

Topic Modelling and Sentiment Analysis on Uber Drivers' Conversations

SMM694 – GROUP 7





TABLE OF CONTENTS



01 GOAL & METHODS



02 TOPIC MODELLING



03 SENTIMENT
ANALYSIS &
RECOMMENDATION



04 STATISTICAL POST-
PROCESSING



AIM

Identification of the problems Uber drivers talk about concerning technology development and the affective correlates of these problems

METHODOLOGY



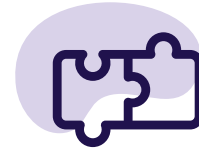
TOPIC MODELLING

Identification of the most relevant themes in the corpus of text containing Uber drivers message's exchange



SENTIMENT ANALYSIS

Identification of the affective states and emotions in uber drivers' conversation



STATISTICAL POST-PROCESSING

Identification of the impact of development in Uber technology on affective states and emotions in uber drivers over times and stages

DATASET

The Complaints community within the Uber People Forum has been used to perform web scraping. Particularly, 12,631 articles have been used to perform the following analysis.

The reason behind the choice of this community relies on the fact that through the analysis of articles concerning Uber drivers' complaints, the company can identify and strategically plan to overcome the main weaknesses regarding the technologies or the general problems contributing to employees' job unsatisfatcion.



TOPIC MODELLING

Model Results

TOPIC 0

[(0,
'0.053*"ride" + 0.037*"trip" + 0.029*"driver" + 0.029*"pay" + 0.025*"time" + 0.023*"lyft"
+ 0.023*"drive" + 0.019*"rider" + 0.019*"hour" + 0.018*"request")],

TOPIC 1

(1,
'0.029*"pax" + 0.027*"minute" + 0.021*"cancel" + 0.020*"pick" + 0.020*"drive" +
0.019*"wait" + 0.018*"app" + 0.018*"rating" + 0.015*"time" + 0.015*"start"),

TOPIC 2

(2,
'0.041*"car" + 0.023*"people" + 0.016*"tip" + 0.011*"good" + 0.011*"bad" +
0.011*"passenger" + 0.010*"thing" + 0.009*"feel" + 0.008*"nt" + 0.008*"order"),

TOPIC 3

(3,
'0.104*"uber" + 0.045*"driver" + 0.014*"email" + 0.012*"support" + 0.010*"account" +
0.010*"company" + 0.009*"issue" + 0.008*"send" + 0.008*"year" + 0.008*"service")]

Model Interpretation

TOPIC 0 - Financial Issues

Noticeable, Lyft is a Uber's competitor, which might indicate that Uber drivers complain about pay per ride, pay per hour, number of requests received or number of trips made compared to competitors. This topic indicates that Uber drivers are financially unsatisfied and company's improvements are required.

TOPIC 1 - Pre-Trip Issues

Uber drivers complain about cancellation rate, waiting rate, the app technology matching a customer with a low rating. Particularly, "minute" and "time" could be correlated to problems concerning the Uber GPS technology choosing a longer route or failing in the identification of traffic routes, so making a trip to last more minutes.

