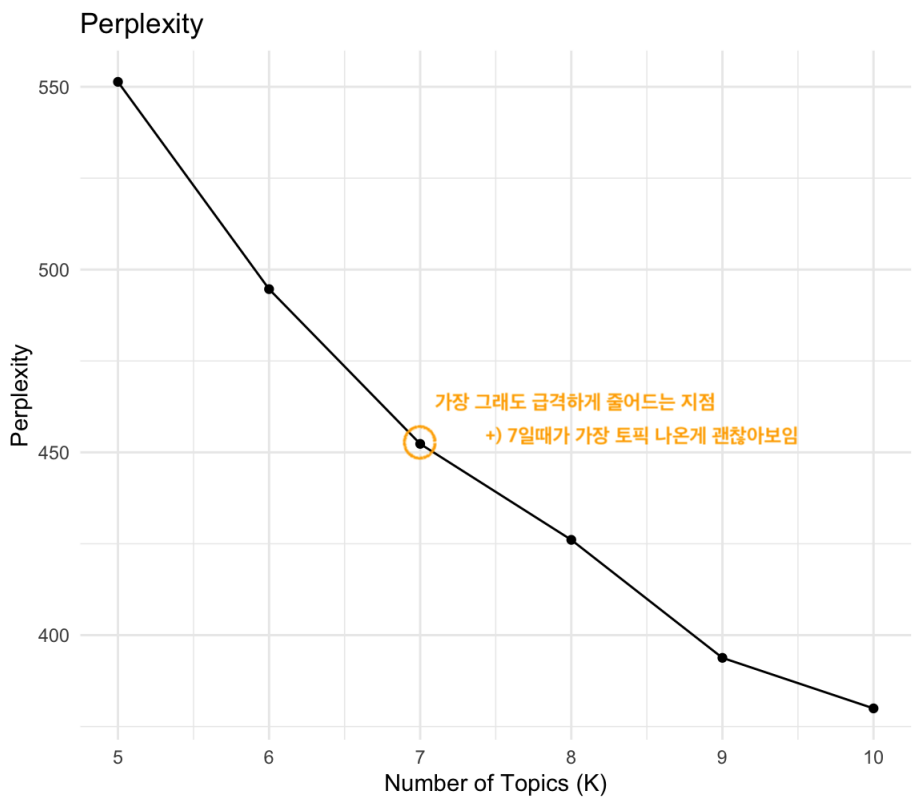
LDA

- 최종 엑셀 데이터 수 : **1527**
- (불용어 생략)

**LDA 모델 파라미터

```
lda_model <- LDA(dtmLDA, k = 7, method = "Gibbs", control = list(
  seed = 1234,
  alpha = 0.1,  # alpha 값을 0.1로 설정
  delta = 0.05,   # eta 값을 0.05로 설정 (토픽이 적은 수의 단어에 집중)
  iter = 5000,  # 반복 횟수를 5000으로 설정
  burnin = 1000,  # burn-in 기간을 1000으로 설정
  thin = 500    # 샘플링 얇게 할 빈도 설정
))
```




```
      Topic 1         Topic 2     Topic 3         Topic 4        Topic 5         Topic 6      Topic 7
 [1,] "sensor"        "location"  "network"       "network"      "tag"           "vessel"     "image"
 [2,] "cargo"         "transport" "communication" "route"        "rfid"          "display"    "trailer"
 [3,] "monitor"       "track"     "wireless"      "machine"      "lock"          "monitor"    "scan"
 [4,] "security"      "route"     "terminal"      "schedule"     "seal"          "ais"        "crane"
 [5,] "radiation"     "freight"   "radio"         "map"          "track"         "drone"      "parcel"
 [6,] "temperature"   "logistic"  "track"         "optimization" "radio"         "beacon"     "transport"
 [7,] "transport"     "robot"     "transceiver"   "cluster"      "transport"     "radar"      "scanner"
 [8,] "smart"         "asset"     "monitor"       "marketplace"  "smart"         "automatic"  "xray"
 [9,] "bulk"          "gps"       "location"      "preference"   "communication" "underwater" "laser"
[10,] "communication" "monitor"   "satellite"     "search"       "battery"       "onboard"    "cargo"
```

이 정도 Topic이면 괜찮을까요

asset / underwater / battery / marketplace

토픽 보고 관련 특허보고 Topic 정의했습니다.(우선, 교수님이 체크해주신 특허 위주로 먼저 보고 뽑았습니다. -> 추후에 나머지 특허도 찾아볼 예정)
전체적인 시스템
Topic 1 : 스마트 화물 관리 및 보안 시스템
- Cooling system and methods for cooling interior volumes of cargo trailers (sensor, cargo, monitor, temperature, transport)
- Shipping container monitoring based on door status (sensor, cargo, monitor, transport, communication)
- System and method to enhance security of shipping containers (radiation, bulk)
- Identification of material type and condition in a dry bulk material storage bin (monitor, security)

Topic 2 : 스마트 물류 추적 및 자산 관리 시스템
- Information transmission and processing systems and methods for freight carriers (location, transport , monitor, GPS)
- Cellular global tracker for freight (location, track, freight, GPS)
- *~~Marine asset security and tracking (MAST) system* (location, transport, track, asset)

Topic 3 : (무슨 기술을 활용한) 무선 네트워크를 통한 소통?
- Method for tracking containers using a low-rate wireless personal area network system (network, communication, radio, track, location)
- Maritime machine-to-machine (M2M) application priority management (network, communication, radio, satellite)
- Method, system, and device for managing network entry and BTS handover in land-to-sea wireless networks (network, wireless)
- Method for transmission of target information over a network (network, track, monitor)

Topic 4 : 경로 및 스케줄 최적화(구체적인 기술)
- System and method for coordinating transport of cargo (schedule, map, optimization)
- Vector-based harbor scheduling (gps, schedule, map, optimization)

Topic 5 : RFID-Tag를 이용한 추적 및 모니터링 시스템(1- 5 사실상 합쳐지는게 맞다 그래서 사실 상 토픽 수를 줄여봐도 될 것 같다. 나눈다고 한다면) 지상용(육상용)
- Temperature tracking and monitoring system used for commodities transportation (tag, rfid, track, transport, communication)
- Tamper evident cargo container seal bolt lock (rfid, lock, seal, battery)
- System and method for transport security control and tracking (tag, rfid, seal, track, transport, communication)

