
Transfer Learning on Stack Exchange Tags

 [Kaggle link](#) 

Outline

- task definition
- dataset
- evaluation
- rules
- some tips

Task definition

- 給予文章的 title 與 content, 預測該文章的 tag

id	1
title	How do I install a new, non load bearing wall in my basement ?
content	<p><p>I'm looking to finish my basement and simply want to wall off the concrete blocks that make up the wall. I want to insulate it as well so that it becomes a usable family room when I'm done. How do I go about installing a wall given wooden beams in the ceiling and a concrete floor?</p></p>
tags	remodeling basement carpentry

Data set

- train set: cooking, crypto, robotics, biology, travel, diy
- test set: physics

cooking.csv	.zip (3.33 mb)
crypto.csv	.zip (3.64 mb)
robotics.csv	.zip (1.03 mb)
biology.csv	.zip (3.35 mb)
sample_submission.csv	.zip (210.38 kb)
travel.csv	.zip (4.67 mb)
diy.csv	.zip (7.18 mb)
test.csv	.zip (25.12 mb)

Evaluation

- 使用 Mean F1-Score

$$F1 = 2 \frac{p \cdot r}{p + r} \text{ where } p = \frac{tp}{tp + fp}, r = \frac{tp}{tp + fn}$$

- 例如正確的 tag 是 (a, b, c), 預測結果是 (a, e)
 $p = 1 / (1 + 1) = 0.5, r = 1 / (1 + 2) = 0.33, F1 = 2 / 5 = 0.4$

- 答案的格式

id,tags

1,physics poetry

2,physics poetry chemistry

3,physics electrons

Rules

- 跟隨該比賽的規則
- 一天可以上傳五次
- 不能夠手動標答案, 或是搜尋標準答案
- 不可以使用額外的資料

Some Tips

- 該使用哪些 feature? 除了常見的word2vec之外, 傳統nlp所會使用到的tf-idf、POS tagging、位置的資訊...都可以用看看, 效果可能會比較好
- 資料有很多的雜訊, 有html語法、數學公式、超連結...如何妥善地進行資料前處理, 是很重要的
- 有些 tags 不會出現在文章中, 該如何找出來可以思考看看
- 需要考慮phrase
- 該怎麼切出validation?

tags 沒有出現在文章中

id	1
title	How do I install a new, non load bearing wall in my basement?
content	<p>I'm looking to finish my basement and simply want to wall off the concrete blocks that make up the wall. I want to insulate it as well so that it becomes a usable family room when I'm done. How do I go about installing a wall given wooden beams in the ceiling and a concrete floor?</p>
tags	remodeling basement carpentry

需要考慮phrase

id	17
title	How can I hanging something on an exterior wall with vinyl siding ?
content	<p>My house has vinyl siding. If I want to hang something on an exterior wall, can I just attach it with screws drilled through the siding? Or do I need to do anything to prevent water from seeping in around the screws?</p>
tags	vinyl-siding

如果你完全想不到要用哪一種模型再往下看

試過的一種model

- feature: tf, idf, tf-idf, 在句子中的哪個位置, 是否出現在標題, 在這個主題中出現過幾次, 是否原本是大寫
- 濾掉stop word
- 當成二元分類問題, 使用gradient descent, 給一個字的feature, 判斷他是不是keyword