05_20

우선, BERTopic -> GTM 아직 못했구요,,, 내일 최대한 이른 저녁까지는 해보겠습니다.
(LDA -> GTM 해봤으니깐 금방할 수 있을 것 같아요. 근데, 정제하구나서 문서 수가 줄어들어서 BERTopic에서 불용어가 튀어나와서 BERTopic결과는 더 안좋을 것 같아요. 대신 LDA 짱짱으로 만들었어요)

그리고 **LDA -> GTM은 다 했는데** 구현하는데 오래걸려서 결론을 아직 못냈습니다.**(내용 이해도 할겸 이거 좀 도와주세요,,,⭐⭐⭐)**
시간 진짜 많이 갈아넣었는데 아직 못했어요,,,

- 주변에 있는 공백끼리 묶어서 이런저런 Topic 들 보면서 이런 기술이 부족하고 연구필요하다는 결론
  - 아래 참고 논문인데 우선 1차 발표때 요런 느낌으로 결론낼까하는데 저기처럼 저 단어 들어간 특허까지는 있으면 챙겨 넣어줘도 좋을 거 같고?

Table 4.11 주변 특허(VA 3, VA 5) 키워드

	keyword
VA 3 주변 특허 (32개)	**sensor**, method, **application**, system, patient, device, **apparatus**, computer, **monitoring**, plurality, time, condition, record, user, unit, embodiment, information, parameter, processor, control, function, measurement
VA 5 주변 특허 (84개)	**sensor**, device, system, method, event, information, **monitoring**, patient, user, condition, **apparatus**, example, **drug**, communication, health, plurality

sensor, monitor, monitoring 등의 키워드가 등장하는 것을 보았을 때, 종합적으로 대상을 감지하고 측정하고 모니터링 하는 기술에 대한 개발이 필요하다고 할 수 있다. VA 3과 VA 5의 주변에 위치한 실제 특허의 예시는 Table 4.12와 Table 4.13에 나타난 것과 같다.

- 33 -

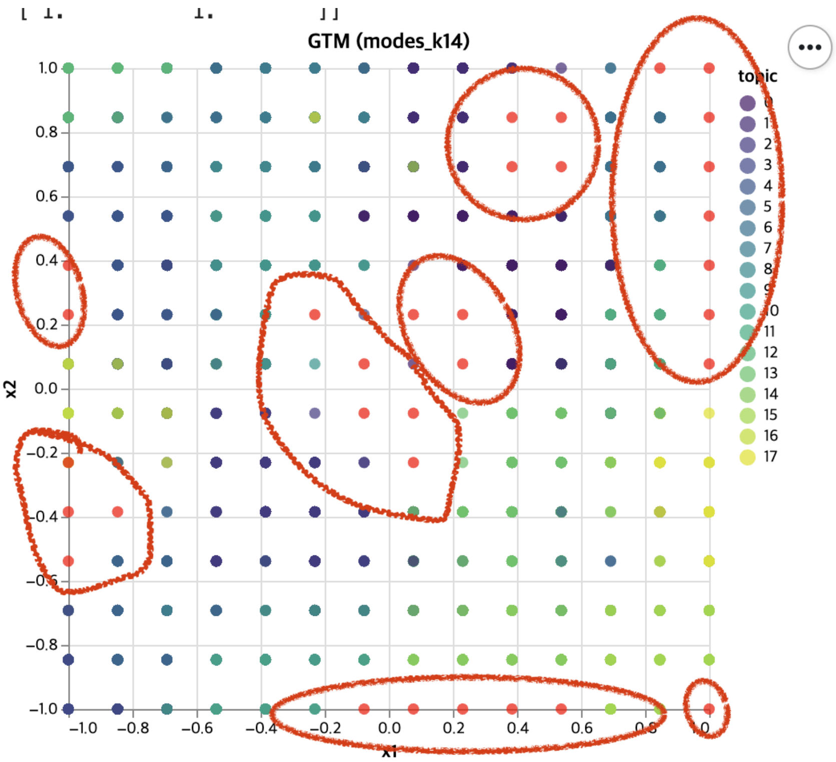
Table 4.12 주변 특허 예시 - VA 3

특허 번호	제목
8667290	Efficient, high volume digital signature system for medical and business applications
9173567	Triggering user queries based on sensor inputs
9339193	Physiological adaptability system with multiple sensors
10172593	Pregnancy state monitoring
10182763	Intelligent assistive mobility device

VA 3 주변에 있는 실제 특허의 일부 예시를 보면 대용량 디지털 서명 시스템, 센서 입력에 따른 사용자 질의응답 관련 기술, 다중 센서가 있는 시스템, 환자의 상태 모니터링, 지능형 보조 이동 장치 등이 존재한다. 따라서 해당 공백 영역에서는 센서를 통해 환자의 상태를 파악하고 입력값을 다양화하거나, 사용자들의 질의응답을 자동화시켜 헬스케어 시스템의 효율성을 높이는 소프트웨어 및 하드웨어 전반에 대한 기술 개발이 필요하다고 할 수 있다.

  -

- 근데, 여러 방법을 해가지구 여러 버전으로 결론 내야합니다.? **신경망, knn, 랜덤 포레스트 회귀, 가우시안** 요 네개 결론 한 명당 하나씩만..
- 밑에 그림에 빨간점이 공백인데 저 점이 무슨 Topic에서 나왔는지가 나와있거던요 비슷한 위치에 있는 애들끼리 묶어가지구
- 위에 논문 처럼 결론 좀 가능할까요.,..(아래 내용 쭉 읽어보시면 뭐 하는거구나 알 수 있습니다.)
- 각각 결론 내는데 30분 ~1시간정도 걸릴거에요 !!
- 대충 비슷한거 묶어서 토픽 확인하고 이런 기술인데 이런게 공백이다라고 결론만 내주세요 ! ! !
-



-
- 요 부분처럼 한거에요

공백 영역의 유망성을 평가하기 위해 Fig. 4.1의 결과로부터 인접한 영역을 하나의 공백 영역으로 간주한다. 함께 평가되는 영역은 Fig. 4.3과 Table 4.3에 나타난 것과 같다. 본 연구에서는 각 공백 영역 하나하나를 vacant로, 함께 평가하는 영역은 VA(Vacant Area)로 정의하였으며 인접 vacant를 하나로 묶은 결과 총 10개의 VA가 정의되었다.