TOPIC 2 - Post-Trip Issues

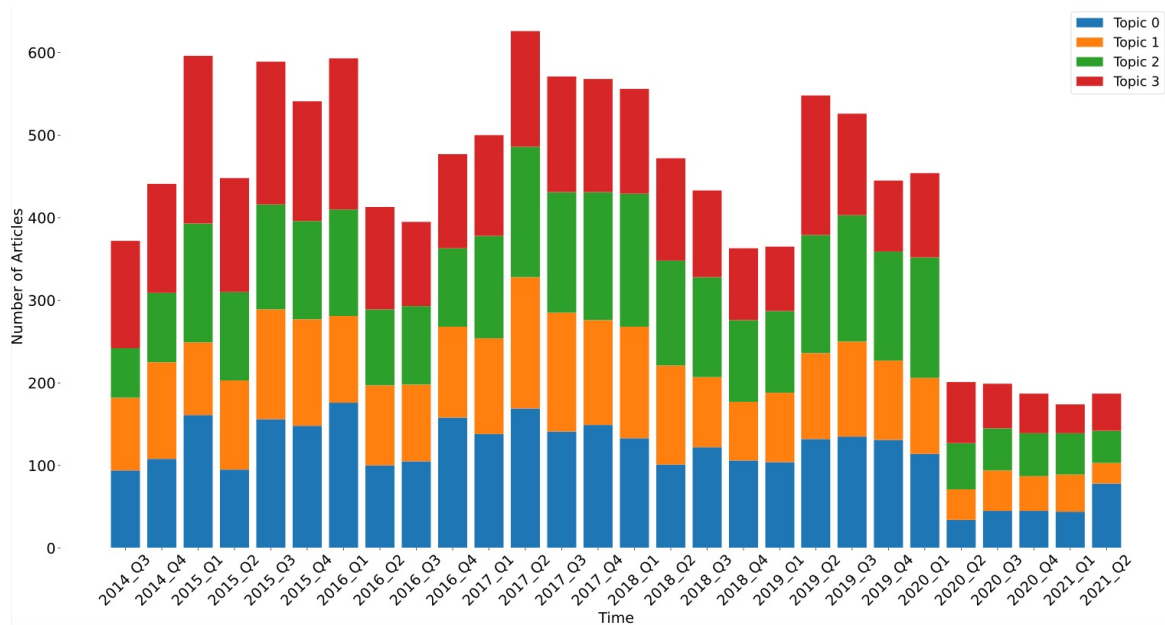
It seems that Uber drivers cares about tips from clients. Also, drivers rate people as "good" or "bad" clients, who probably affect their job satisfaction in terms of feelings concerning a trip. Drivers' low ratings guarantee that the same match will be avoided as a consequence of bad experience with passengers.

TOPIC 3 - Drivers' General Enquiries

Uber drivers' general complains concern the company's support or services. It could concerns delays from the support team in replying to drivers' email concerning issues with the account. From the employees' perspective, it seems that Uber should improve their support and service teams to improve the drivers' job satisfaction.

TOPIC MODELLING

Graph 1 – Topics' Predominance Over Time



Graph 1 shows the predominance of the topics (0, 1, 2, 3) analysed quarterly over the period between 2014 and 2021.

Since the UberPool and UberFresh launch in 2014, the Uber trip rate constantly increased until 2019 which explains the complaints predominance in the given period. Particularly, analysing the graph from a seasonality perspective, it is possible to notice that more complaints have been recorded in the first quarter of each year, except in 2017. The increase in complaints from 2016 to 2017 is explainable as the number of active users increased from 19 million in 2016 to 68 million in 2017.

The decrease in the complaint's prevalence from Q1 to Q2 in 2020 has not been generated by financial, services or support improvements in the company but by the reduction in number of trips from 1.6 billion (Q1) to 737 million (Q2) due to the Covid-19 pandemic.

TOPIC MODELLING

The following recently posted two text data have been manually collected from the Complaints Community to test that the model performs well with unseen data:

"
Scrubber emailed me and sent msg thru app on Friday July 2nd; \$100 for 3 completed rides over 4th of July weekend. Figured do 3 quick trips and then done and go home w \$100 + fares. Scruber paid the fares but not the promo. I sent msg & they replied 'due to some outages, we are experiencing delays in response times & prioritizing emergencies.... That was 2 days ago & still no response despite repeated msgs to them. Scruber strikes again.'"

Model Results

[(0, 0.350480109739369), (1, 0.22565157750342935), (2, 0.1598079561042524), (3, 0.26406035665294925)]



Noticeable, the main topic of the article is related to topic 0, financial issues. The model successfully classified the article to topic 0 as in the result topic 0 is associated with the highest probability!

"
Anyone else get this I'm in Chicago. No way I am waiting 7 minutes for a pax. Since Uber has not been paying for the first 2 minutes after arrival, I am only waiting those 2 minutes from this point forward, then leaving. Uber knows that riders are waiting longer for drivers than ever before; now they are going to give them 7 MORE minutes after we arrive?? Good luck with that Uber and pax.'"

Model Results

[(0, 0.21987480438184664), (1, 0.3591549295774648), (2, 0.2136150234741784), (3, 0.20735524256651017)]



Noticeable, the main topic of the article is related to topic 1, mainly pre-trip issues. The model successfully classified the article to topic 1 as in the result topic 1 is associated with the highest probability!