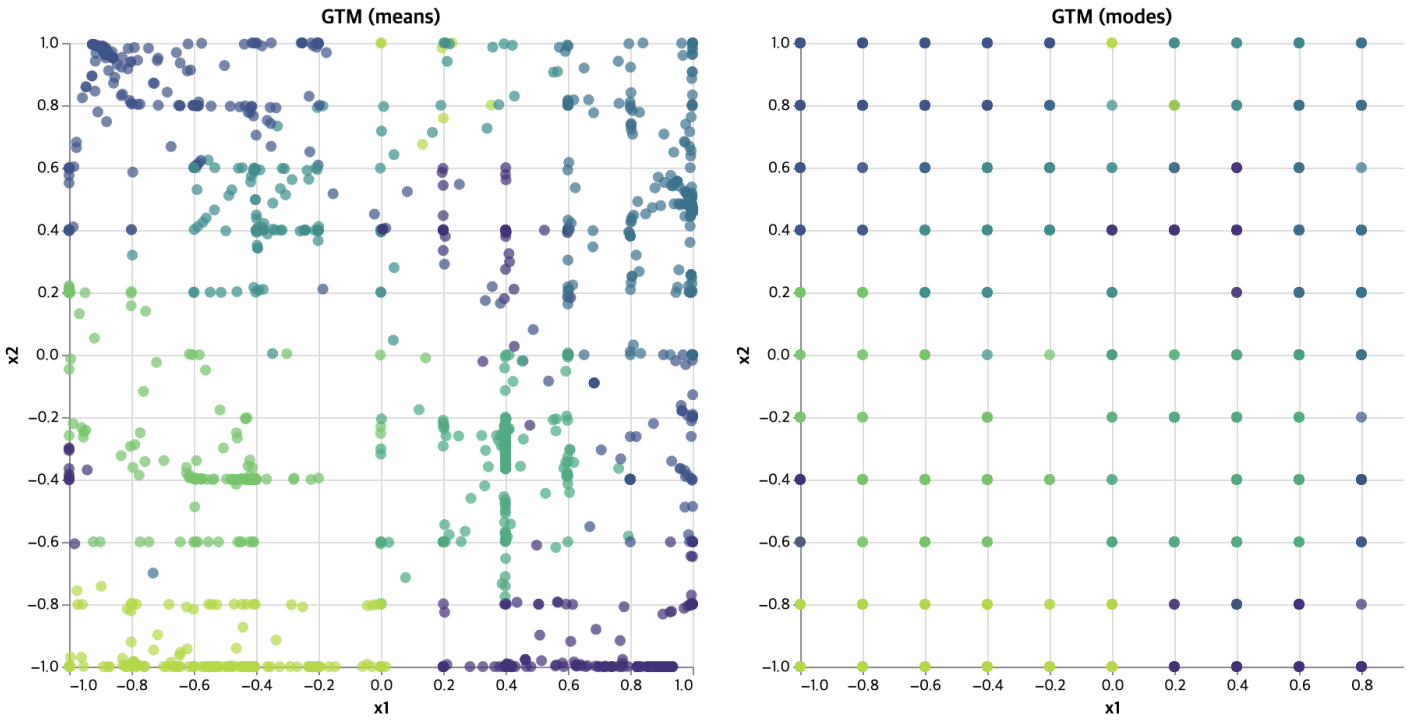
Topic 6 : 드론, 비콘, AIS를 활용한 모니터링? (AIS = 수중 GPS)
- System and method for extending GPS to divers and underwater vehicles (beacon, underwater)
- System and method for transmitting and receiving radar data using the automatic identification system (display, ais, radar)
- Data retrieval and transmitting marine exploration vessel systems (vessel, monitor, drone)

Topic 7 : X-ray를 활용한 화물 이미지 인식
- X-ray inspection system that integrates manifest data with imaging/detection processing (image, scanning, x-ray, cargo)
- Cargo container X-ray apparatus (image, crane transport, x-ray, cargo)