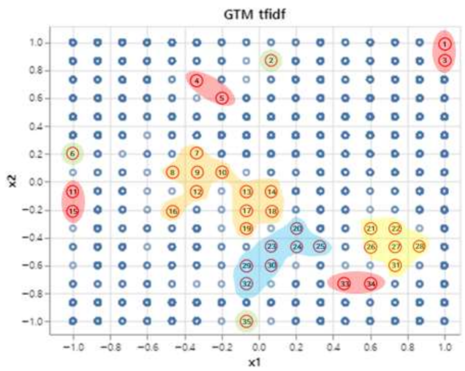


Fig. 4.3 공백 영역(VA)

Table 4.3 공백 영역(VA) 구분

	vacant
VA 1	vacant 1, vacant 3
VA 2	vacant 2
VA 3	vacant 4, vacant 5
VA 4	vacant 6
VA 5	vacant 7, vacant 8, vacant 9, vacant 10, vacant 12, vacant 13, vacant 14, vacant 16, vacant 17, vacant 18, vacant 19
VA 6	vacant 11, vacant 15
VA 7	vacant 20, vacant 23, vacant 24, vacant 25, vacant 29, vacant 30, vacant 32
VA 8	vacant 21, vacant 22, vacant 26, vacant 27, vacant 28, vacant 31
VA 9	vacant 33, vacant 34
VA 10	vacant 35

-

이제 내용 시작

.

.

.

LDA

달라진 건 불용어가 추가 + 토픽 수

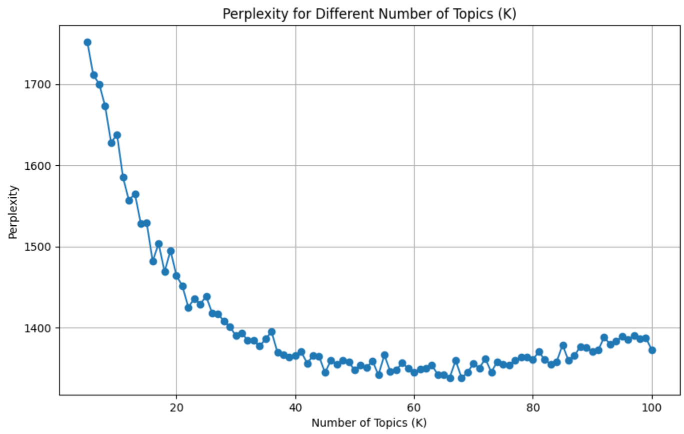
나머지는 기존과 동일.

요게 추가한 불용어구여

```
stop_words = [
"first", "second", "device", "includes", "portion", "one", "third", "least", "layer", "receiving",
"system", "storage", "server", "computer", "processing", "client", "may", "communication", "terminal",
"information", "station", "access", "vehicle", "transport", "transportation", "unit", "method",
"container", "load", "material", "door", "item", "product", "package", "order", "carrier", "plurality",
"set", "time", "value", "determining", "event", "state", "detection", "condition", "mode", "status",
"message", "path", "address", "end", "assembly", "member", "side", "body", "service", "connection",
"resource", "associated", "unit", "configured", "processor", "mean", "position", "water", "transaction",
"identifier", "supply", "entity", "block", "embodiment", "component", "invention", "present", "used",
"use", "multiple", "port", "test", "channel", "aspect", "signal", "power", "electrical", "element",
"object", "image", "user", "content", "display", "code", "computing", "includes", "based", "model",
"object", "using", "action", "image", "process", "type", "generated", "determine", "flow", "air", "said",
"heat", "system", "method", "embodiment", "device", "various", "example", "include", "described",
"herein", "identification", "reader", "within", "shipping", "transmission", "base", "transmitted",
"transmitting", "signal", "apparatus", "frequency", "section", "sample", "wall", "structure", "substrate",
"housing", "interior", "opening", "comprising", "provides", "specific", "low", "allows", "high", "box",
"large", "provided", "module", "remote", "communication", "center", "management", "via", "equipment",
"video", "software", "subsystem", "controller", "instruction", "input", "memory", "operation", "output",
"interface", "parameter", "delivery", "item", "inventory", "label", "destination", "article", "source",
"light", "measurement", "receiver", "transfer", "medium", "electronic", "part", "step", "line", "cable",
"mooring", "hull", "zone", "area", "movement", "inspection", "interface", "connected", "external",
"host", "shipment", "chain", "provider", "request", "customer", "good", "available", "management",
"link", "source", "routing", "segment", "frame", "support", "mechanism", "arm", "coupled", "application",
"configuration", "file", "response", "policy", "distributed", "internet", "key", "local", "call", "platform",
"medium", "profile", "document", "new", "interaction", "also", "voice", "user", "unit", "signal", "port",
"bus", "core", "including", "system", "physical", "component", "method", "device", "wherein", "record",
"unique", "service", "account", "party", "payment", "received", "selected", "object", "sensing",
"detecting", "detect", "distance", "mean", "group", "cell", "number", "according", "aspect", "disclosure",
"comprises", "reference", "moving", "field", "electrically", "internal", "rack", "drive", "along",
"direction", "disposed", "mounted", "upper", "movable", "bottom", "temperature", "valve", "volume", "tube",
"reaction", "communication", "gateway", "message", "link", "connectivity", "receive", "transceiver",
"transmit", "item", "facility", "vehicle", "good", "site", "handling", "speed", "actuator", "propulsion",
"unmanned", "virtual", "command", "program", "instruction", "execution", "threshold", "level", "stored",
"alert", "active", "table", "central", "context", "quality", "lock", "truck", "seal", "open", "content",
"rate", "cost", "value", "period", "current", "amount", "constraint", "generate", "task", "rule",
"attribute", "agent", "pattern", "feature", "analysis", "perform", "point", "travel", "region", "collection",
"main", "enclosure", "locking", "top", "plate", "pair", "extending", "closed", "device", "engine",
"receive", "store", "operating", "related", "master", "operational", "method", "value", "medium",
"representation", "user", "providing", "graphical", "displaying", "screen", "disclosed", "vehicle", "plan",
"freight", "arrival", "station", "space", "natural", "outlet", "instruction", "generating", "system",
"activity", "cause", "contract", "card", "public", "port", "unit", "programmable", "element", "two",
"corresponding", "register", "buffer", "line", "object", "image", "determined", "capture", "characteristic",
"imaging", "result", "communication", "function", "terminal", "appliance", "provide", "link", "payload",
"respective", "path", "manager", "transducer", "array", "acoustic", "adapted", "solution", "relationship",
"result", "asset", "transmitter", "motor", "control", "desired", "charging", "compartment", "work",
"loading", "move", "loaded", "weight", "signal", "circuitry", "filter", "electromagnetic", "aspect",
"basis", "slot", "change", "primary", "index", "item", "good", "product", "order", "distribution",
"package", "store", "unit", "beam", "system", "method", "motion", "generator", "point", "attached",
"relative", "embodiment", "aspect", "resource", "device", "service", "technology", "communication",
"transmits", "positioning", "encoded", "containing", "condition", "range", "driver", "setting", "sequence",
"station", "different", "particular", "trip", "object", "content", "view", "receptacle", "conductive",
"form", "line", "element", "message", "receives", "session", "relay", "sending", "terminal", "search",
"industrial", "representing", "record", "query", "entity", "whether", "event", "value", "detected",
"report", "metric", "proxy", "expected", "delay", "component", "hardware", "connection", "communicate",
"particular", "magnetic", "business", "well", "control", "phase", "thereby", "thus", "required", "holding",
"identified", "learning", "interest", "indicator", "instruction", "navigation", "deployment", "mast",
"port", "link", "size", "vehicle", "order", "storing", "block", "good", "medium", "recipient", "option",
"consumer", "shipper", "item", "good", "system", "method", "order", "product", "service", "device",
```

