

The Embedded Topic Model

INSA
TOULOUSE

Zineb Bennis



NIELS HENRIK ABEL
1802 – 1829

MATEMATIKER, BERØMT FOR
BANEBRYTENDE ARBEIDER INNEN
LIGNINGSTEORI, UENDELIGE REKKER
OG ELLIPTISKE FUNKSJONER



Table of contents

1 ASRS

2 LDA

3 Réseaux de neurones

4 Word Embedding et Word2Vec

5 ETM

6 Results

- Top words
- Comparison ETM/LDA
- Topic Coherence

Aviation Safety Reporting System

- Plusieurs millions de rapport.
- Mise à jour régulière
- Référence dans la littérature de NLP
- Enjeu de sécurité aérienne



Aviation Safety Reporting System

'After preflighting the aircraft we taxied out and took off. After the gear handle was retracted the main gear came up but the nose gear stayed down. We ran the emergency check list for the gear retraction. When the nose gear would not come up, we put the gear back down and returned for landing. On my post flight I found a bolt sticking only in the right side nose gear pin hole. It had no flag or handle on it like the aircraft gear pins. When looking up into the left side of the nose gear well you could see the left pin hole clear without seeing the pin sticking out of the right side pin hole. It must have been put in by a vendor or service center. Using only aircraft model type specific gear pins with flags would have avoided this occurrence and would prevent it from happening again.'

Table of contents

1 ASRS

2 LDA

3 Réseaux de neurones

4 Word Embedding et Word2Vec

5 ETM

6 Results

- Top words
- Comparison ETM/LDA
- Topic Coherence

Latent Dirichlet Allocation

- Un document = Une distribution de sujet
- Un sujet = Une distribution de mots
- Génération de "faux" documents
- Estimation du maximum de vraisemblance



Latent Dirichlet Allocation

- Un document = Une distribution de sujet
 - Un sujet = Une distribution de mots
 - Génération de "faux" documents
 - Estimation du maximum de vraisemblance
1. Draw topic proportion $\theta_d \sim \text{Dirichlet}(\alpha_\theta)$.
 2. For each word n in the document:
 - (a) Draw topic assignment $z_{dn} \sim \text{Cat}(\theta_d)$.
 - (b) Draw word $w_{dn} \sim \text{Cat}(\beta_{z_{dn}})$.

Table of contents

1 ASRS

2 LDA

3 Réseaux de neurones

4 Word Embedding et Word2Vec

5 ETM

6 Results

- Top words
- Comparison ETM/LDA
- Topic Coherence

Réseaux de neurones

- Interconnexion des neurones
- Les neurones "s'allument"
- Liaisons par des poids et des biais
- Entrainement à travers la Retropropagation

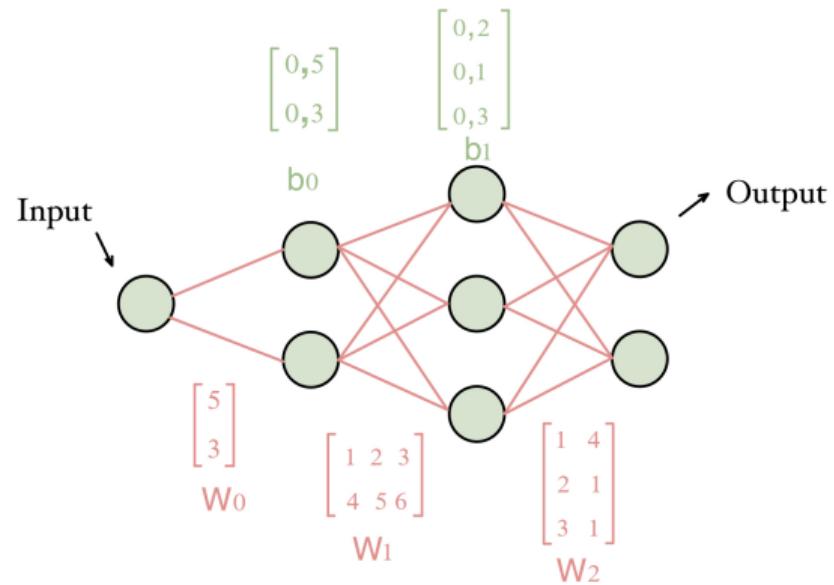
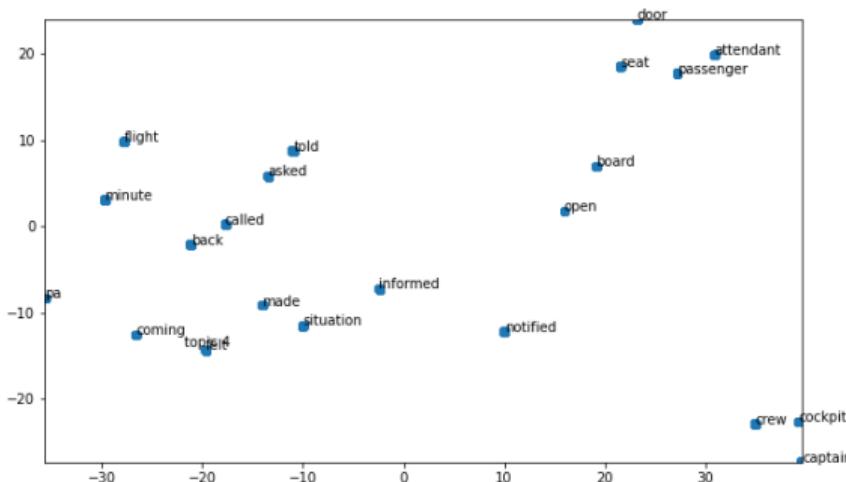