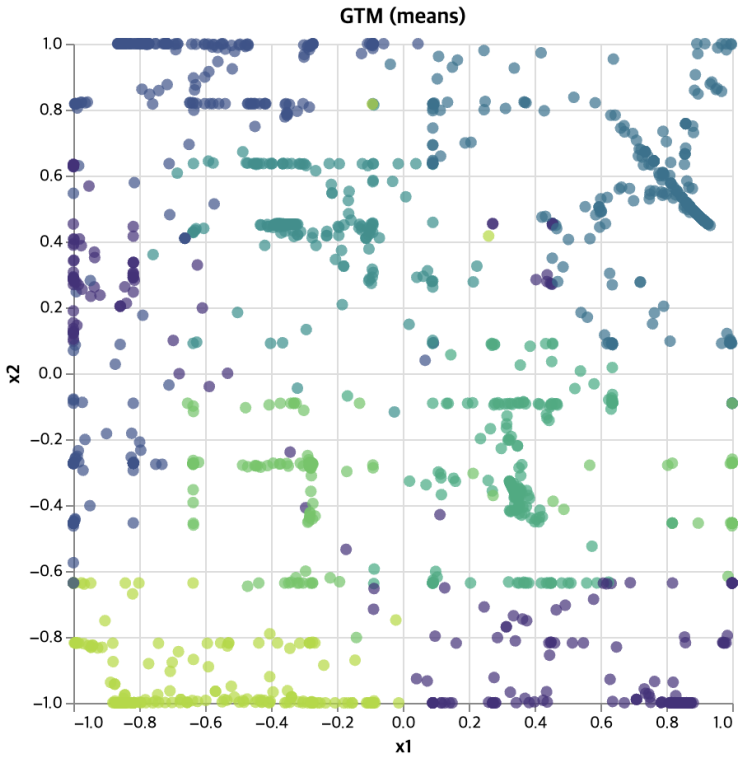
GTM
k 값(차원 수) : 11

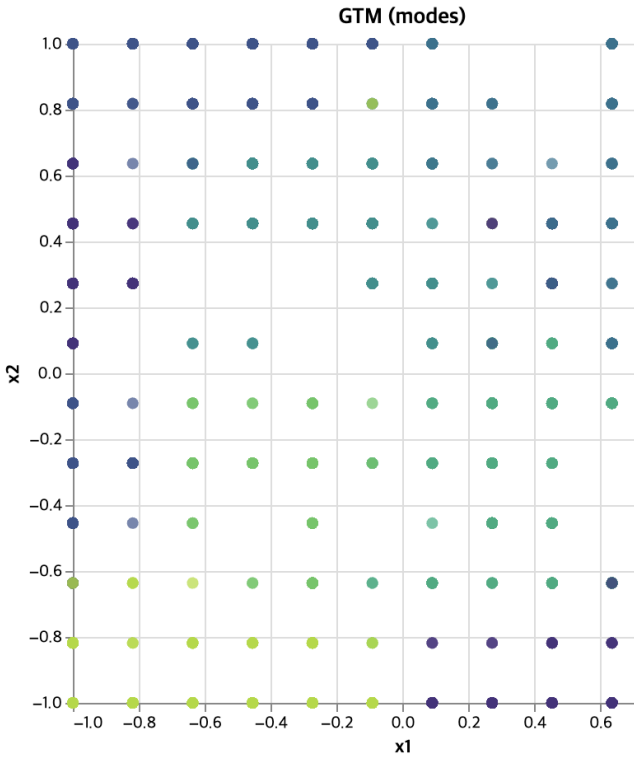
- 토픽이 7개라 차원 수도 비슷하게 7이나 +- 1로 해봤는데 10 이하으로는 modes 차트가 꽉 차서 공백이 안나왔어요.
- 토픽 안에 들어있는 주제들이 세부적으로 나누면 토픽 당 1-2개씩 있는 거 같아서 차원수는 좀 더 늘린 11로 정했습니다.