"message", "topology", "connectivity", "address", "edge", "value", "parameter", "measured", "obtained",
"calculated", "calculating", "measuring", "difference", "candidate", "instruction", "processor", "result",
"action", "behavior", "executed", "operation", "vehicle", "object", "condition", "characteristic", "factor",
"transit", "updated", "force", "inner", "wheel", "outer", "shaft", "bag", "aspect", "thing", "another",
"dynamic", "responsive", "positioned", "mean", "located", "port", "unit", "block", "element", "fabric",
"switch", "station", "interconnected", "sample", "solid", "density", "medium", "shared", "signaling",
"detector", "threat", "contact", "printing", "piece", "breach", "column", "building", "audio", "vessel",
"predetermined", "cycle", "controlled", "interval", "clock", "symbol", "reception", "component", "module",
"single", "application", "resource", "entity", "entry", "rule", "transaction", "store", "indicating",
"sent", "error", "value", "given", "risk", "assigned", "defined", "determines", "system", "object",
"station", "arranged", "device", "transporting", "method", "resource", "service", "embodiment",
"application", "instruction", "block", "operation", "execute", "instance", "creating", "property",
"communication", "connect", "aspect", "server", "usage", "list", "user", "item", "transaction", "option",
"element", "intermediate", "diagnostic", "loop", "site", "fire", "respect", "control", "front", "ring",
"pod", "coupling", "operated", "mean", "plane", "relates", "performed", "included", "generation",
"interference", "flexible", "member", "mounting", "low", "cap", "signature", "upon", "authorization",
"certificate", "criterion", "rule", "communication", "signal", "unit", "capable", "port", "message",
"stream", "switch", "component", "operation", "condition", "modular", "product", "shipped", "reading",
"visual", "log", "record", "participant", "content", "composite", "separation", "containment", "device",
"communication", "connection", "application", "system", "service", "resource", "signal", "value", "spatial",
"noise", "series", "band", "point", "operation", "area", "content", "comprise", "wind", "auxiliary",
"incident", "treatment", "event", "result", "rule", "technique", "attack", "category", "process", "object",
"instruction", "image", "selection", "terminal", "displayed", "medium", "method", "page", "transaction",
"mean", "characteristic", "acquisition", "element", "sea", "additional", "variable", "leg", "exemplary",
"operable", "guide", "low", "cavity", "member", "alarm", "proximity", "asset", "presence", "aspect",
"relate", "priority", "next", "example", "unit", "port", "message", "link", "communicating", "good",
"vehicle", "without", "condition", "coordinate", "component", "module", "b", "line", "analog", "design",
"chassis", "c", "thing", "supporting", "item", "station", "order", "package", "pickup", "connecting",
"bar", "control", "floating", "rotatable", "actual", "asset", "system", "method", "condition", "case",
"certain", "requirement", "made", "rule", "tunnel", "device", "established", "designated", "transfer",
"signal", "receiver", "matching", "channel", "broadcast", "element", "encoding", "service", "thing",
"aspect", "application", "point", "timing", "line", "user", "communication", "component", "server",
"programmed", "good", "header", "need", "human", "order", "functionality", "capability", "medium",
"embodiment", "couple", "inside", "conduit", "formed", "member", "exterior", "cover", "content", "collect",
"characteristic", "operation", "fleet", "initial", "portion", "ambient", "item", "good", "product",
"order", "link", "path", "enterprise", "result", "medium", "transaction", "operation", "unit", "terminal",
"station", "coverage", "instruction", "port", "module", "processor", "telephone", "adapter", "modem",
"compute", "resource", "failure", "among", "component", "discharge", "unit", "device", "service",
"resource", "application", "communication", "medium", "measurement", "meter", "forming", "requirement",
"fault", "interconnection", "condition", "event", "aspect", "consumption", "embodiment", "characteristic",
"feature", "asset", "reporting", "member", "moved", "value", "performing", "strength", "item", "method",
"generates", "change", "point", "thing", "reduce", "car", "low", "surface", "manage", "good", "networking",
"peer", "entity", "establish", "window", "flight", "message", "instruction", "message", "module",
"processor", "read", "job", "serial", "independent", "switch", "mean", "content", "person", "region",
"manifest", "mean", "system", "substantially", "user", "system", "mean", "thing", "readable", "applied",
"aspect", "medium", "event", "great", "thereof", "generally", "user", "allow", "define", "face", "estimate",
"plan", "predict", "contain", "deliver", "extend", "enable", "data", "medium", "member", "engage", "unit",
"record", "embodiment", "item", "write", "low", "thing", "monitor", "make", "result", "characteristic"
]

토픽 수



K : Topics (토픽수)