SENTIMENT ANALYSIS

Semantic Axis Method (SAM)

The aim is to analyze people characterized by actively hostile speech and behaviors compared to people with a passively submit attitude

STEP 1 : Harvard General Inquires Category Dictionaries (Hostile, Submit, Active, Passive)

STEP 2 : got word embedding for each word included in dictionary

STEP 3 : got the positive and negative centroids, mean word vector associated with each tokenized word

STEP 4 : got the semantic axis

$(\text{Hostile} + \text{Active}) - (\text{Submit} + \text{Passive})$

STEP 5 : analyzed the position of words in the semantic axis

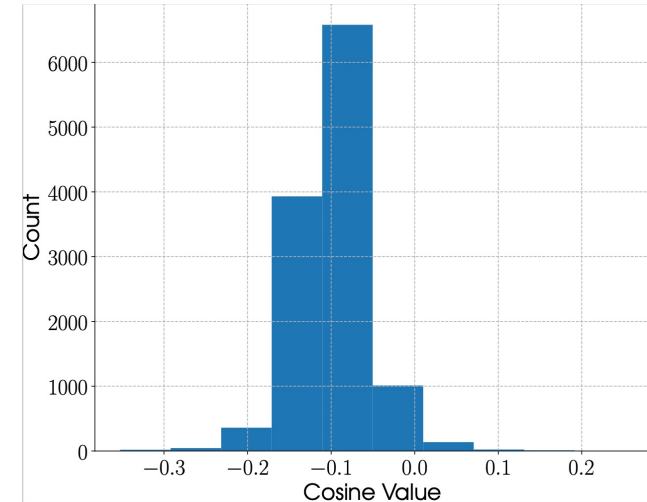


Interpretation

Graph 2 shows that the average cosine similarity is less than 0, meaning that most of the articles in the community tend to be characterized by a *submit* and *passive* communication style. Many researches and statistics demonstrate that people practise passive and assertive communication style on forums and social medias. In fact, users tend to maintain a passive and vulnerable attitude. *K. Hemmings-Jarrett, J. Jarrett and M. B. Blake (2017)* demonstrated that individuals differ ideologically and behaviorally when online or offline. As the nature of discussions and communication styles evolve, so do the dynamics within collective groups. Mainly, changes in engagement can be influenced by a more detailed choice of words used in online discussions.



Graph 2 – Average Cosine Similarity



SENTIMENT ANALYSIS

SAM Results and Limitation



Positive Cosine Similarity : Hostile and Active

The greatest values for positive average cosine similarity*

"Uber block system needs to review. They blocked drivers before any proven guilty."

"Uber programmers strike again. They moved the "X" for declining a ride partially behind the clock on iPhones. Incompetence? Maybe. Intentional? Probably."

"I'm not a number.. I'm a free man."

This phrase is highly hostile and active!

Negative Cosine Similarity: Submit and Passive

The lowest values for negative average cosine similarity*

"Doe anyone know the list of comfort vehicles in Kansas City? My Honda CRV is not qualified anymore. Thank you."

This phrase is very submitted and passive!

"You're not my friend, Mia. You appreciate the support? I contacted you, to do a job. What is this Lyft love you speak of? GFY with a rusty cheese grater."

"Basically a 1 MI bonus per ride. UBER, you are less and less disappointing the more I realize; you're evil and stoopit. "

Limitation: given that Harvard Inquires Category dictionaries have been used to perform sentiment analysis modelling, the model fails to recognise typo or slang words and metaphors. In fact, even though this phrase shows hostile speech, the model classified it with a negative cosine similarity.

*Average Cosine Similarity from each tokenised word in the article

SENTIMENT ANALYSIS

Recommendation



Uber should prioritise the tracking concerning hostile and active information:

- Many studies demonstrated that the sentiment of a message on social media affect its virality, defined as the volume and the speed of message diffusion. Particularly, *Tsugawa and Ohsaki* (2015) analysed the speed of diffusion based on message sentiment that revealed how the polarity of message sentiment affects its virality. Results showed that negative messages are likely to be reposted more rapidly and frequently than positive and neutral messages (for example, Donald Trump's tweets spread quickly and led to insurrection at Capitol Hill).
- Over the past years, interest in online hate speech detection and, particularly, the automation of this process has continuously grown, along with the societal impact of the phenomenon. Investing and employing advanced technologies to detect hate speech among Uber drivers' forum communities is crucial for the company because it would allow to detect its employees' job satisfaction as well as their sentiment and identification with the company's culture and continuous technology development. The implementation of Natural Language Processing tools would allow the Service, Support and Management teams to strategically plan ways to improve their employees' wellbeing and job satisfaction.