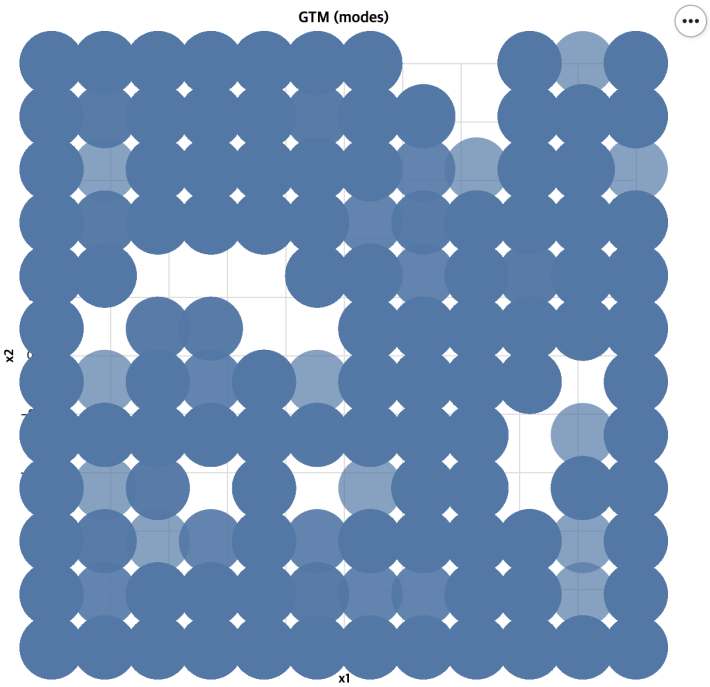
k = 11



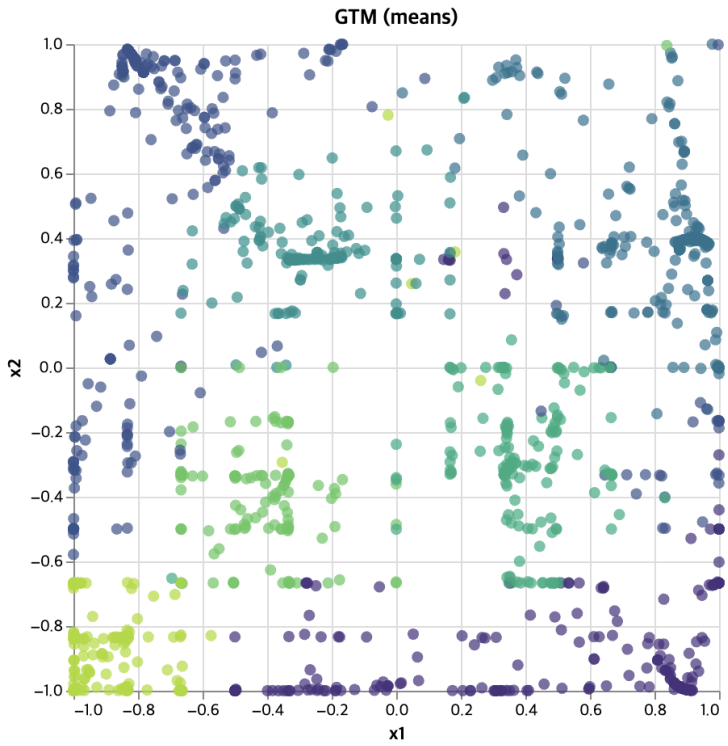
k = 12**

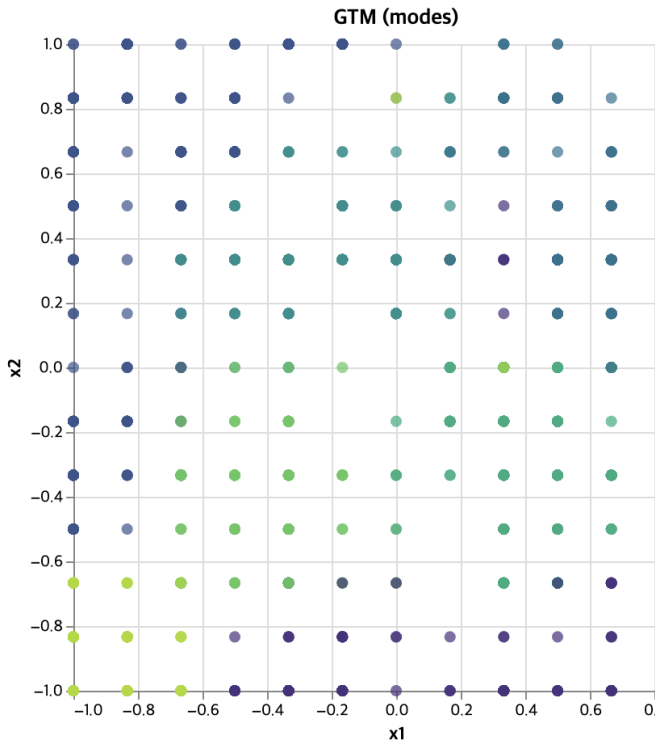

GTM (means)
x1
x2

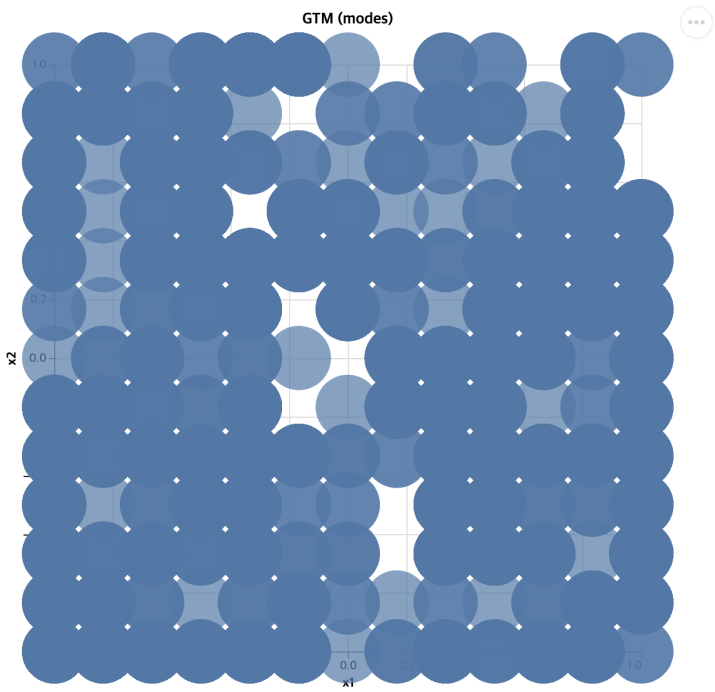

GTM (modes)
x1
x2


GTM (modes)
x2
x1

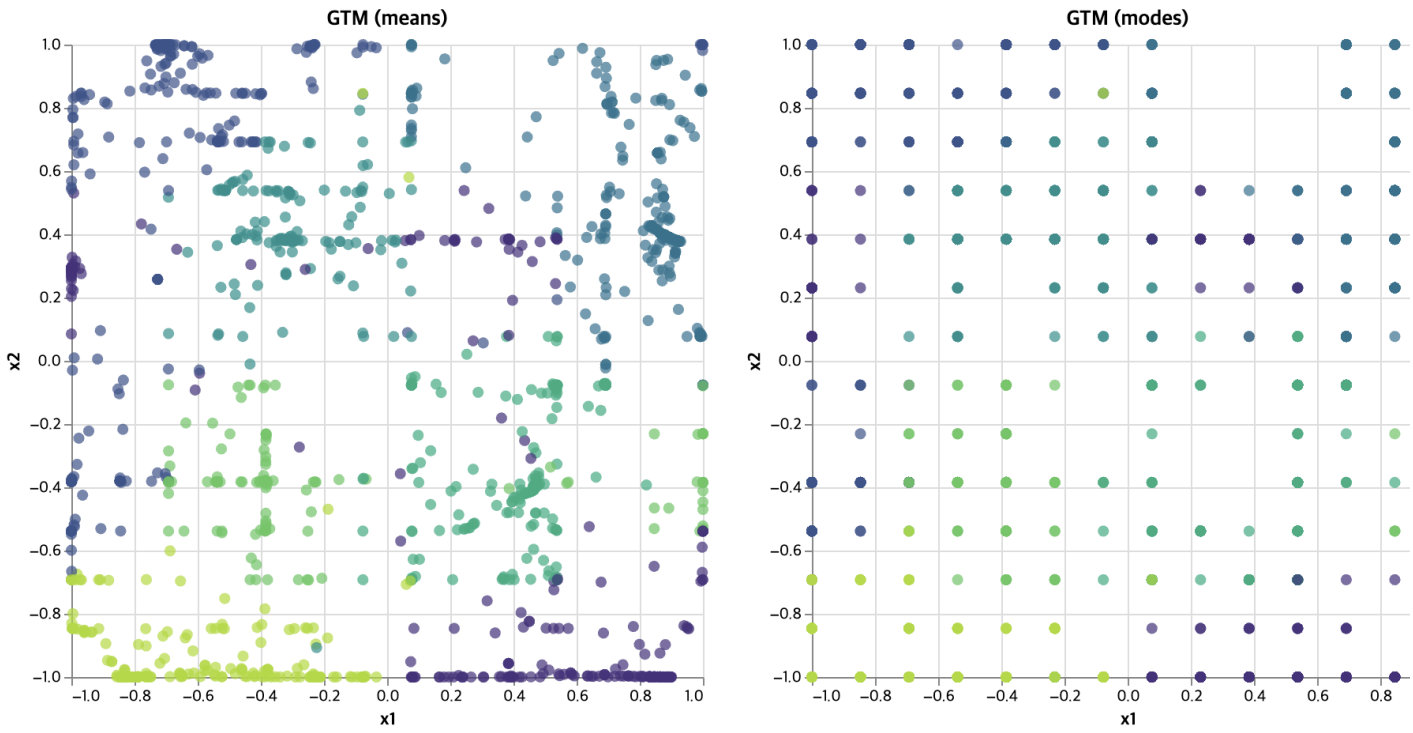
k = 13







k = 14

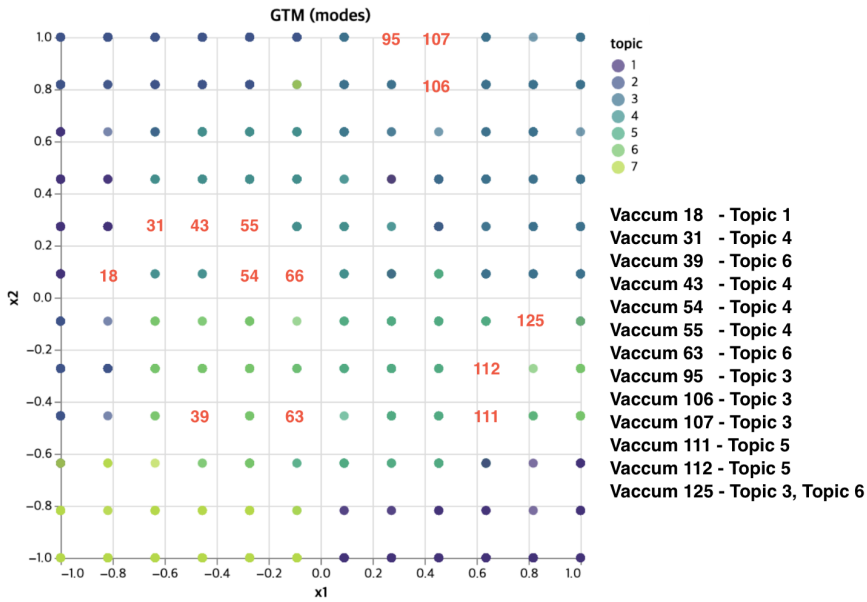


- k = 11일때는, 공백이 잘드러나지 않는 것 같아 제외 시켰고,
- k = 12, 13, 14는 공백 위치가 비슷하게 잘(?) 형성된거 같아서 **k값을 12랑 13** 두 개로 진행해봤습니다.
  GTM 다른 파라미터들 있던데 (기저 함수, 폭 파라미터, 가중치 정규화 인자 등)은 조절 할 필요있는지 궁금??