1000회 반복 기준, 5부터 100까지 1단위로 LDA 돌렸을때의 복잡도(Perplexity)를 그래프화 시킨거구요

그래서 급격하게 감소하다 완만해지는 지점을 최적의 토픽 수로 잡았구 18로 정했습니다

그렇게 해서 나온 토픽들

```
Topic 0 : signal circuit optical antenna rf radio connector switch chip digital
Topic 1 : value schedule aspect wireless uplink ue map determination bandwidth synchronization
Topic 2 : digital iot blockchain smart scan image identify device portable hash
Topic 3 : route traffic network method performance parameter model plan system algorithm
Topic 4 : network device protocol message machine address router ip switch resource
Topic 5 : system device method component environment service application embodiment integrate communication
Topic 6 : fluid gas pressure liquid chamber tank cool resonator fuel pump
Topic 7 : location track target ship gps sonar geographic object radar identify
Topic 8 : system element component monitor improve hub reduce maintenance maintain safety
Topic 9 : mobile database location record transfer terminal point change method update
Topic 10 : item good logistics method service system order product embodiment parcel
Topic 11 : energy mean radiation battery electric module circuit wire sensor emit
Topic 12 : network node wireless communication method system mobile radio device mesh
Topic 13 : sensor tag rfid monitor cargo monitoring track signal wireless security
Topic 14 : container vehicle item autonomous system location robot unit conveyor automate
Topic 15 : cargo trailer rail secure axis rotate lift deck anchor vertical
Topic 16 : network packet security traffic secure authentication communication endpoint encrypt identity
Topic 17 : marine satellite ship vessel control sail sensor fuel steer operator
```

나름 야무지게 나온거 같구요
(아쉬울 수 있지만 GTM이랑 아다리가 맞아야되서 이렇게라도 나오게 하려고 불용어 빼면서 LDA만 2000번 넘게 돌렸습니다ㅠㅠ)

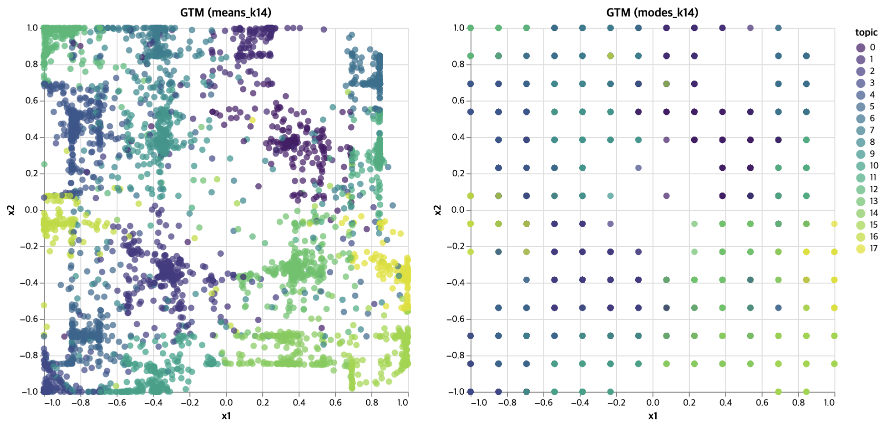
.

.

.

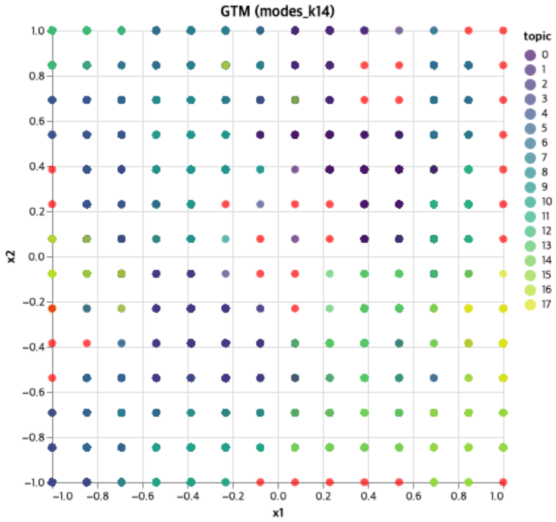
GTM

이후에 얻은 LDA 분포를 GTM을 통해 저차원 투영이죠. 시각화 시켰구여.



왼쪽이 모든 문서를 시각화(점) 찍은거구, 오른쪽은 그걸 군집화해서 (최대값에) 점찍었다구 생각하면됩니다.

이후에 공백 영역을 따냈어요. **빨간 점이 공백영역**



이제 공백영역이 어떤 토픽인지를 알아야하는데, 이걸 어떻게 알아내느냐.

우선 세 가지를 알아야 합니다.(내가 이해한 바로는 그래요)

1. **잠재 공간(Latent Space)**

- 잠재 공간이란 데이터의 고차원 특성을 저차원 공간으로 축소한 것
- GTM(Generative Topographic Mapping)은 고차원 데이터를 저차원 공간으로 매핑하여 데이터의 구조를 시각화하고 분석하는 데 도움을 줌

2. **책임 분포 (Responsibilities)**

- 각 데이터 포인트가 특정 잠재 공간 점(모델의 가우시안 함수)에서 생성될 확률
- 이를 통해 각 데이터 포인트가 잠재 공간의 어느 부분에 속하는지 알 수 있다.
- GTM에서 책임 분포는 가우시안 함수의 가중치로 계산

3. **역투영 (Inverse Projection)** - 역매핑이라구도 하죠

- 저차원 잠재 공간에서의 좌표를 고차원 데이터 공간으로 변환하는 과정
- 이를 통해 잠재 공간에서 선택한 위치가 원래 데이터 공간에서 어떤 의미를 가지는지 이해할 수
- GTM에서는 잠재 공간에서의 좌표를 책임 분포를 통해 원래 데이터 공간으로 변환

그래서 GTM에서 빈 영역의 토픽을 어떻게 찾느냐

빈 영역(잠재 공간에서의 특정 좌표)에 대한 책임 분포를 계산하여 이 영역이 어떤 토픽과 연관이 있는지 알아낼 수 있음.

