Experiments- LDA vs Lda2vec

Topic 0: government law gun public use states state right encryption rights
Topic 1: space data research program technology center nasa launch earth science
Topic 2: year team game season hockey win players league la vs
Topic 3: test st south posting miles fit search american west th
Topic 4: believe question point israel argument law different answer claim mean
Topic 5: problem sale offer problems condition keyboard driver best works work
Topic 6: information available list software package email university mail send computer
Topic 7: drive card disk scsi db mac hard video drives monitor
Topic 8: key chip bit window using keys use memory mhz chips
Topic 9: god jesus church christian religion christ world life bible gods
Topic 10: thanks email help post anybody looking advance hi bike send
Topic 11: car price buy ground cars work power pay cost wire
Topic 12: ago power game play water pts games hot baseball dave
Topic 13: maxaxaxaxaxaxaxaxaxaxaxaxaxaxaxax john dod according bob worked rd ask drives comes
Topic 14: stephanopoulos read cause yes evidence health person men medical course
Topic 15: true man example word truth bible david assume moral wrong
Topic 16: dos windows mb ms black os period ram pp appears
Topic 17: file use program files image server code set display using
Topic 18: good right better little sure make probably lot look bad
Topic 19: mr armenian war armenians turkish jews said went president children

Topic 0 jesus god orthodox faith verse worship belief christianity christian mormon
Topic 1 bike tire honda battery brake ride valve aluminum intake jeep
Topic 2 game team play win baseball score playoff mets goal phillies
Topic 3 shuttle nasa space payload computational planetary satellite mission launch lunar
Topic 4 armenian troop cyprus israel soldier agdam bayonet terror bullet he
Topic 5 drive controller battery dock connector ink card slot pin vram
Topic 6 key encrypt cryptosystem encryption cryptography secure rsa crypto eavesdrop decrypt
Topic 7 ride revolver bike bikers go jeep i dog a tank
Topic 8 homicide gun handgun firearm fbi unconstitutional federal smuggle drug weapon
Topic 9 israel armenian plo cyprus troop lebanese arab palestinian syria turkish
Topic 10 armenian bayonet arena team turkish tartar nhl playoff game defenseman
Topic 11 widget server contrib font editor format mit pixmap toolkit binary
Topic 12 controller card disk bios rom vram drive isa vga trident
Topic 13 cd shuttle sale frequency antenna disc space shipping offer model
Topic 14 game win play baseball score he team playoff goal mets
Topic 15 shuttle homicide space payload nasa launch mission gonorrhea rocket armenian
Topic 16 key cryptography devguide encrypt ripem encryption rsa eff cryptosystem crypto
Topic 17 jesus messiah verse orthodox god psalm apostle faith luke divine
Topic 18 quack blood lyme jesus infant orthodox substance god candida msg
Topic 19 font icon window server editor format disk card terminal i

Classification accuracy : 37%
Baseline:5%
Classification accuracy : 47.8%

Dummy example

```
texts[0]
```

u'Iphone ipad iPhone iphone4 Ipad apple_iphone iphones iPhone_3GS 3gs iPhone4 Iphone_3G htc android_phones droid_x ipod ipod_touch iphone_3g samsung 3GS iPad itouch mytouch tmo iPhone_3gs iPhone_3g MyTouch_4G iphone_3gs Apple_Iphone iPhone_3G appstore iphone_3G sony_ericsson iphone_3GS iphone_4g Motorola_Atrix_4G HTC_Droid_Incredible Apple_ipad Iphone_OS treo'