공백 영역 식별

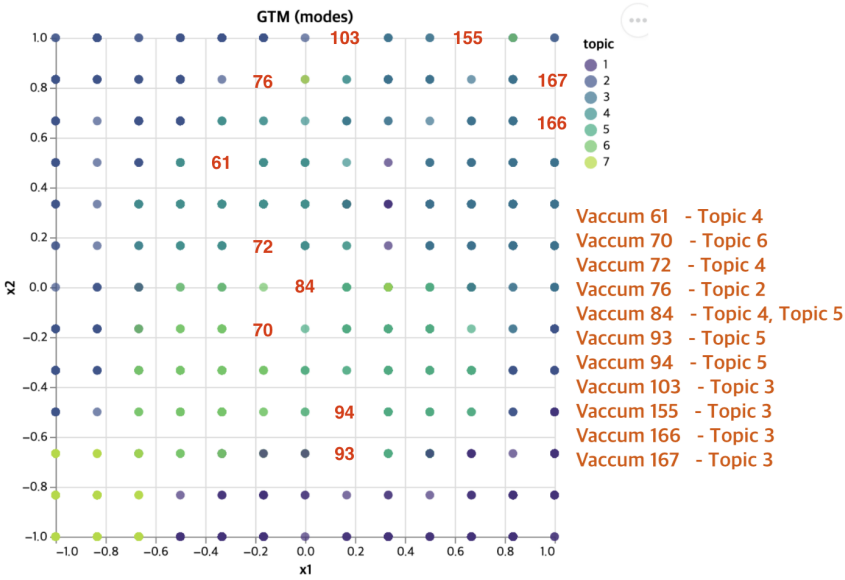
k = 12 -> 144

	1	2	3	4	5	6	7
18	0.420891193	0.122845397	0.075489567	0.226306865	0.033981907	0.135571676	-0.015084153
31	0.27975799	-0.072787107	0.040838799	0.64544499	0.005208847	0.102833632	-0.001293846
39	0.014664263	0.138194399	0.031577212	-0.075733128	-0.065926465	0.802369203	0.154860917
43	0.050139161	-0.113758115	0.000756021	0.92803447	0.000892115	0.132123061	0.001811936
54	-0.0221935	-0.091852444	-0.016689714	0.731117993	0.01933826	0.380639445	-0.000367864
55	-0.035757649	-0.099545958	-0.015679582	0.984803474	0.01709589	0.131357188	0.017721488
63	0.03504841	0.011921411	0.019921861	-0.06951813	0.319656156	0.575639896	0.107328795
66	0.020865052	-0.027374427	-0.022414581	0.581663221	0.147650727	0.271224753	0.028374796
95	-0.015995671	-0.086400261	0.878200393	0.079751619	0.060393472	0.016586576	0.067455739
106	0.045348229	-0.107358251	0.876635775	0.083519968	0.032214998	0.024891866	0.04475494
107	-0.022440867	-0.113902558	0.932990144	0.067362938	0.06003697	0.020338391	0.055609118
111	0.121381695	0.009325403	0.038896495	0.035136005	0.603017664	0.187048178	0.005203424
112	-0.012529406	0.030838998	0.112523222	0.026835347	0.638293292	0.19290266	0.011140973
125	-0.010304948	0.010256753	0.383963435	0.019832008	0.229885542	0.350445056	0.015925078



k = 13 -> 169

	1	2	3	4	5	6	7
61	0.011352544	0.177096467	0.027924239	0.800422715	0.020235343	-0.048865497	0.011837522
70	-0.035601318	-0.055407916	0.004618974	0.196308231	0.189355144	0.668219068	0.032500458
72	0.043400544	-0.06697508	0.009417818	0.711013337	0.067189665	0.183662461	0.052285096
76	0.008946718	0.493559089	0.161892025	0.282529036	0.019451115	0.010783137	0.0228412
84	0.051084328	-0.042326096	-0.003294578	0.317497858	0.310387986	0.284375678	0.082264888
93	0.309737365	-0.026672145	0.082079	0.012901511	0.472381876	0.14482561	0.004748403
94	0.097336998	-0.038396876	0.019961448	-0.000242331	0.710835936	0.198909935	0.011593588
103	-0.003379067	0.193249787	0.459057702	0.328346666	0.036158029	-0.029676922	0.016234332
155	0.03682045	0.142522888	0.475201414	0.037542746	0.019038713	0.257561221	0.031304419
166	0.017627887	0.062225571	0.765040894	-0.024882555	0.009408352	0.159519149	0.011056064
167	0.040415347	0.142585873	0.570573676	-0.048110429	0.009240058	0.256679911	0.028607917



k = 12, 13일때 모두 공백 기술로 지목된 토픽은 얼추 비슷하게 나온 듯 합니다. (Topic 3,4,5가 좀 공백,,?)