Table of contents

- 1 ASRS
- 2 LDA
- 3 Réseaux de neurones
- 4 Word Embedding et Word2Vec
- 5 ETM
- 6 Results
 - Top words
 - Comparison ETM/LDA
 - Topic Coherence

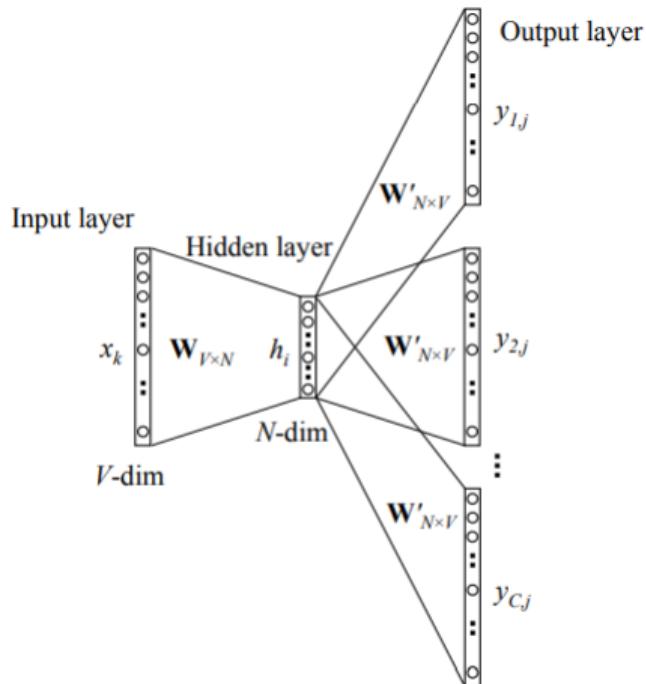
Word Embedding

- Projection de mots dans un space numérique
- Les mots similaires possèdent des vecteurs qui sont relativement proches.
- Représentation graphique possible



Word2Vec

- Par Google en 2013
- Réseau de neurones à une couche
- Prédiction des mots contexte à partir du mot cible



Word2Vec

- Par Google en 2013
- Réseau de neurones à une couche
- Prédiction des mots contexte à partir du mot cible
- "Opérations" numériques possibles

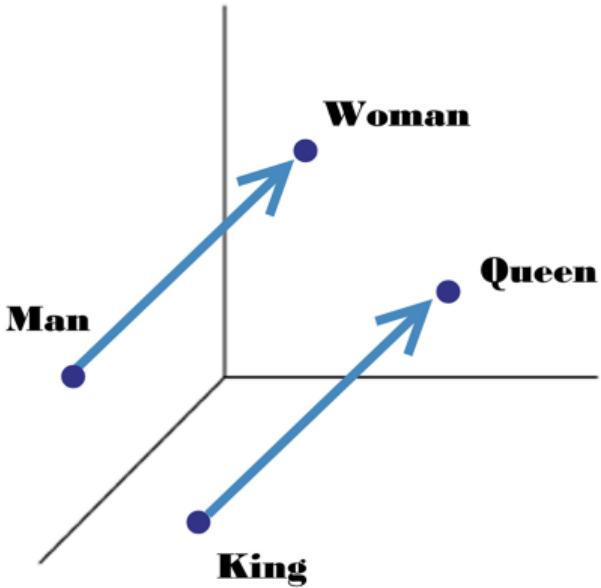


Table of contents

- 1 ASRS
- 2 LDA
- 3 Réseaux de neurones
- 4 Word Embedding et Word2Vec
- 5 ETM
- 6 Results
 - Top words
 - Comparison ETM/LDA
 - Topic Coherence

The Embedded Topic Model

- Combinaison de LDA et Word2Vec
 - Utilisation de réseaux de neurones
-
1. Draw topic proportions $\theta_d \sim \mathcal{LN}(0, I)$.
 2. For each word n in the document:
 - a. Draw topic assignment $z_{dn} \sim \text{Cat}(\theta_d)$.
 - b. Draw the word $w_{dn} \sim \text{softmax}(\rho^\top \alpha_{z_{dn}})$.