1. **잠재 공간 좌표 선정**: 빈 영역의 잠재 공간 좌표를 선정
2. **가우시안 함수 계산**: 빈 영역 좌표에서의 가우시안 함수 값을 계산
3. **책임 분포 계산**: 가우시안 함수 값을 사용하여 각 데이터 포인트가 빈 영역 좌표에 대해 가질 책임 분포를 계산
4. **토픽 추정**: 책임 분포에서 가장 높은 값을 가지는 토픽을 선택

그래서

이걸 GTM 모델이 잠재 공간 좌표와 문서-토픽 분포를 정확하게 학습했다고 믿는 전제하에 아래와 같이 둬도 된다고 (gpt)가 그러더라구요

latent*coords(잠재 공간) = dgtm_modes_k14['x1', 'x2'].values*
-> 요건 GTM 군집한 그래프(GTM 맨 위 오른쪽 그래프)의 x1, x2
responsibilities(책임 분포) = lda_model.doc_topic
-> 요건 LDA topic 분포(각 문서가 각 토픽에 대해 가지는 확률)인데 이걸 바탕으로 gtm했으니깐 책임 분포로 이걸 둬도 되나봐영.

그러니깐, 정리하자면 잠재공간(x1, x2)과 책임분포(LDA에서 뽑은 분포)을 가정해두고, 내가 추출한 빈 영역의 잠재 공간 좌표를 설정한 다음에 **여러 방법**을 통해서 빈 영역에 대한 책임 분포를 계산 하는거에요 -> 요러면 빈 영역 좌표에 토픽이 딱 할당되죠

근데 왜 또 여러 방법이냐.
(가우시안으로 하려고 했는데 몇 시간 삽질해도 안되서 다른 방법 찾았음. 근데, 다하구나서 가우시안으로도 됐씁니다.)

그래도 좋은 점은 평가기준에 아주 부합하게 되었습니다👍🏻

- **1차 설계 발표**
  - 팀당 20분 내외 발표
  - 연구 주제에 대한 이해도 (10%)
    - ➢ 주제를 선정한 근거가 타당한가?
    - ➢ 연구 주제를 명쾌히 설명하였는가?
  - 연구의 창의성(원천성) 및 도전성 (20%)
    - ➢ 기존 연구 대비 독창성을 명쾌히 제시하였는가?
  - 연구 내용 및 방법의 적합성 (30%)
    - ➢ 해결방안을 적절히 개선하였는가?
  - 연구 내용의 효과성 (40%)
    - ➢ 제안한 연구의 결과물을 다각도로 제시하였는가?

GTM(Generative Topographic Mapping)에서 원래 사용되는 책임 분포 계산 방식은 가우시안 혼합 모델(Gaussian Mixture Model, GMM)에서 영감을 받아 각 데이터 포인트가 특정 잠재 공간 점에서 생성될 확률을 나타냄

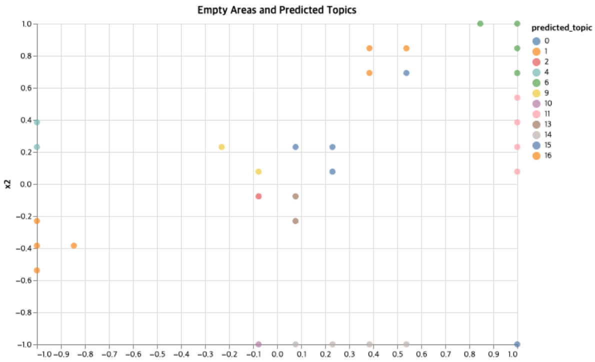
이 과정은 일반적으로 가우시안 함수의 가중치를 기반으로 수행 (이래서 정석은 가우시안)
(참고했던 GTM 논문들에 있던 수식이 이 가우시안 수식이였어요. 그래서 나중에 GTM 좀 더 보강해야될듯?)

그러나 실제 응용에서는 다양한 머신러닝 모델을 사용하여 유사한 목적을 달성할 수 있습니다.
신경망, k-최근접 이웃(k-NN), 랜덤 포레스트 회귀(RandomForestRegressor)와 같은 다른 방법을 사용하여 GTM의 잠재 공간 좌표를 예측하는게 가능.

이게 아까 찍었던 빨간점(빈 영역)에 토픽을 할당시키는거 (지금은 우선 가장 최대값으로 할당했는데 보통 다른 논문들에서는 임계값을 정하고 그 이상인 토픽들을 다 채택해서 이 영역은 뭐 이런 기술인데 공백이다 이런 느낌으로 감. 그리고 주변에 같이 있는 공백끼리 엮기도 함.)