=> 이제 결론을 내야하는데, 1차 발표때처럼 **주변 공백 묶어서 결론을 내면 되는지? 아님 같은 토픽끼리??** 3, 4 , 5 토픽 기술이 부족하다...?
=> 토픽마다 좀 개성이 생겨서 묶을 필요는 없을 것 같기도 하고

참고한 논문에서는 토픽이 아닌 키워드를 추출해서 일정 임계점 이상인 키워드들을 모아서 기술 공백 키워드라고 설명하던데
==토픽이니 하나만 택해서 설명하면 되는지? ==

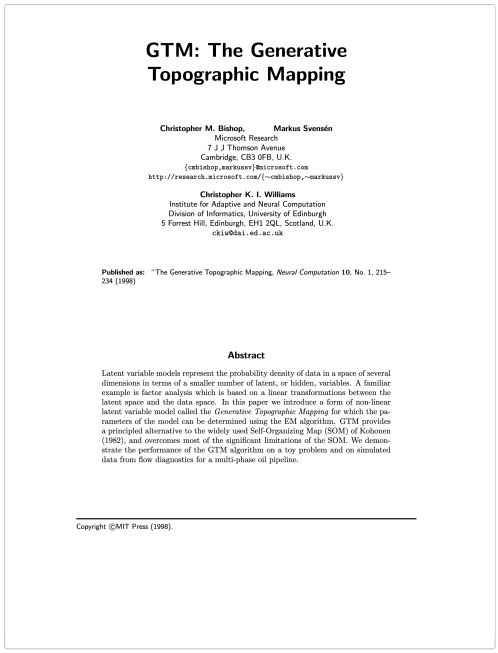
그리고, 이렇게 GTM 통해서 뽑을 수 있으니, 굳이 신경망이나, 랜덤포레스트, GPR 등등 할 필요없어졌는지

Table 5
The final result of vacuum interpretation.

Patent vacuum no.	Keywords
1	Apparatus, exposure, lens
2	Apparatus, exposure, alignment
3	Apparatus, exposure, radiation, reticle, wavelength, alignment
4	Apparatus, exposure, radiation, reticle, laser, axis, calibration,
5	Apparatus, exposure, reticle, axis, wavelength, lens
6	Apparatus, exposure, reticle, laser, axis, path, wavelength, lens, alignment
7	Exposure, radiation, laser, axis, path, wavelength, grid, lens, frequency
8	Apparatus, exposure, reticle, axis, path, temperature, calibration, numerical aperture, grid, dose, sigma, alignment
9	Apparatus, exposure, reticle, axis, path, temperature, interferometer, calibration, dose, lens, sigma, alignment
10	Immersion, apparatus, exposure, radiation, reticle, laser, axis, wavelength, temperature, interferometer, lens, aberration, refraction, curvature
11	Apparatus, exposure, reticle, laser, axis, path, wavelength, temperature, modulator, calibration, numerical aperture, maskless lithography, dose, sigma, alignment, pulse
12	Apparatus, exposure, reticle, laser, axis, path, wavelength, modulator, deflection, maskless lithography, dose, lens, frequency, modulation, polarization, alignment, pulse
13	Apparatus, exposure, radiation, reticle, euv, laser, axis, path, wavelength, temperature, interferometer, calibration, detector, lens, frequency, refraction, polarization, alignment

---

그리고 GTM 논문에 수식이 들어가야할 것 같은데, 어떻게 넣어야할지 고민
=> GTM 원문 보고 이해해서 저희 말로 정리해서 논문에 넣어야되나?
=> 다른 GTM 논문 참고해서 비슷하게 하면 표절일까요,,,?



GTM: The Generative Topographic Mapping

Christopher M. Bishop, **Markus Svensén**
Microsoft Research
7 J J Thomson Avenue
Cambridge, CB3 0FB, U.K.
{cmbishop,markusv}@microsoft.com
http://research.microsoft.com/{~cmbishop,~markusv}

Christopher K. I. Williams
Institute for Adaptive and Neural Computation
Division of Informatics, University of Edinburgh
5 Forrest Hill, Edinburgh, EH1 2QL, Scotland, U.K.
ckiw@dai.ed.ac.uk

Published as: "The Generative Topographic Mapping, *Neural Computation* **10**, No. 1, 215–234 (1998)

Abstract

Latent variable models represent the probability density of data in a space of several dimensions in terms of a smaller number of latent, or hidden, variables. A familiar example is factor analysis which is based on a linear transformations between the latent space and the data space. In this paper we introduce a form of non-linear latent variable model called the *Generative Topographic Mapping* for which the parameters of the model can be determined using the EM algorithm. GTM provides a principled alternative to the widely used Self-Organizing Map (SOM) of Kohonen (1982), and overcomes most of the significant limitations of the SOM. We demonstrate the performance of the GTM algorithm on a toy problem and on simulated data from flow diagnostics for a multi-phase oil pipeline.

Copyright ©MIT Press (1998).

Expert Systems with Applications 39 (2012) 2489–2500

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Development of a GTM-based patent map for identifying patent vacuums

Changho Son, Yongyoon Suh, Jeonghwan Jeon, Yongtae Park *

Department of Industrial Engineering, School of Engineering, Seoul National University, San 56-1, Sillim-Dong, Kwanak-Gu, Seoul 151-742, Republic of Korea

ARTICLE INFO

Keywords:
GTM
Patent map
Patent vacuum
Keyword vector

ABSTRACT

The patent map has long been considered as a useful tool for mining latent technological information. Among others, the detection of patent vacuums, defined as unexplored areas of new technologies, deserves intensive research. However, previous studies for identifying patent vacuums on the patent map have been subjected to some limitations, stemming from the subjective and manual identification of patent vacuums. To address these limitations, this paper proposes a generative topographic mapping (GTM)-based patent map, which aims to automatically identify a patent vacuum. Since GTM is a probabilistic approach of mapping multidimensional data space onto a low-dimensional latent space and vice versa, it contributes to the automatic detection and interpretation of patent vacuums. The proposed approach consists of three stages. Firstly, text mining is executed in order to transform patent documents into keyword vectors as structured data. Secondly, the GTM is employed to develop the patent map, subsequently leading to the discovery of patent vacuums, which are expressed as blank areas in the map. Lastly, the meaning of each patent vacuum is interpreted by the inverse mapping of patent vacuums onto the original keyword vector. The case study is conducted with lithography technology-related patents. We believe the proposed approach not only saves time and effort for identifying patent vacuums, but also increases objectivity and reliability.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

A patent map has been widely used for identifying the possibilities and opportunities for new technology (Grandstrand, 1999). Since patents are useful sources of knowledge about technical progress and innovative activity (Basberg, 1987; Ernst, 2003; Griliches, 1990; Jaffe, Trajtenberg, & Fogarty, 2000; Li, Wang, & Hong, 2009), the patent map is a guaranteed useful proxy measure for technological power (Park, Yoon, & Lee, 2005), has been employed as the representative tool used to grasp diverse features of individual patents and identify complex relationships among patents (Yoon, Yoon, & Park, 2002). Since patent maps are presented in visual forms such as charts, tables, or graphs, significant amounts of technological information can be acquired in informative and easy ways. More importantly, patent maps have been employed to identify patent vacuums, which are regarded as an unexplored area of technology that deserves intensive investigation for new technology development. In previous studies, two representative types of patent maps have been used for identifying patent vacuums: a principle component analysis (PCA)-based patent map (Lee, Yoon, & Park, 2009) and a self-organizing map (SOM)-based patent map (Yoon et al., 2002).

