Adapt BP structural method to go from article -> article for ANY link type

Try sklearn imbalanced package in content baseline model to see if it improves credibility results (not just predicting majority class).

Look at histogram of coefficients to find a natural breaking point between liberal and conservative word informativeness.

Liberal words are saying more of what conservative words don’t say? (words less common in conservative media…or, liberal is default and conservatives use this dialect)?

Holdout Label Distribution: Counter({0: 11268, 1: 6937})

holdout test score: 0.824388904147

Vocabulary Length : 235729

Most informative liberal words :

('advertisement', -10.17163901812614)

('editing', -9.9218754515685212)

('advertisements', -8.2819278634411884)

('picture', -8.1236538818153683)

('getty', -7.071959115270162)

('rs', -5.7731143854716276)

-symbol for rupee the indian currency, also goes along with ‘crore’ since large amounts of money in india and Nepal are often written in terms of crores

-possibly, spearman correlation coefficient abbreviation

('reporting', -5.6185451286651924)

('image', -5.5441509808495688)

('euobserver', -4.9754045689145583)

('says', -4.9478341420640621)

('images', -4.7016897881407056)

('aug', -4.6377209278867113)

('montreal', -4.5583092270091372)

-canada

('chicago', -4.5451269988345331)

('abc', -4.5281050495741999)

-liberal sources site abc?

('pinterest', -4.1429485737272609)

('notifications', -4.0186398575376288)

('adds', -3.9943278769035859)

('bc', -3.9863856708901877)

-canada

('quebec', -3.8839046454831516)

-canada

Most informative conservative words :

('sputnik', 21.802664985580787)

-russian newspaper?

('dubai', 8.396253662106103)

('follow', 7.8489006158333474)

('afp', 7.6593911697427339)

-americans for prosperity

('daesh', 6.6779923896555466)

-arabic for ISIS

('al', 6.3054427978702154)

-alabama.com newspaper

('fortune', 5.511696671791591)

('colorado', 5.4366343110221917)

('more', 4.951891960990233)

('uae', 4.9421990912411662)

-united arab ermerits

('caller', 4.7161035875702328)

-daily caller magazine or news radio

('jiji', 4.5389878233400953)

('oilpricecom', 4.149536833154861)

('manhattan', 3.9183461965851234)

('ap', 3.9106206659794771)

('tlr', 3.8693285152300017)

-texas law review?

('wa', 3.8530491862409137)

-washington or Western Australia?

('moscow', 3.7712801715225797)

('blasio', 3.744493556679648)

('yomiuri', 3.5565356430563648)

pip install --user https://github.com/explosion/spacy-models/releases/download/en\_core\_web\_sm-1.2.0/en\_core\_web\_sm-1.2.0.tar.gz#egg=en\_core\_web\_sm-1.2.0