Table of contents

1 ASRS

2 LDA

3 Réseaux de neurones

4 Word Embedding et Word2Vec

5 ETM

6 Results

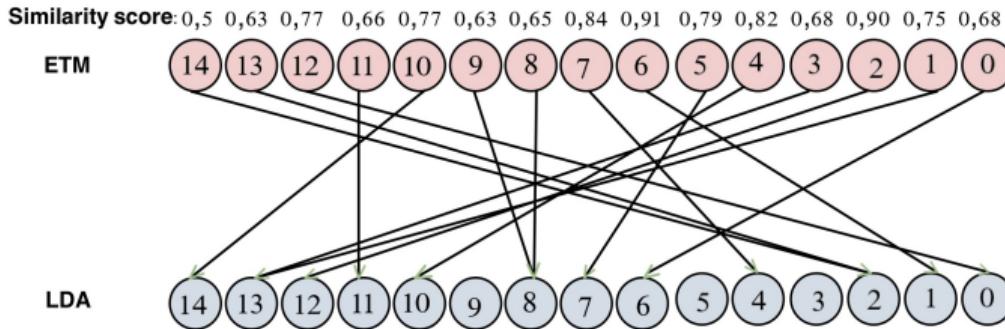
- Top words
- Comparison ETM/LDA
- Topic Coherence

Top Words

Subject	top words
Landing	runway, aircraft, tower, traffic, airport, pilot, final, pattern, turn downwind, landing, base, mile, land, position, approach
Cabin and crew	flight, passenger, attendant, captain, door, cabin, cockpit seat, crew, aircraft, back, told, smoke, asked
Mechanic maintenance	maintenance, aircraft, check, mel, mechanic, found, item, part, zzz(sleep), problem, flight, logbook, manual, inspection, system, work, control
Mechanic and brakes	aircraft, gear, landing, brake, nose, damage, light, ground, wing, runway, main, ramp, gate, stop, back, propeller, normal, airplane, side
Clearance and radio	controller, clearance, frequency, radio, call, told, asked, contact, clear, communication, back, tower, air traffic control, called, control, heard
Fuel and Fire	engine, fuel, emergency, landing, light, zzz (sleep), problem, checklist, tank, system, pressure, power, declared, fire, procedure, failure
Speed and turbulence	knot, speed, aircraft, flap, degree, autopilot, control, captain, first officer, power, turbulence, normal, wind, warning, airspeed, foot
Time	flight, time, day, hour, crew, captain, minute, company, duty, trip, weight, first officer, operation, pilot

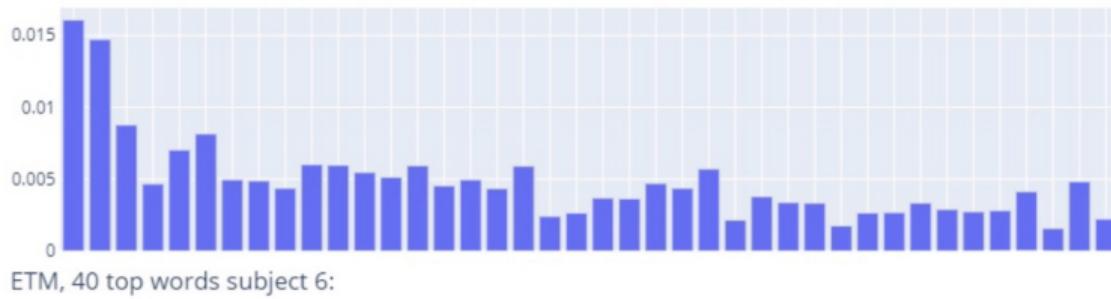
Comparison ETM/LDA

- Top words for topic 6 ETM :
maintenance, aircraft, check, mel, mechanic, found, item, part, zzz, problem, flight, logbook, manual, inspection, system, word, installed.
- Top words for topics 1 LDA :
maintenance, aircraft, check, found, mechanic, problem, flight, inspection, installed, time, logbook, manual, item, work, part, mel.

