A longitudinal study about gradual changes in the Iranian Online Public Sphere pre and post of ‘Mahsa Moment’: Focusing on Twitter

Sadegh Jafari
Iran University of Science and Technology
sadegh_jafari@comp.iust.ac.ir

Amin Fathi
Iran University of Science and Technology
aminbaybon@gmail.com

Abolfazl Hajizadegan
University of Tehran
a.hajizadegan@ut.ac.ir

Amirmohammad Kazameini
Vector Institute
Amirmohammad.kazemeini@vectorinstitute.ai

Sauleh Eetemadi
Iran University of Science and Technology
sauleh@iust.ac.ir

Abstract

Mahsa Amini’s death shocked Iranian society. The effects of this event and the subsequent tragedies in Iran not only in realspace but also in cyberspace, including Twitter, were tremendous and unimaginable. We explore how Twitter has changed after Mahsa Amini’s death by analyzing the sentiments of Iranian users in the 90 days after this event. Additionally, we track the change in word meaning and each word’s neighboring words. Finally, we use embedding clustering methods for topic modeling.

1 Introduction

Clashes broke out throughout Iran after Mahsa Amini, a 22-year-old Kurdish Iranian woman, died on 16 September 2022 after being detained by "morality police" and taken to a "re-education center" allegedly for not abiding by the country’s conservative dress code. Although Iranian officials have said that Mahsa Amini died of a heart attack, according to a United Nations report, Amini collapsed at a detention center in Tehran on 13 September 2022, in the custody of Iran’s morality police and then died three days later after being transferred to a hospital. The report said Amini was "severely beaten" by Iranian authorities during her detention. (UN, 2022)

During a crisis, people and the media take over the flow of information, process it, and react to it. The effects of this situation may harm the mental health of the affected population. Mahsa Amini’s death caused widespread reactions on several social networks among Iranian and non-Iranian users; for example, although Twitter is banned in Iran and people are having trouble accessing it, MahsaAmini and its Persian-translated hashtag became one of the most repeated hashtags on Twitter and broke a historical record.(BBC, 2022)

The content effects of this tragic incident on Twitter, especially among Iranian users during the 90-day period following Mahsa Amini’s death, were analyzed. Twitter data, including tweets with the hashtag "#mahsa_amini" and relevant hashtags, were collected from September 21, 2022, through December 19, 2022. The dataset comprises a total of 1,944,056 tweets in various languages, primarily Persian and English. After preprocessing the tweets, the Persian dataset was utilized to assess the sentiment of Iranian users and illustrate how events during this period, such as the onset of executions, impacted the emotions of Iranian Twitter users. Subsequently, word embeddings were examined to assess the extent to which the meaning of Persian words in tweet content evolved due to the societal changes triggered by Mahsa Amini’s death. Cosine similarity was computed between the embedding vector of each word using the original BERT(Devlin et al., 2018) model and a fine-tuned BERT model for this purpose.

Moreover, an analysis of neighboring words for each word before and after Mahsa Amini’s death was conducted. The findings concerning the key protest slogan in Iran, "woman, life, freedom," revealed significant changes in the neighboring words of "woman" and "life" on Twitter following Mahsa Amini’s death. However, this incident did not lead to notable alterations in the neighboring words of "freedom." The paper’s final section employed topic modeling as an unsupervised machine-learning technique to automatically cluster words within the English and Persian tweet datasets obtained from Twitter.

2 Related Works

Previously, researchers have computationally investigated diachronic language change in various ways. Sagi et al. (2009) use a variation of latent semantic analysis to identify semantic change of specific words from early to modern English. Mihalcea and Nastase (2012) take a supervised learning approach and predict the time period to which a word belongs given its surrounding context. Kim et al.

Proceedings of the 4th Workshop on Computational Approaches to Historical Language Change, pages 46–52
December 6, 2023

Language	Number of tweets
Persian	1,445,537
English	317,046
Arabic	54,106
Urdo	28,880
German	13,919

Table 1: the number of tweets of the five most frequent languages in the dataset.