.
.
.
.
.
.
발표할때는 각각 무슨 무슨 방법인지 말해줘야할 것 같기두 하고

- **신경망 (Neural Networks)**
  - 얘는 좀 잘나온 편 주변애들끼리 서로 비슷하고 means_k14랑 비교해도 그럴듯



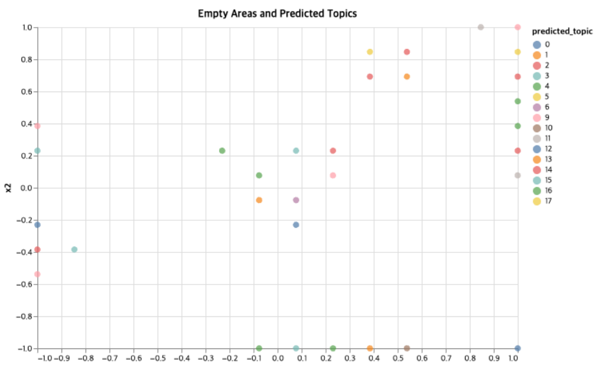
  -
  - **이거 PPT에 넣을때는 동그라미 점 위에 토픽 위에 숫자 적어주면 좋을듯, 그리고 사진은 다시 드리께여 Topic 1부터 시작하게 좀 바꾸려구여**
  - 아래 문장은 그냥 시각화 한거 말로 설명한겁니다(혹시나 나중에 필요한데 없을까봐 찍어둔거)
  - 저기 해당하는 좌표가 topic 몇번이 가장 많이 나왔다 이런 느낌 책임 분포 확률은 뭐다 라고 써져있습니다.
  -

```
Empty area 0 at [-1.         -0.53846154] is most likely associated with topic 16 with responsibility 0.42376264929771423
Empty area 1 at [-1.         -0.38461538] is most likely associated with topic 16 with responsibility 0.5384473204612732
Empty area 2 at [-1.         -0.23076923] is most likely associated with topic 16 with responsibility 0.6041637659072876
Empty area 3 at [-1.          0.23076923] is most likely associated with topic 4 with responsibility 0.42052653431892395
Empty area 4 at [-1.          0.38461538] is most likely associated with topic 4 with responsibility 0.5881344676017761
Empty area 5 at [-0.84615385 -0.38461538] is most likely associated with topic 5 with responsibility 0.3842390477657318
Empty area 6 at [-0.07692308 -1.        ] is most likely associated with topic 10 with responsibility 0.4347069561481476
Empty area 7 at [ 0.07692308 -1.        ] is most likely associated with topic 14 with responsibility 0.4487851560115814
Empty area 8 at [ 0.23076923 -1.        ] is most likely associated with topic 14 with responsibility 0.6445967555046082
Empty area 9 at [ 0.38461538 -1.        ] is most likely associated with topic 14 with responsibility 0.7363776564598083
Empty area 10 at [ 0.53846154 -1.        ] is most likely associated with topic 14 with responsibility 0.6589601039886475
Empty area 11 at [ 1. -1.] is most likely associated with topic 15 with responsibility 0.8486230969429016
Empty area 12 at [-0.23076923  0.23076923] is most likely associated with topic 9 with responsibility 0.4906749725341797
Empty area 13 at [-0.07692308 -0.07692308] is most likely associated with topic 2 with responsibility 0.2487557828426361
Empty area 14 at [-0.07692308  0.07692308] is most likely associated with topic 9 with responsibility 0.2605195939540863
Empty area 15 at [ 0.07692308 -0.07692308] is most likely associated with topic 13 with responsibility 0.19354087114334106
Empty area 16 at [ 0.07692308 -0.23076923] is most likely associated with topic 13 with responsibility 0.289632648229599
Empty area 17 at [0.07692308 0.23076923] is most likely associated with topic 0 with responsibility 0.36470094323158264
Empty area 18 at [0.23076923 0.07692308] is most likely associated with topic 0 with responsibility 0.4743216931819916
Empty area 19 at [0.23076923 0.23076923] is most likely associated with topic 0 with responsibility 0.6694310903549194
Empty area 20 at [0.38461538 0.69230769] is most likely associated with topic 0 with responsibility 0.2416241317987442
Empty area 21 at [0.38461538 0.84615385] is most likely associated with topic 1 with responsibility 0.4663020074367523
Empty area 22 at [0.53846154 0.69230769] is most likely associated with topic 0 with responsibility 0.19122248888015747
Empty area 23 at [0.53846154 0.84615385] is most likely associated with topic 1 with responsibility 0.2832971513271332
Empty area 24 at [0.84615385 1.        ] is most likely associated with topic 6 with responsibility 0.7835562229156494
Empty area 25 at [1.         0.07692308] is most likely associated with topic 11 with responsibility 0.4233115315437317
Empty area 26 at [1.         0.23076923] is most likely associated with topic 11 with responsibility 0.6516543626785278
Empty area 27 at [1.         0.38461538] is most likely associated with topic 11 with responsibility 0.6411468982696533
Empty area 28 at [1.         0.53846154] is most likely associated with topic 11 with responsibility 0.44324934482574463
Empty area 29 at [1.         0.69230769] is most likely associated with topic 6 with responsibility 0.5418524742126465
Empty area 30 at [1.         0.84615385] is most likely associated with topic 6 with responsibility 0.8041566014289856
Empty area 31 at [1. 1.] is most likely associated with topic 6 with responsibility 0.9103043079376221
```

- **장점**: 비선형 관계를 잘 모델링할 수 있으며, 복잡한 데이터 패턴을 학습하는 데 효과적입니다.
- **단점**: 학습에 많은 데이터와 계산 자원이 필요하며, 하이퍼파라미터 튜닝이 복잡할 수 있습니다.

.
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.

- **k-최근접 이웃 (k-Nearest Neighbors, k-NN)** -
  - 이건 잘안나온편 주변에 붙어 있는데도 불구하고 토픽이 다 다르게 나옴
  - 이게 그냥 내가 눈으로 보고 판단한거긴한데 발표할때 어떻게 말할지는 조금 의문



-

-

-

```
Empty area 0 at [-1.         -0.53846154] is most likely associated with topic 16 with responsibility 0.25165505020043033
Empty area 1 at [-1.         -0.38461538] is most likely associated with topic 16 with responsibility 0.6014832019711468
Empty area 2 at [-1.         -0.23076923] is most likely associated with topic 16 with responsibility 0.6014832019711468
Empty area 3 at [-1.          0.23076923] is most likely associated with topic 4 with responsibility 0.45427071416694104
Empty area 4 at [-1.          0.38461538] is most likely associated with topic 4 with responsibility 0.6556616066276827
Empty area 5 at [-0.84615385 -0.38461538] is most likely associated with topic 5 with responsibility 0.41899594980884125
Empty area 6 at [-0.07692308 -1.        ] is most likely associated with topic 10 with responsibility 0.5105583243851143
Empty area 7 at [ 0.07692308 -1.        ] is most likely associated with topic 10 with responsibility 0.4942209421661169
Empty area 8 at [ 0.23076923 -1.        ] is most likely associated with topic 14 with responsibility 0.5866743653472828
Empty area 9 at [ 0.38461538 -1.        ] is most likely associated with topic 14 with responsibility 0.6625466001624399
Empty area 10 at [ 0.53846154 -1.        ] is most likely associated with topic 14 with responsibility 0.4745944322591211
Empty area 11 at [ 1. -1.] is most likely associated with topic 15 with responsibility 0.7524454684760937
Empty area 12 at [-0.23076923  0.23076923] is most likely associated with topic 9 with responsibility 0.337065425301025
Empty area 13 at [-0.07692308 -0.07692308] is most likely associated with topic 2 with responsibility 0.36812502800403857
Empty area 14 at [-0.07692308  0.07692308] is most likely associated with topic 9 with responsibility 0.30735592782651294
Empty area 15 at [ 0.07692308 -0.07692308] is most likely associated with topic 13 with responsibility 0.332351698345197
Empty area 16 at [ 0.07692308 -0.23076923] is most likely associated with topic 13 with responsibility 0.37541777449988173
Empty area 17 at [0.07692308 0.23076923] is most likely associated with topic 0 with responsibility 0.437574933446222
Empty area 18 at [0.23076923 0.07692308] is most likely associated with topic 0 with responsibility 0.3948764607810425
Empty area 19 at [0.23076923 0.23076923] is most likely associated with topic 0 with responsibility 0.5098589990180146
Empty area 20 at [0.38461538 0.69230769] is most likely associated with topic 0 with responsibility 0.31853129930862806
Empty area 21 at [0.38461538 0.84615385] is most likely associated with topic 1 with responsibility 0.5247880959775718
Empty area 22 at [0.53846154 0.69230769] is most likely associated with topic 0 with responsibility 0.30486329794003547
Empty area 23 at [0.53846154 0.84615385] is most likely associated with topic 1 with responsibility 0.2563163458020645
Empty area 24 at [0.84615385 1.        ] is most likely associated with topic 6 with responsibility 0.7353457473037716
Empty area 25 at [1.         0.07692308] is most likely associated with topic 17 with responsibility 0.48523121006056075
Empty area 26 at [1.         0.23076923] is most likely associated with topic 11 with responsibility 0.5850199704833748
Empty area 27 at [1.         0.38461538] is most likely associated with topic 11 with responsibility 0.48638383918673234
Empty area 28 at [1.         0.53846154] is most likely associated with topic 17 with responsibility 0.48523121006056075
Empty area 29 at [1.         0.69230769] is most likely associated with topic 17 with responsibility 0.48523121006056075
Empty area 30 at [1.         0.84615385] is most likely associated with topic 6 with responsibility 0.7353457473037716
Empty area 31 at [1. 1.] is most likely associated with topic 6 with responsibility 0.7353457473037716
```