However, two significant limitations exist in both types of patent maps. The first limitation originates from detecting patent vacuums from the patent map. In previous studies, patent vacuums in the patent map have been detected by the subjective ways, depending on the knowledge and experience of researchers. Since there is no clear standard for detecting vacuums, they have been characterized as the relatively sparse or empty areas in the patent map. Thus, no alternative exists, and patents vacuums must be identified in this work by the subjective judgment of researchers. Consequently, patent vacuums might be detected differently depending on each researcher's knowledge and experience, even in a single patent map.

The second limitation constricting previous patent maps corresponds to the interpretation of identified vacuums. After detecting patent vacuums, the vacuum should be interpreted as a real-world technological opportunity, which is as a key part of patent vacuum mapping. However, the interpretation has relied on manual work by researchers, such as investigating the surrounding patents of target vacuums. Therefore, the interpretation of patent vacuums possesses an inevitable weakness in regards to efficiency and effectiveness. In terms of efficiency, an ample amount of time and effort must be devoted to interpreting the patent vacuum as real-world technology. In terms of effectiveness, quite naturally, a significant subjectivity problem arises since the interpretations vary depending on the knowledge and experience of researchers.

* Corresponding author. Tel.: +82 2 880 8358; fax: +82 2 889 8560.
E-mail addresses: c130810@snu.ac.kr (C. Son), yoo2000@snu.ac.kr (Y. Suh), hwan63@snu.ac.kr (J. Jeon), parkyt@cybernet.snu.ac.kr (Y. Park).

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.06.101

BERTopic

이번에 데이터 1500여개로 줄이면서 데이터 수가 줄면 줄수록 뭔가 토픽이 제대로 안나오는 느낌...

```python
# 임베딩 모델
embedding_model = SentenceTransformer("all-MiniLM-L6-v2")

# UMAP 모델
umap_model = umap.UMAP(
    n_neighbors=15,  # 각 데이터 포인트가 고려하는 이웃의 수.
    n_components=2,  # 차원 축소 후의 출력 차원 수. 2로 설정하여 2차원 시각화
    min_dist=0.05,   # UMAP이 데이터를 끌어당기는 정도. 낮을수록 데이터 포인트가 더 가깝게 위치
    spread=1.0       # UMAP이 전체 데이터의 글로벌 구조를 유지하는 정도. 값이 높을수록 전체 구조가 퍼지게
)

# HDBSCAN 모델
hdbscan_model = hdbscan.HDBSCAN(
    min_cluster_size=5,  # 클러스터의 최소 크기. 너무 작은 클러스터를 방지
    prediction_data=True, # 클러스터 예측 데이터를 저장할지 여부.
    min_samples=5,       # 각 데이터 포인트가 핵심 포인트로 간주되기 위한 최소 샘플 수.
    alpha=1.0            # 클러스터 밀도를 계산할 때 사용할 거리 척도.
)

# CountVectorizer 설정
vectorizer_model = CountVectorizer(
    stop_words=stop_words,  # 불용어 목록.
    ngram_range=(1, 1)     # n-그램의 범위. 바이그램 (1, 2)로도 해봤는데 딱히...
)

# BERTopic 모델 초기화 및 훈련
topic_model = BERTopic(
    language="english",
    calculate_probabilities=True,
    # nr_topics=7,                      # 주제의 수 제한. 모델이 최대 7개의 주제를 찾도록 제한
    top_n_words=10,                   # 각 주제의 상위 단어 수.
    min_topic_size=5,                 # 주제의 최소 크기.
    vectorizer_model=vectorizer_model,  # 벡터화 모델. CountVectorizer를 사용하여 텍스트를 벡터화
    embedding_model=embedding_model,  # 임베딩 모델. 문장을 벡터로 변환하는 모델.
    umap_model=umap_model,            # UMAP 모델. 차원 축소를 수행하는 모델.
    hdbscan_model=hdbscan_model,      # HDBSCAN 모델. 클러스터링을 수행하는 모델.
    ctfidf_model=None,                # c-TFIDF 모델. 기본 설정.
    verbose=True
)
```

BERTopic 우선 결과가 너무 안좋은거 같습니다.