(2014) use word2vec (Mikolov et al., 2013) to assay the change of words across time. Hamilton et al. (2016) develop a robust methodology for quantifying semantic change by evaluating word embeddings (PPMI (Marek et al., 2011), SVD (Stewart, 1993), word2vec) against known historical changes. Xie et al. (2020) investigate the change in moral sentiment among the public using longitudinal corpora. We use a transformer base language model to calculate word embeddings to specify our dataset's context. Also, we use a model to predict the semantics of sentences, which can help us find the reason for the change of words.

3 Data

We use the "snscrape" python library to crawl Twitter data tweets "mahsa_amini" and relevant hashtags from 2022-9-21 through 2022-12-19 . Our dataset contains about 2 million tweets in different languages. To pre-process the tweets, we removed usernames, hashtags, and URLs. According to this pre-processed dataset and the results obtained from the "langdetect" python library, the number of tweets of the five most frequent languages in the dataset is illustrated in Table 1.

4 Sentiment analysis

For this aim, we use “persiannlp/mt5-small-parsinlu-sentiment-analysis” transformer-based model (Daniel Khashabi, 2020) to predict the sentiment of sentences in Persian tweets; the chart of the results is shown in figure 1. The events that have happened have led to the growth of negative feelings among Iranian Twitter users. For example, as can be seen in the chart, negative sentiments among Iranian users increased significantly from December 5 until December 10. Calls for strikes and protests on December 5, 6, and 7, as well as the media coverage of the execution of "Mohsen Shekari", who was the first known executed person over anti-government protests, on December 8,

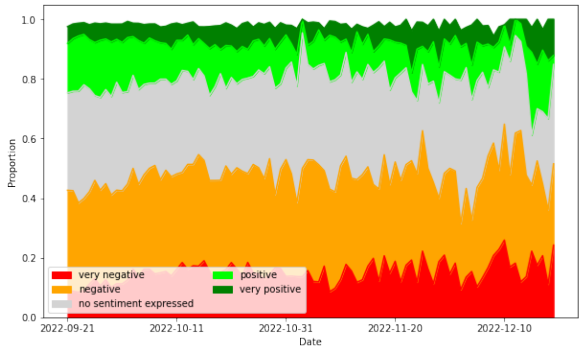

Figure 1: Illustration of result chart for sentiment analysis, the horizontal axis represents 90 days after Mahsa Amini's dead, and the vertical axis represents the population percentage. In the above chart, red, orange, gray, pale green, and deep green, respectively, represent "very negative sentiment", "negative sentiment", " no sentiment expressed", "positive sentiment" and "very positive sentiment".

caused a wave of negative emotions among Iranian users so that the proportion of negative emotions in Persian tweets reached its maximum level in these 90 days. There are also some impulses in the chart at certain times; for example, on November 25, videos of shooting protesters in Zahedan city and anti-riot police forces celebration in the streets after the victory of the Iranian national football team against Wales in Qatar's world cup provoked many adverse reactions on Twitter.

5 Word embedding analysis

In this section, the analysis of word embeddings after and before Mahsa Amini's death is undertaken. First, we calculate the embedding of the Persian tweets using the "HooshvareLab/BERT-base-parsBERT-uncased" (Mehrdad Farahani, 2020) transformer-based model, we use BERT because it offers contextualized embeddings, enabling us to analyze how various word senses change in meaning across different contexts. BERT model calculate the context-aware embedding, so each word can have multiple embeddings depending on the context of the text. We calculate the average of all these word embeddings to get each word's unique embedding.

5.1 Find embedding of words before fine-tuning the model

First, we calculate embedding for each word in the corpus and then remove the tokens that are stop-words or function-words, or subwords(BERT creates some subwords in its tokenizer, for example, ##ing). After that, we choose the 1000 most frequent tokens.