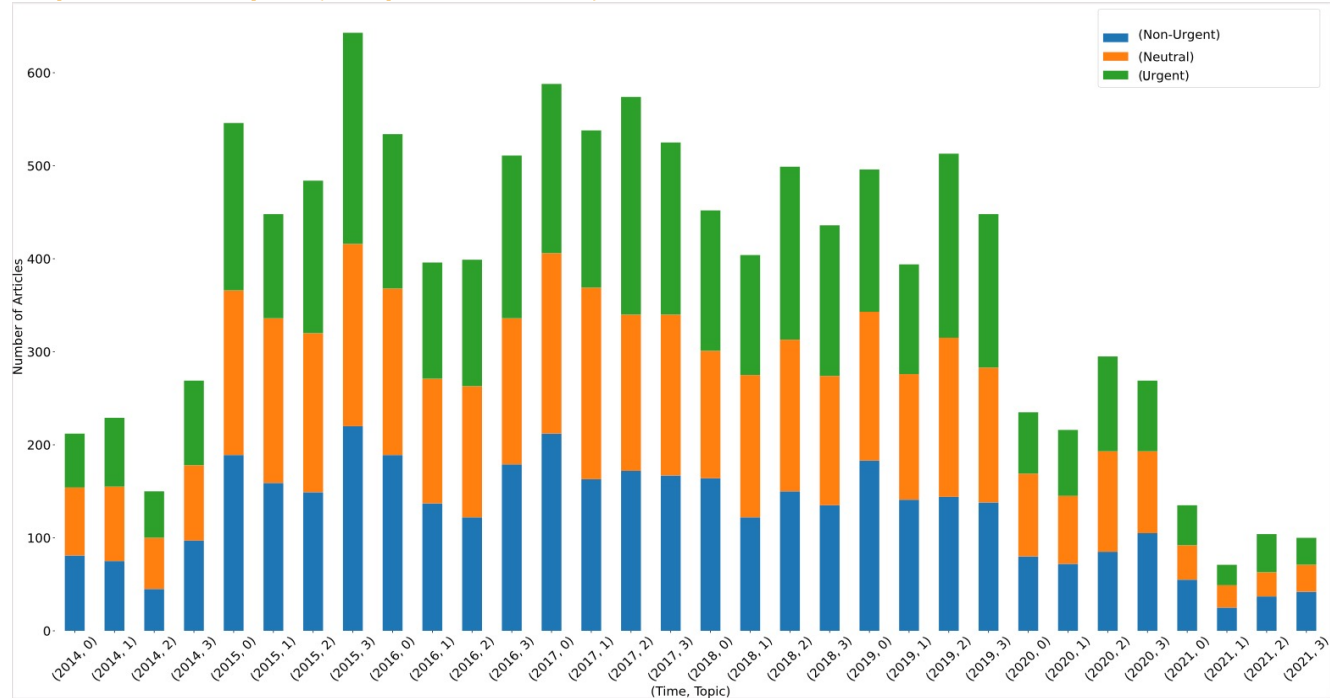


STATISTICAL POST-PROCESSING



Graph 3 – How Topics (Complaint's Themes) changed over Time

Graph 3 shows topics (0, 1, 2, 3) based on three urgency levels (Non-Urgent, Neutral and Urgent), that has been calculated based on whether the article showed or hostile and active, or neutral, or submit and passive characteristics, over the period between 2014 and 2021. Particularly, the topic seems to be characterized by a normal distribution concerning the urgency levels, as none of the level is predominant over time, but all three levels are homogenously prevalent for each topic and over time.



STATISTICAL POST-PROCESSING



Graph 3 - Interpretation

Noticeable, between 2014 and 2019, complaints concerning topics (0, 1, 2, 3) increased dramatically, but from 2020 complaints drastically decreased which is probably due to the reduction in Uber trip rate by 70% in US and by 77% in two of its largest European markets (London and Paris) during mid-March 2020 because of the Covid-19 pandemic.