- **장점**: 구현이 간단하고, 데이터의 지역적 패턴을 잘 반영합니다.

- **단점**: 계산 비용이 높을 수 있으며, 고차원 데이터에서 성능이 저하될 수 있습니다.

.

.

.

.

.

.

.**VDES**

.

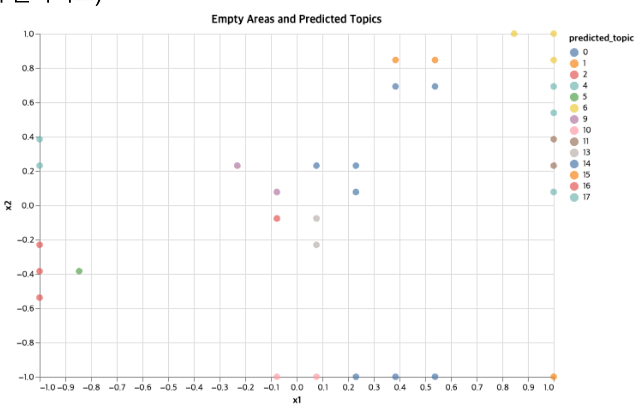
.

.

.

- **랜덤 포레스트 회귀 (RandomForestRegressor)**

  - 요건 중간정도? 나온편(내 눈으로 봤을 때, 구니깐 이게 각각의 방법마다 결과 해석이 완전 달라짐 할당된 Topic이 달라서 공백으로 판단하는 게 달라져요)



-

```
Empty area 0 at [-1.         -0.53846154] is most likely associated with topic 16 with responsibility 0.25165505020043033
Empty area 1 at [-1.         -0.38461538] is most likely associated with topic 16 with responsibility 0.6014832019711468
Empty area 2 at [-1.         -0.23076923] is most likely associated with topic 16 with responsibility 0.6014832019711468
Empty area 3 at [-1.          0.23076923] is most likely associated with topic 4 with responsibility 0.45427071416694104
Empty area 4 at [-1.          0.38461538] is most likely associated with topic 4 with responsibility 0.6556616066276827
Empty area 5 at [-0.84615385 -0.38461538] is most likely associated with topic 5 with responsibility 0.41899594980884125
Empty area 6 at [-0.07692308 -1.        ] is most likely associated with topic 10 with responsibility 0.5105583243851143
Empty area 7 at [ 0.07692308 -1.        ] is most likely associated with topic 10 with responsibility 0.4942209421661169
Empty area 8 at [ 0.23076923 -1.        ] is most likely associated with topic 14 with responsibility 0.5866743653472828
Empty area 9 at [ 0.38461538 -1.        ] is most likely associated with topic 14 with responsibility 0.6625466001624399
Empty area 10 at [ 0.53846154 -1.        ] is most likely associated with topic 14 with responsibility 0.4745944322591211
Empty area 11 at [ 1. -1.] is most likely associated with topic 15 with responsibility 0.7524454684760937
Empty area 12 at [-0.23076923  0.23076923] is most likely associated with topic 9 with responsibility 0.337065425301025
Empty area 13 at [-0.07692308 -0.07692308] is most likely associated with topic 2 with responsibility 0.36812502800403857
Empty area 14 at [-0.07692308  0.07692308] is most likely associated with topic 9 with responsibility 0.30735592782651294
Empty area 15 at [ 0.07692308 -0.07692308] is most likely associated with topic 13 with responsibility 0.332351698345197
Empty area 16 at [ 0.07692308 -0.23076923] is most likely associated with topic 13 with responsibility 0.37541777449988173
Empty area 17 at [0.07692308 0.23076923] is most likely associated with topic 0 with responsibility 0.437574933446222
Empty area 18 at [0.23076923 0.07692308] is most likely associated with topic 0 with responsibility 0.3948764607810425
Empty area 19 at [0.23076923 0.23076923] is most likely associated with topic 0 with responsibility 0.5098589990180146
Empty area 20 at [0.38461538 0.69230769] is most likely associated with topic 0 with responsibility 0.31853129930862806
Empty area 21 at [0.38461538 0.84615385] is most likely associated with topic 1 with responsibility 0.5247880959775718
Empty area 22 at [0.53846154 0.69230769] is most likely associated with topic 0 with responsibility 0.30486329794003547
Empty area 23 at [0.53846154 0.84615385] is most likely associated with topic 1 with responsibility 0.2563163458020645
Empty area 24 at [0.84615385 1.        ] is most likely associated with topic 6 with responsibility 0.7353457473037716
Empty area 25 at [1.         0.07692308] is most likely associated with topic 17 with responsibility 0.48523121006056075
Empty area 26 at [1.         0.23076923] is most likely associated with topic 11 with responsibility 0.5850199704833748
Empty area 27 at [1.         0.38461538] is most likely associated with topic 11 with responsibility 0.48638383918673234
Empty area 28 at [1.         0.53846154] is most likely associated with topic 17 with responsibility 0.48523121006056075
Empty area 29 at [1.         0.69230769] is most likely associated with topic 17 with responsibility 0.48523121006056075
Empty area 30 at [1.         0.84615385] is most likely associated with topic 6 with responsibility 0.7353457473037716
Empty area 31 at [1. 1.] is most likely associated with topic 6 with responsibility 0.7353457473037716
```

-

- **장점**: 앙상블 학습 방법으로 과적합을 방지하고, 복잡한 데이터 패턴을 잘 학습합니다.