5.2 Find embedding of words after fine-tuning the model

Then we fine-tune the BERT model with a sample of 300,000 Persian tweets from 2022-9-21 through 2022-12-19 for three epochs with learning-rate=2e-5 and weight-decay=0.01. After that, we repeat step 5.1 and calculate the embeddings of each word.

5.3 Calculate self-similarity

Finally, due to this method's popularity, we calculate the cosine similarity between the embedding vector of the word with the original BERT model (known as emb_{before}) and the embedding vector of the word with a fine-tuned BERT model (known as emb_{after}). The similarity metric is defined as

$$sim = \frac{emb_{before} \cdot emb_{after}}{||emb_{before}|| \; ||emb_{after}||}. \tag{1}$$

If the self-cosine similarity of a word is 1, that word is not changed at all. However, if the self-cosine similarity of a word is near 0, it shows that it is changed so much after this period. Considering the interconnected relationship between language and culture in every society, tracking the changes in the meaning of words is essential for analyzing society's culture. The most significant changes were in profanity; for example, the function of the sexual slurs changed, and Twitter users used them in a political context throughout these 90 days. During emotionally charged periods, like protests or reactions to tragic events such as Mahsa Amini's death, individuals may vividly express their emotions, occasionally resorting to profanity. Emotions such as anger, frustration, and grief can increase online profanity usage, allowing individuals to vent their feelings. Furthermore, the transformation of profanity's role from a personal expression to a tool in political protests highlights language's adaptability to shifting societal dynamics. This linguistic evolution mirrors the changing landscape of public discourse amid social and political unrest, with individuals increasingly using strong language to underscore their positions on contentious issues. Also, the meaning of words such as "woman", "life", "freedom" and "protest" changed a lot. Table 2 displays words with the most significant meaning changes according to the self-cosine similarity metric.

Persian word	English translation	cosine similarity value
ژیان	life (in Kurdish)	0.394
شیش	six	0.440
کون	ass	0.496
گای	f*ck	0.526
کیر	d*ck	0.530
خایه	male balls	0.535
ژن	woman (in Kurdish)	0.549
نذاری	don't allow	0.559
جنده	bitch	0.562
آبادی	prosperity	0.567

Table 2: Ten words with the lowest self-cosine similarity scores, which are derived from a pool of the 1000 most frequently used words.

Persian word	English translation	NSV value
ژیان	life (in Kurdish)	0.045
ژن	woman (in Kurdish)	0.043
گای	f*ck	0.041
پشم	fur	0.035
صدا	voice	0.035
عن	sh*t	0.035
آبادی	prosperity	0.035
گوز	fart	0.035
کون	ass	0.034
کیر	d*ck	0.034

Table 3: Among the 1000 most frequently used words, ten words with the highest NSV scores are identified. The NSV metric typically ranges from 0 to 1. However, for these ten words in the table, their NSV values are extremely low, nearing 0. This is due to the NSV metric's nature, as it calculates a word's similarity to itself. When used to compare two nearly unrelated words, the metric's value significantly increases.

5.4 Calculate neighbor square value

In this section, we want to use another way to measure each word's embedding space changes. The problem with self-cosine similarity is that a word and its neighbors might move to new same neighborhood points in the embedding space, so in this situation, self-cosine similarity shows this word changed, but we know that only our coordinate is changed. We should compare embeddings in the same coordinate using a new metric, NSV (neighbor square value).

In algorithm 1, we want to find k words most similar to the desired word. First, we calculate the cosine similarity for all words with input words and then save them in the neighbors dictionary(the key is a word, and the value is cosine similarity). Finally, we return the k words with the highest cosine similarity with our word.

Persian word	English translation	NSV rank	cosine similarity rank
شصت	sixty	981	77
هفتاد	seventy	951	91
سی	thirty	973	128
هشتاد	eighty	918	89
دویست	two hundred	888	100
سیصد	three hundred	882	103
نود	ninety	854	78
چهارصد	four hundred	876	118
بیست	twenty	915	160
پنجاه	fifty	863	137

Table 4: Rankings pertain to the 1000 most commonly used words, as evidenced by the substantial semantic changes observed at higher ranks within the cosine similarity matrix. In contrast, the NSV rank positions these words at the bottom, indicating minimal semantic alterations.