Topic 0 denoting the Uber drivers' complaints regarding financial issues have been the predominant problems discussed in the Uber People forum from 2014 until 2018. However, after a decrease in terms of predominance in 2019, financial issues still represents a major topic in the drivers' conversations in 2021. A PwC research reported that in 2018 Uber changed its pay structure in some cities to prioritise trip time over distance driven, which led to an increase in the pay-out for short trips, and adopted a beta driver rewards program providing cash back on gas purchase, which explains the decrease of driver's complaints about topic 0 in 2019. Even though, Uber offered to 70,000 UK drivers a minimum hourly wage, holiday pay and pensions, the company's employees hourly wage is roughly at the 10th percentile of all salary workers' wages. In fact, Uber drivers earn less than what 90 percent of workers earn in most of the company's markets, which explains the predominance of topic 0 in 2021.

The predominance of topics 1, 2 and 3, mainly pre and after trip and general issues, characterize the period between 2014 and 2020. However, the decrease in these topics' prevalence is due to development in the Uber Service & Support management. In 2020, Uber launched the Hyderabad Tech Centre, an in-house team working on technologies, aiming to solve the fundamental problems by using psychological and social science techniques to support riders and drivers and influence when, where and how long drivers work. Uber is investing in data science and analytics to analyse problems' patterns and to determine where and how the company should strategically focus support and service developments and improvements to make employees satisfied.

REFERENCES

- Blackall, M., 2021. *'A lot are sceptical': Uber drivers' cautious welcome over worker status.* [online] the Guardian. Available at: <<https://www.theguardian.com/technology/2021/mar/18/a-lot-are-sceptical-uber-drivers-cautious-welcome-over-worker-status>> [Accessed 11 July 2021].
- Bonet, O., Cuadros, M., García-Pablos, A. and Perez Miguel, N., 2018. *Hate Speech Dataset from a White Supremacy Forum.* [online] ResearchGate. Available at: <https://www.researchgate.net/publication/327621135_Hate_Speech_Dataset_from_a_White_Supremacy_Forum> [Accessed 11 July 2021].
- Exploring Your Mind. 2019. *Do You Use Passive Communication? - Exploring your mind.* [online] Available at: <<https://exploringyourmind.com/do-you-use-passive-communication/>> [Accessed 9 July 2021].
- Hemmings-Jarret, K., Jarret, J. and Blake, M., 2017. *Evaluation of User Engagement on Social Media to Leverage Active and Passive Communication.* [online] Ieeeexplore.ieee.org. Available at: <https://ieeexplore.ieee.org/abstract/document/8029234?casa_token=8IkX4dZHc8cAAAAA-BvSwEVi-AeKvVkJKa7ee_r1cUR5x62t1Baja-SB9HHzp_fHdFfz8Y4ysmJCqUQC0PWZuheal> [Accessed 12 July 2021].
- Mathur, N., 2020. *Ubersets up fintech team in Hyderabad.* [online] mint. Available at: <<https://www.livemint.com/companies/news/uber-establishes-uber-money-team-in-hyderabad-11581409688769.html>> [Accessed 10 July 2021].
- Mishel, L., 2018. *Uber and the labor market: Uber drivers' compensation, wages, and the scale of Uber and the gig economy.* [online] Economic Policy Institute. Available at: <<https://www.epi.org/publication/uber-and-the-labor-market-uber-drivers-compensation-wages-and-the-scale-of-uber-and-the-gig-economy/>> [Accessed 11 July 2021].
- Morshed, S., Khan, S., Bari Tanvir, R. and Nur, S., 2021. *Impact of COVID-19 pandemic on ride-hailing services based on large-scale Twitter data analysis.* [online] ScienceDirect. Available at: <<https://www.sciencedirect.com/science/article/pii/S2226585621000200>> [Accessed 12 July 2021].
- Rapier, G., 2018. *Uber's new rates are designed to make driver payout more consistent, but some say it has decreased their earnings.* [online] Business Insider. Available at: <<https://www.businessinsider.com/new-uber-pay-rates-hurt-earnings-drivers-say-2018-12?r=US&IR=T>> [Accessed 10 July 2021].
- Tsugawa, S. and Ohsaki, H., 2015. *(PDF) Negative Messages Spread Rapidly and Widely on Social Media.* [online] ResearchGate. Available at: <https://www.researchgate.net/publication/301417650_Negative_Messages_Spread_Rapidly_and_Widely_on_Social_Media> [Accessed 10 July 2021].

**THANK YOU FOR
YOUR ATTENTION!**