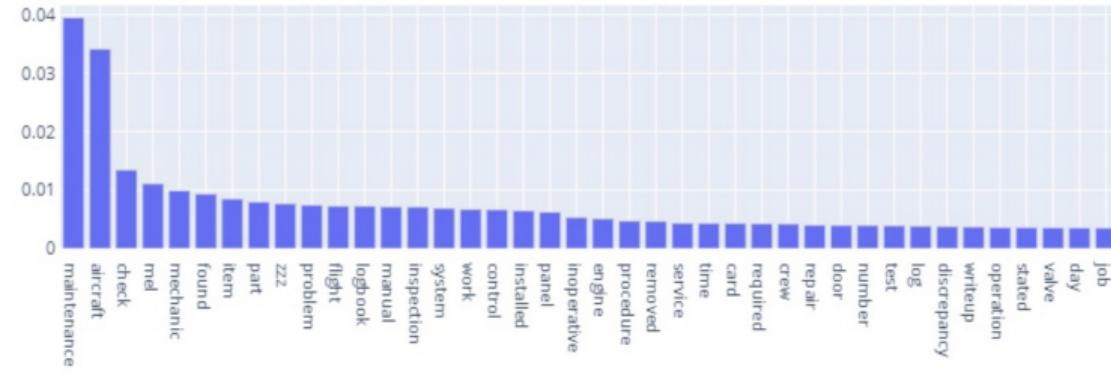


Comparison ETM/LDA

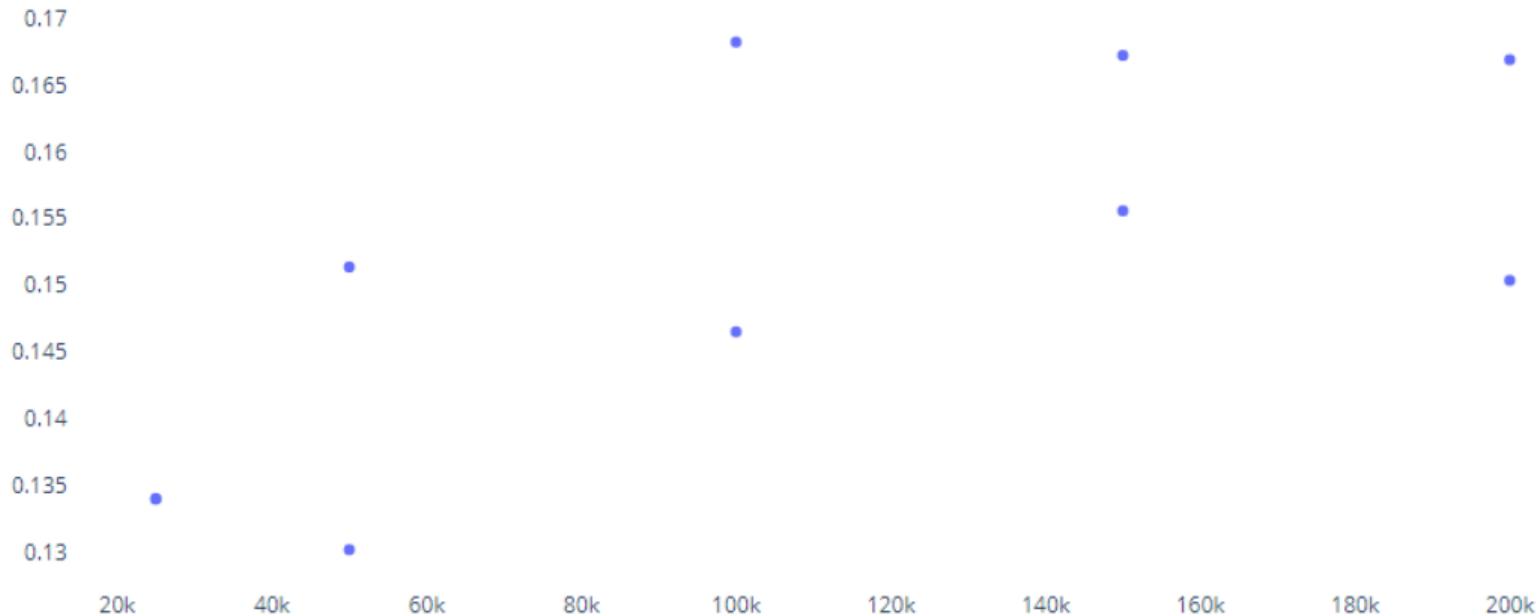
LDA, 40 top words subject 1:



ETM, 40 top words subject 6:



Topic Coherence



Zineb Bennis

The Embedded Topic Model