- **단점**: 많은 트리를 사용할 경우 계산 비용이 증가할 수 있습니다.

.

.

.

.

.

.

.

.

.

.

.

.

.

.

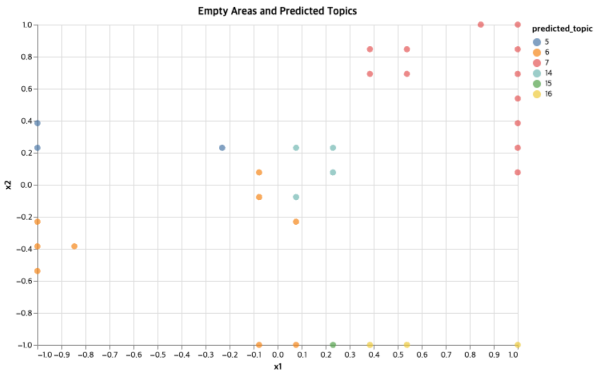
.

.

.

.

- **가우시안 프로세스 회귀 (Gaussian Process Regression)**
- 요것도 정석답게 좀 잘나온듯? 이게 젤 잘나왔나 싶기도 하고?

- 


```
Empty area 0 at [-1.         -0.53846154] is most likely associated with topic 6 with responsibility 0.09291025484242944
Empty area 1 at [-1.         -0.38461538] is most likely associated with topic 6 with responsibility 0.09050081217998063
Empty area 2 at [-1.         -0.23076923] is most likely associated with topic 6 with responsibility 0.08803460613294754
Empty area 3 at [-1.          0.23076923] is most likely associated with topic 5 with responsibility 0.0879242854605097
Empty area 4 at [-1.          0.38461538] is most likely associated with topic 5 with responsibility 0.08994892914552938
Empty area 5 at [-0.84615385 -0.38461538] is most likely associated with topic 6 with responsibility 0.08803742138856506
Empty area 6 at [-0.07692308 -1.        ] is most likely associated with topic 6 with responsibility 0.08120064910459117
Empty area 7 at [ 0.07692308 -1.        ] is most likely associated with topic 6 with responsibility 0.07793203716685407
Empty area 8 at [ 0.23076923 -1.        ] is most likely associated with topic 15 with responsibility 0.07739472280364605
Empty area 9 at [ 0.38461538 -1.        ] is most likely associated with topic 16 with responsibility 0.08226742473992983
Empty area 10 at [ 0.53846154 -1.        ] is most likely associated with topic 16 with responsibility 0.08807172292092508
Empty area 11 at [ 1. -1.] is most likely associated with topic 16 with responsibility 0.1064177207662386
Empty area 12 at [-0.23076923  0.23076923] is most likely associated with topic 5 with responsibility 0.07289166155887579
Empty area 13 at [-0.07692308 -0.07692308] is most likely associated with topic 6 with responsibility 0.07088165752908673
Empty area 14 at [-0.07692308  0.07692308] is most likely associated with topic 6 with responsibility 0.06906868876847132
Empty area 15 at [ 0.07692308 -0.07692308] is most likely associated with topic 14 with responsibility 0.0689920787080701
Empty area 16 at [ 0.07692308 -0.23076923] is most likely associated with topic 6 with responsibility 0.06996650038280777
Empty area 17 at [0.07692308 0.23076923] is most likely associated with topic 14 with responsibility 0.06679706346518373
Empty area 18 at [0.23076923 0.07692308] is most likely associated with topic 14 with responsibility 0.06977974144823311
Empty area 19 at [0.23076923 0.23076923] is most likely associated with topic 14 with responsibility 0.06856133354023773
Empty area 20 at [0.38461538 0.69230769] is most likely associated with topic 7 with responsibility 0.07792063065544674
Empty area 21 at [0.38461538 0.84615385] is most likely associated with topic 7 with responsibility 0.08130254464837344
Empty area 22 at [0.53846154 0.69230769] is most likely associated with topic 7 with responsibility 0.08326194618153246
Empty area 23 at [0.53846154 0.84615385] is most likely associated with topic 7 with responsibility 0.08694111114443463
Empty area 24 at [0.84615385 1.        ] is most likely associated with topic 7 with responsibility 0.10277957680263064
Empty area 25 at [1.         0.07692308] is most likely associated with topic 7 with responsibility 0.08156552236338355
Empty area 26 at [1.         0.23076923] is most likely associated with topic 7 with responsibility 0.08601034642114087
Empty area 27 at [1.         0.38461538] is most likely associated with topic 7 with responsibility 0.09054348085925078
Empty area 28 at [1.         0.53846154] is most likely associated with topic 7 with responsibility 0.09513562886740505
Empty area 29 at [1.         0.69230769] is most likely associated with topic 7 with responsibility 0.09975648118747571
Empty area 30 at [1.         0.84615385] is most likely associated with topic 7 with responsibility 0.1043754853747673
Empty area 31 at [1. 1.] is most likely associated with topic 7 with responsibility 0.10896259560263433
```

- **장점**:
  - **확률적 모델링**: 데이터의 불확실성을 효과적으로 다룰 수 있습니다.
- **단점**:
  - **계산 비용**: 고차원 데이터나 대규모 데이터셋의 경우, 가우시안 함수의 계산이 매우 비용이 많이 들 수 있습니다. - O(n^3) 시간복잡도 개큼
  - **모델 선택**: 가우시안 함수의 너비(또는 스케일) 파라미터를 적절히 선택하는 것이 어려울 수 있습니다.

요기까쥐,,