貼標進度 = **FAILURE**) 紀錄貼標進行進度，例如執行 **2021** 整年貼標，任務在預測 **2021-03-18** 資料時失敗，因此檢查點紀錄 **2021-03-18**。
3. 錯誤訊息：預設為 **None**，若任務失敗 (貼標進度 = **FAILURE**) 紀錄錯誤原因。

其他使用須知1

1. 一次執行最佳任務數為 3，建議不要同時貼超過 5 個來源資料貼標任務。
2. 任務狀態刷新須重新整理網頁，有時可能需要等幾秒。
3. 查看抽樣結果會從後端資料庫抓取目標抽樣資料，可能會花幾秒時間。
4. 建議執行完任務紀錄任務ID或是不要修改與刪除完成之任務，可以在任務頁面左上角 操作 > 查看任務資訊 取得任務資訊列表。
5. 任務執行時間會根據來源資料量而定，約 10 min – 500 min
 - 2020 整年 男女作者貼標 wh_forum_02 約執行 180 min，wh_plurk 約執行 12 min
6. 時間配置錯誤 (例如起始時間比結束時間晚)，來源資料任務會直接顯示錯誤

任務ID	來源	貼標進度	任務狀態
0e0ee3265bd711eca43c04ea56825bad	wh_bbs_01	SUCCESS	finish



其他使用須知2

1. 貼標進度說明:

- PENDING=貼標執行中
- SUCCESS=貼標完成(待產出貼標率等資訊)
- FAILURE=貼標失敗
- BREAK=使用者中斷貼標任務

2. 任務狀態說明:

- None=任務未完成
- Finish=任務完成
- No_data=該來源範圍無資料