Algorithm 1 Find k nearest neighbors

Input: word, embeddings, k
Output: neighbors

$neighbors \leftarrow \emptyset$
for token **to** embeddings.tokens **do**
  $cs \leftarrow CosineSimilarity(word, token)$
  $neighbors \cdot add((token, cs))$
end for
return $neighbors \cdot SortByValue(k)$

Algorithm 2 get neighbor cosine similarity matrix

Input: word, embeddings, k
Output: matrix

$$matrix_{kk} = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}$$

$neighbors \leftarrow FindNearestNeighbors(word, embeddings, k)$

for i=0 **to** k **do**
  for j=0 **to** k **do**
    $vector_i \leftarrow neighbors[i] \cdot vector$
    $vector_j \leftarrow neighbors[j] \cdot vector$
    $matrix[i][j] \leftarrow CosineSimilarity(vector_i, vector_j)$
  end for
end for
return $matrix$

Word	**Neighboring words**	
	before	**after**
(Woman) زن	(women) زن‌ها (man) مرد (girl) دختر (lady) خانم (sister) خواهر	(man) مرد (life) زندگی (women) زن‌ها (woman in Kurdish) ژن (girl) دختر
(Life) زندگی	(future) آینده (home) خانه (woman) زن (family) خانواده (country) وطن	(woman) زن (homeland) میهن (life in Kurdish) ژیان (man) مرد (prosperity) آبادی
(Freedom) آزادی	(liberation) آزادی (free) آزاد (justice) عدالت (peace) صلح (victory) پیروزی	(liberation) آزادی (free) آزاد (victory) پیروزی (justice) عدالت (peace) صلح

Table 5: Analyzing the top 5 neighboring words for 'woman,' 'life,' and 'freedom' before and after Mahsa Amini's death reveals significant changes.
While 'woman' and 'life' were influenced, 'freedom' remained consistent. This reflects the historical significance of freedom movements in Iran, dating back to the 1979 revolution. Before her passing, Twitter discussions about 'woman' covered diverse topics, including lifestyle and relationships. After her death, the focus shifted to critical subjects related to her case and women's rights. 'Women in Kurdish' among the related words shows the broader discussion encompassing regional and ethnic aspects.

In algorithm 2, we want to find the neighbor cosine similarity matrix. Each element of this matrix shows the cosine similarity between $neighbor_i$ and $neighbor_j$ of the input word.

$$NSV = \frac{\sum_{i=0}^{k} \sum_{j=0}^{k} (m_b[i][j] - m_a[i][j])^2}{k^2} \quad (2)$$

Finally, we calculate the NSV. m_b is matrix before finetuning and m_a is matrix after finetuning. NSV is between 0 to 1. 0 means that the embedding space of our word does not change, and 1 means that our word has completely changed. Table 3 displays words with the most significant meaning changes according to the NSV metric. Comparing the two methods, as discussed earlier, makes it clear that the NSV metric provides more consistent results. For instance, as seen in table 4, the self-cosine-similarity metric for numeric words suggests significant changes, implying that numeric words undergo substantial alterations. However, the NSV values for numeric words are remarkably low, indicating minimal changes, aligning with our expectations.

6 Topic Modeling

For topic modeling, we used the LDA (Blei et al., 2003) technique for English tweets and HDBSCAN (McInnes et al., 2017) for Persian tweets; the reason for this is the better performance of the LDA method on English tweets and the HDBSCAN method on Persian tweets. We remove the numbers, double spacing, and stopwords to clean the tweet dataset using the NLTK library. After that, we convert the text tweets into vectors. We also filtered any words that appeared in more than 90% tweets or less than 25 tweets. For the number of clusters, we ran the model for k = 5,7,10 clusters for English tweets and k = 4, 6,9 clusters for Persian tweets, and the results showed better performance on k=7 for English tweets and on k=4 for Persian tweets, which are illustrated in table 6 and table 7 .