Out of vocabulary document - LDA

Raw text

Word movers distance

```
ldamodel[dictionary.doc2bow(test)]
```

```
[(0, 0.049999999999999989),
 (1, 0.049999999999999989),
 (2, 0.049999999999999989),
 (3, 0.049999999999999989),
 (4, 0.049999999999999989),
 (5, 0.049999999999999989),
 (6, 0.049999999999999989),
 (7, 0.049999999999999989),
 (8, 0.049999999999999989),
 (9, 0.049999999999999989),
 (10, 0.049999999999999989),
 (11, 0.049999999999999989),
 (12, 0.049999999999999989),
 (13, 0.049999999999999989),
 (14, 0.049999999999999989),
 (15, 0.049999999999999989),
 (16, 0.049999999999999989),
 (17, 0.049999999999999989),
 (18, 0.049999999999999989),
 (19, 0.049999999999999989)]
```

```
ldamodel[dictionary.doc2bow(old_to_new_words.values())]
```

```
[(0, 0.079884178031323957),
 (1, 0.020918364843171573),
 (2, 0.10753779636012975),
 (3, 0.02102270053600307),
 (4, 0.032369788495238631),
 (5, 0.10296038767407101),
 (6, 0.033368268092897509),
 (7, 0.085216772329985743),
 (8, 0.04344668736274937),
 (9, 0.051882563915244519),
 (10, 0.068147178644059794),
 (11, 0.029245712861690077),
 (12, 0.022246926497340464),
 (13, 0.036422945788688994),
 (15, 0.024532956434806821),
 (16, 0.044112822493258062),
 (17, 0.06890910722408887),
 (18, 0.039163518859255883),
 (19, 0.088101119472159729)]
```

Out of Vocabulary document - Lda2vec

```
texts[0]
```

u'Iphone ipad iPhone iphone4 Ipad apple_iphone iphones iPhone_3GS 3gs iPhone4 Iphone_3G htc android_phones droid_x ip od ipod_touch iphone_3g samsung 3GS iPad itouch mytouch tmo iPhone_3gs iPhone_3g MyTouch_4G iphone_3gs Apple_Iphone i Phone_3G appstore iphone_3G sony_ericsson iphone_3GS iphone_4g Motorola_Atrix_4G HTC_Droid_Incredible Apple_ipad Ipho ne_OS treo'

Top words in topic 0 out_of_vocabulary government militia law nsa citizens crypto encryption laws privacy
Top words in topic 1 out_of_vocabulary chip apple serial scsi chips device buy board phones
Top words in topic 2 out_of_vocabulary board scsi apple card drive mhz meg quadra bike
Top words in topic 3 out_of_vocabulary fired city team hawks teams car shot oil hockey
Top words in topic 4 out_of_vocabulary window cursor <SKIP> mouse i screen bitmap application void
Top words in topic 5 out_of_vocabulary hell i god eternal jesus satan christians think just
Top words in topic 6 out_of_vocabulary god jesus eternal christ lord sin christians hell sabbath
Top words in topic 7 out_of_vocabulary scsi apple quadra drive meg card board vlb <SKIP>
Top words in topic 8 out_of_vocabulary armenian soldiers azerbaijan azerbaijani city azeri turkish armenians turks
Top words in topic 9 out_of_vocabulary <SKIP> window widget xt file bitmap motif application windows
Top words in topic 10 crypto nsa rsa privacy encryption cryptography administration eff clipper patent
Top words in topic 11 out_of_vocabulary patients hiv disease clinical # law encryption sexual persons
Top words in topic 12 out_of_vocabulary planetary space nasa propulsion satellite shuttle lunar spacecraft available
Top words in topic 13 out_of_vocabulary he hell said mamma i sad koresh sumgait lord
Top words in topic 14 out_of_vocabulary planetary nasa distributed <SKIP> propulsion information available privacy sp ace
Top words in topic 15 out_of_vocabulary jews arab nazi jewish israeli greek occupied israel turkish
Top words in topic 16 out_of_vocabulary god arab jews christians jewish jesus christianity israel christian
Top words in topic 17 planetary out_of_vocabulary lunar division spacecraft launched satellite nasa missions team

```
plt.bar(np.arange(20), dat['doc_topic_dists'][0])
```

<Container object of 20 artists>