- BERTopic이 문맥을 이해하니깐 불용어로 표제어 / 조사 이런걸 다빼버려서 그런가 싶어서 텍스트 전처리 과정을 생략하고도 돌려봤는데
- 데이터 수가 1500개라 그런지 모든 토픽에 불용어가 껴버려서 쓸 수가 없었습니다.(to, the, and, ~)
- 그래서 불용어는 제거하구 파라미터만 할 수 있는대로 이리저리 다 바꿔봤습니다.
  `nr_topics` 제한 없이 10개 언저리 정도 토픽 나오도록 파라미터 다 꺼내서 건들여보고, 민감도 분석도 해봤는데
  토픽의 질이 별로 좋지 않아보이고, 파라미터를 미세하게 조절하는데도 불구하고 **극과 극으로 결과가 나옴**
  아래 처럼 토픽이 제대로 안나오던가 or 토픽이 너무 많이 나오던가

```
   Topic  Count                                       Name  \
0     -1     23                     -1_tag_line_system_storage
1      0   1447                0_system_container_include_data
2      1     38                       1_rfid_tag_reader_device
3      2     19  2_refrigeration_temperature_control_container


Topic -1: tag line display system laser label traction intrinsic support hawser
Topic 0: system container include data device method information least network vehicle
Topic 1: rfid tag reader device antenna system include data identification radio
Topic 2: refrigeration temperature control container system power sensor reefer controller air
```

```
    Topic  Count                                Name  \
0      -1    506   -1_system_container_method_network
1       0     55      0_marine_vessel_underwater_ship
2       1     36      1_portion_conveyor_rack_section
3       2     35            2_rfid_tag_reader_antenna
4       3     35  3_transaction_marketplace_good_sale
..    ...    ...                                  ...
64     63      6      63_line_drilling_traction_float
65     64      5  64_semantic_deep_profile_enterprise
66     65      5     65_time_base_receiver_particular
67     66      5           66_tag_laser_contract_grid
68     67      5          67_tilt_feedback_dolly_scan
```

```
Topic -1: system container method network device item data include use information
Topic 0: marine vessel underwater ship ais information radio unmanned transmit subsea
Topic 1: portion conveyor rack section wall bearing end storage support ring
Topic 2: rfid tag reader antenna nfc read chip identification device radio
Topic 3: transaction marketplace good sale custom purchase company information product user
Topic 4: container solvent track communication support wpan locate error handle device
Topic 5: track location gps component driver telematics object event voice vehicle
Topic 6: vessel sail ship position marine rudder hull speed control route
Topic 7: terminal transmission signal wireless relay sequence trace basis pr say
Topic 8: assembly frame support arm wheel rear container lift bed transverse
Topic 9: lock door seal bolt arm cover closure container mechanism circuit
Topic 10: constraint order set route car shipment transportation voyage cost asset
Topic 11: sensor monitoring monitor module condition transmitter data touch wireless communication
Topic 12: inventory sortation item package station delivery bin sort bay pickup
Topic 13: robot robotic depot container location shelf destination item conveyor transport
Topic 14: refrigeration temperature control reefer preservation refrigerate air controller nutritional power
Topic 15: thing move dma examples network anomaly disclosure various autonomous infrastructure
Topic 16: image layer parcel region map pixel digital affinity container screenshot
Topic 17: vehicle autonomous electrical truck allocation transportation area configure collaborative passenger
Topic 18: track solenoid pump movement label indicia line body display tow
Topic 19: sensor door controller monitor cargo monitoring module container strap security
Topic 20: collector usb data structure collection circuit collect peripheral power line
Topic 21: serviceable shield package quality embodiment locker logistics ice customer delivery
Topic 22: address content file request processor issue least data cache network
Topic 23: image optical vessel longitudinal satellite sensor marker imagery object cusp
Topic 24: identity mobile device consumer access personally technology master identifiable environment
Topic 25: optical node optic mch width endpoint viability leak super network
Topic 26: screen incident status package ctu information vehicle action pertain reporting
Topic 27: service network connectivity premise connection client endpoint exploitable policy provision
Topic 28: transportation bid service provider vehicle subcomponent mission location demand transport
Topic 29: crane gantry land container dunnage load arrival arrangement lift spreader
Topic 30: supply drug chain alias logistics provider reorder disclosure identifier merchant
Topic 31: image develop toner form voltage latent photo unit member mode
Topic 32: charge question bus survey empty handheld cogeneration power incident data
Topic 33: transportation document instruction pgv poli set schema documentation hu load
Topic 34: power fault accessory electrical electrically test appliance device conductor designate
Topic 35: kit ship occurrence set cost geographic method shipping shipment event
Topic 36: wireless audio amplification signal network timer antenna factor indication two
Topic 37: shipment prediction international tote consignee delivery truckload full shipper order
Topic 38: fire vehicle entity group transporter identification provide associate mads sads
Topic 39: case sense detector multipurpose container breach chemical intrusion security matrix
Topic 40: load metric arrival transport section interest actual delivery validation cycle
Topic 41: radiation neutron ionize detector radioactive source detect nuclear level fissile
Topic 42: block srgb slice prefix igp sid game entertainment path tb
Topic 43: vehicle cargo movement transport intermodal guided travel region zone entry
Topic 44: demand logistic chain supply good campaignable fragment forecast path policy
Topic 45: terminal call adapter telemetry local user offline parent node comparison
Topic 46: delivery sort destination prober agv compartment item parcel parent postal
Topic 47: logic fabric block interoperability instruction dissimilar transfer execution editor switch
Topic 48: radiation quasistatic field powder source detector search bulk material irradiation
Topic 49: database page result optimization pluggable node diversity feature transaction reside
Topic 50: filter signal noise band reduction property unit ac overtone shelf
Topic 51: gps receiver asset track range module dgps within wireless housing
Topic 52: waveform ue transmission synchronization aspect cbg crc decode disclosure feedback
Topic 53: che screening environment screen pibo platform execution manufacture job continuous
Topic 54: mutation cargo vibration container manifest security agency threat collection csu
Topic 55: chain supply vendor task orchestration financial agent event software pricing
Topic 56: inspection guideway scan cargo optical hazardous spacers nonintrusive step electromagnetic
Topic 57: lidar light optical mount anchor head pulse scatter camera cantilever
Topic 58: humidity vibration timekeeping container unit tag upper division isolation plate
Topic 59: flash stop memory permit pcd dram mt boot transfer pickup
Topic 60: storage virtualization controller tier decentralize io volume logical array host
Topic 61: carrier compute schedule shipment datum page transport web optimization recipient
Topic 62: electric charge power supply craft motor harbor housing water invention
Topic 63: line drilling traction float actuator hawser buoyant dock end weight
Topic 64: semantic deep profile enterprise contract node feature model observation learning
Topic 65: time base receiver particular network position interconnected station statistical easily
Topic 66: tag laser contract grid gridsat smart environmental identifier intermediate sensor
Topic 67: tilt feedback dolly scan apparatus dimension section memory logistics data
```

그래서, 우선 토픽 수 제한을 걸고 뽑았는데

```
   Topic  Count                                  Name  \
0     -1    577          -1_system_data_container_device
1      0    350             0_network_device_data_system
2      1    248             1_item_vehicle_method_system
3      2    195         2_container_cargo_include_system
4      3     90                3_vessel_marine_ship_sail
5      4     25           4_image_container_include_form
6      5     16  5_radiation_source_detector_quasistatic
7      6     15        6_carrier_aspect_compute_waveform
8      7      6                 7_shaft_actuate_lock_fit
9      8      5             8_light_optical_object_lidar

Topic -1: system data container device include method information item use provide
Topic 0: network device data system signal include information method wireless communication
Topic 1: item vehicle method system include delivery data order storage provide
Topic 2: container cargo include system lock sensor control device position door
Topic 3: vessel marine ship sail system position image control include least
Topic 4: image container include form layer develop toner unit print map
Topic 5: radiation source detector quasistatic detect field neutron container ionize radioactive
Topic 6: carrier aspect compute waveform srgb ue transmission slice prefix synchronization
Topic 7: shaft actuate lock fit rotate corner strap clamp pawl unit
Topic 8: light optical object lidar head pulse scatter configure portion include
```

내용들이 LDA에 비해 너무 좋지 않은 것 같아서 GTM해도 의미가 없을 것 같아 여기까지만 해봤습니다.
BERTopic은 안하고,, LDA만 진행하면 될까요,,,?