For English tweets, the first topic related to the news about what has passed during this period, which is why the words Republic and Islam (which represent the Islamic Republic) are at the top of this topic. "massacre," "rape," "torture" and "shoot" are also among the frequent words of this cluster. The second topic expresses gratitude for the support of the international community; as seen in table 6, the words "thank" and "support" are among the five most used words in this category. Other keywords in this cluster are "love," "need," "dear," and "appreciate".Topic number 3, more than anything, deals with asking for help from the international community; the words "help", "us", "internet", "human", "right", "world" and "support" are among the most frequent words in this topic.
Topic number 4 is also related to the context and motto of what happened in Iran, as can be seen in table 6, "woman", "life" and "freedom" are in the five most frequent words in this cluster, "brave", "right", "together" and "free" are other key words of this cluster. The fifth topic expresses violence and oppression, "arrest", "beaten", "execute", "gestapo", "moral", "police" and "danger" are the most repeated words of this topic. The sixth topic expresses a more general aspect of protests; words such as "prison", "student", "IRGC", "university", "dictator", and "street" are among the other frequent words of this cluster of words. And finally, the last topic that discussed the death of Mahsa Amini, words such as "sharia", "hijab", "mandatory", "moral", "police", "kill", "Mahsa", "Amini" and "murder" are the most repeated words in this cluster of words.

topic1	topic2	topic3	topic4	topic5	topic6	topic7
Islam	thank	people	freedom	arrest	protest	police
republic	Iranian	Iranian	woman	Islam	Iran	kill
kill	support	please	Iran	beaten	force	Iranian
people	people	human	life	secure	Islam	brutal
regime	voice	help	fight	hijab	death	girl

Table 6: 5 most frequent words for each topic in English tweets.

topic1	topic2	topic3	topic4
(woman) زن	(hope) آرزو	(Mahsa) مهسا	(people) مردم
(life) زندگی	(victory) پیروزی	(my sister) خواهرم	(blood) خون
(freedom) آزادی	(free) آزاد	(Amini) امینی	(life) زندگی
(man) مرد	(Iran) ایران	(Iran) ایران	(war) جنگ

Table 7: 4 most frequent words for each topic in Persian tweets, also, in the results, there was a topic related to tweets of numbers that are not mentioned in the table above; Twitter users have used these numbers for purposes such as mentioning the number of people killed and the days that have passed since Mahsa Amini's death.

In the results obtained in Persian tweets, the first topic contains tweets with the main slogans of the protesters, such as "woman, life, freedom" The second topic has hopeful content for the future, "hope", "victory" and "free" are among the most frequent words in this topic. The third topic included tweets directly related to Mahsa Amini's death and the last topic deals with Persian Twitter users' protests regarding Iran's current situation.

7 Conclusion and feuture work

So far, we have focused on textual analysis of Iranian Twitter accounts before and after 'Mahsa Moment'. We will complete these analyses in the future. In addition, another longitudinal study - as a complementary method – has been left for the future due to a lack of time. After completing and updating our databases about different characteristics of Iranian 'users' and 'influencers' on Twitter, we will test the significance of changes on both sides of 'Mahsa Moment'. The variables that we are gathering are as follows: location, gender, political tendency, number of followers and join date. In the next step, we will analyze the relationship network between the influencers. For this purpose, by using Gephi (Bastian et al., 2009), the relationships graph will be visualized, and the main communities within the Iranian space of Twitter will be detected according to the Louvain (Blondel et al., 2008) method. We also intend to calculate a topic prevalence chart, similar to the one presented in the study by Ebadi et al. (2021), and analyze its findings. We will employ DeLong et al. (2023) method instead of cosine similarity because this paper serves as a more accurate predictor than cosine similarity based on embeddings when using BERT in a sense-disambiguation related task. We also aim to conduct our analysis on the entire dataset of 2 million entries to ensure more accurate results.

