gensim lda 결과 topic 별 기여도가 점수화 되는데 이를 모두 합하면 1이 될까요?

제가 업로드한 gensim을 이용한 lda 토픽 모델링의 코드 중 아래와 같은 결과가 있습니다.

```
import gensim
NUM_TOPICS = 20 # 20개의 토픽, k=20
ldamodel = gensim.models.ldamodel.LdaModel(corpus, num_topics = NUM_TOPICS, id2word=dictionary, passes=15)
#passes : 알고리즘의 동작횟수 - 적절히 지정해주면되는 하이퍼 파라미터
topics = ldamodel.print_topics(num_words=4)
#num_words : 출력하고 싶은 단어의 갯수 지정
for topic in topics:
    print(topic)
```

```
(0, '0.009*"smokeless" + 0.006*"sweden" + 0.005*"adobe" + 0.004*"latin"')
(1, '0.010*"nrhj" + 0.007*"wwiz" + 0.006*"bxom" + 0.006*"gizw"')
(2, '0.026*"jesus" + 0.012*"bible" + 0.010*"christ" + 0.008*"word"')
(3, '0.018*"printf" + 0.014*"color" + 0.014*"scsi" + 0.009*"widgets"')
(4, '0.016*"government" + 0.015*"president" + 0.011*"state" + 0.007*"states"')
(5, '0.012*"armenian" + 0.011*"israel" + 0.010*"jews" + 0.009*"armenians"')
(6, '0.013*"sale" + 0.010*"condition" + 0.010*"offer" + 0.010*"shipping"')
(7, '0.012*"drive" + 0.012*"windows" + 0.011*"thanks" + 0.011*"system"')
(8, '0.009*"said" + 0.007*"people" + 0.007*"time" + 0.007*"know"')
(9, '0.016*"guns" + 0.010*"crime" + 0.007*"control" + 0.007*"henrik"')
(10, '0.017*"space" + 0.007*"nasa" + 0.006*"research" + 0.005*"university"')
(11, '0.006*"islam" + 0.006*"john" + 0.005*"filename" + 0.005*"water"')
(12, '0.008*"bike" + 0.008*"like" + 0.007*"good" + 0.007*"much"')
(13, '0.011*"health" + 0.008*"medical" + 0.006*"administration" + 0.006*"disease"')
(14, '0.014*"encryption" + 0.012*"chip" + 0.011*"keys" + 0.010*"clipper"')
(15, '0.015*"church" + 0.009*"atheism" + 0.008*"atheists" + 0.007*"existence"')
(16, '0.016*"team" + 0.016*"game" + 0.011*"year" + 0.011*"play"')
(17, '0.018*"file" + 0.011*"program" + 0.008*"files" + 0.008*"window"')
(18, '0.008*"like" + 0.007*"would" + 0.007*"power" + 0.006*"used"')
(19, '0.019*"would" + 0.013*"people" + 0.011*"think" + 0.008*"know"')
```

- 결과 해석
- (0, '0.009*"smokeless" + 0.006*"sweden" + 0.005*"adobe" + 0.004*"latin"')
- 0 : 특정 토픽을 의미 우리는 모두 20개의 토픽으로 설정하였으므로 0~19의 수치를 가짐
- 해당 단어 앞에 붙은 숫자는 해당 토픽에서의 기여도를 의미하고 있습니다.

  → 정말 무수히 많은 단어가 문서들 안에 있으니 그 중에서도 높은 순서로 최대 10개까지 단어기여도 표시가 가능합니다.

  print(ldamodel.print_topics()) : 해당 명령어로 토픽당 10개까지의 기여도 표현이 가능
- 따라서 정말 모든 단어의 기여도를 표현한다면 그 총합이 1이겠지만 그건 수치상으로 표현이 거의 불가능하므로 특정 토픽에 대한 단어 기여도를 보여준다고 이해하시면 될

것 같아요🙂

p.s 코드는 깃허브(inventory203) 참고해주세요