Limitations

The research struggled with many limitations. One of the most important ones was the lack of access to appropriate computing resources such as GPU, especially for sentiment analysis (we only analyzed a sample of 100,000 tweets, while the total available tweets were almost 2 million, and the analysis of the embedding space was on 300,000 tweets and we were not able to analyze the whole data). Also, we were looking for gender analysis. For that, we need to crawl new data. But, we could not do this because of financial transaction restrictions due to Iran sanctions. Another limitation was the internet shutdown by the government after the protests which led to slowness and frequent interruptions of the research process.

References

Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. 2009. Gephi: an open source software for exploring and manipulating networks. In *Proceedings of the international AAAI conference on web and social media*, volume 3, pages 361–362.

BBC. 2022. hashtags, a viral song and memes empower iran's protesters.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008.

Siamak Shakeri Pedram Hosseini Pouya Pezeshkpour Malihe Alikhani Moin Aminnaseri Marzieh Bitaab Faeze Brahman Sarik Ghazarian Mozhdeh Gheini Arman Kabiri Rabeeh Karimi Mahabadi Omid Memarrast Ahmadreza Mosallanezhad Erfan Noury Shahab Raji Mohammad Sadegh Rasooli Sepideh Sadeghi Erfan Sadeqi Azer Niloofar Safi Samghabadi Mahsa Shafaei Saber Sheybani Ali Tazarv Yadollah Yaghoobzadeh Daniel Khashabi, Arman Cohan. 2020. ParsiNLU: a suite of language understanding challenges for persian. *arXiv*.

Katherine A DeLong, Sean Trott, and Marta Kutas. 2023. Offline dominance and zeugmatic similarity normings of variably ambiguous words assessed against a neural language model (bert). *Behavior research methods*, 55(4):1537–1557.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Ashkan Ebadi, Pengcheng Xi, Stéphane Tremblay, Bruce Spencer, Raman Pall, and Alexander Wong. 2021. Understanding the temporal evolution of covid-19 research through machine learning and natural language processing. *Scientometrics*, 126:725–739.

William L Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. Diachronic word embeddings reveal statistical laws of semantic change. *arXiv preprint arXiv:1605.09096*.

Yoon Kim, Yi-I Chiu, Kentaro Hanaki, Darshan Hegde, and Slav Petrov. 2014. Temporal analysis of language through neural language models. *arXiv preprint arXiv:1405.3515*.

Kenneth Marek, Danna Jennings, Shirley Lasch, Andrew Siderowf, Caroline Tanner, Tanya Simuni, Chris Coffey, Karl Kieburtz, Emily Flagg, Sohini Chowdhury, et al. 2011. The parkinson progression marker initiative (ppmi). *Progress in neurobiology*, 95(4):629–635.

Leland McInnes, John Healy, and Steve Astels. 2017. hdbscan: Hierarchical density based clustering. *J. Open Source Softw.*, 2(11):205.

Marzieh Farahani Mohammad Manthouri Mehrdad Farahani, Mohammad Gharachorloo. 2020. Parsbert: Transformer-based model for persian language understanding. *ArXiv*, abs/2005.12515.

Rada Mihalcea and Vivi Nastase. 2012. Word epoch disambiguation: Finding how words change over time. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 259–263.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Eyal Sagi, Stefan Kaufmann, and Brady Clark. 2009. Semantic density analysis: Comparing word meaning across time and phonetic space. In *Proceedings of the Workshop on Geometrical Models of Natural Language Semantics*, pages 104–111.

Gilbert W Stewart. 1993. On the early history of the singular value decomposition. *SIAM review*, 35(4):551–566.

UN. 2022. un experts demand accountability for death of mahsa amini, call for end to violence against women.

Jing Yi Xie, Renato Ferreira Pinto Jr, Graeme Hirst, and Yang Xu. 2020. Text-based inference of moral sentiment change. *arXiv preprint arXiv:2001.07209*.