

Did that happen? Predicting Social Media Posts that are Indicative of what happened in a scene: A case study of a TV show

Anietie Andy¹, Reno Kriz⁴, Sharath Guntuku², Derry Tanti Wijaya³, Chris Callison-Burch²
 Penn Medicine¹, Computer and Information Science, University of Pennsylvania², Computer Science, Boston University³,
 Human Language Technology Center of Excellence, Johns Hopkins University⁴
 Anietie.Andy@pennmedicine.upenn.edu, rkriz1@jh.edu, wijaya@bu.edu
 {ccb, sharathg}@cis.upenn.edu

Abstract

While popular Television (TV) shows are airing, some users interested in these shows publish social media posts about the show. Analyzing social media posts related to a TV show can be beneficial for gaining insights about what happened during scenes of the show. This is a challenging task partly because a significant number of social media posts associated with a TV show or event may not clearly describe what happened during the event. In this work, we do the following: (a) collect social media (Twitter) posts associated with a TV show, Game of Thrones, (b) identify the various scenes of the TV show, (c) for each scene, we annotate the social media posts (published during the time period of the scene) that are indicative of what transpired during the scene, and (d) we propose a method to predict the indicative social media posts in each scene. We show that for each of the identified scenes, with high AUC's, our method can predict posts that are indicative of what happened in a scene from those that are not-indicative. Based on Twitter's policy, we will make the Tweeter ID's of the Twitter posts used for this work publicly available; also, we will make the annotations available.

Keywords: Twitter, Game of Thrones, indicative, non-indicative

1. Introduction

Prior works have shown that while popular televised events such as TV shows are airing, some individuals interested in these events publish social media posts associated with the events (Shen et al., 2013; Andy et al., 2017; Andy et al., 2019; Andy et al., 2020). Some of these posts capture user reactions to the various scenes associated with the events. In this work, using social media data associated with an episode of a TV show, Game of Thrones season 7 (GoT), we aim to identify the various scenes associated with the show and use a machine learning model to detect the posts that describe part/all of what happened during each scene.

Similar to prior work (Panda et al., 2018), in this work, we define a scene as comprising of several shots focused on a particular idea/concept.

Prior work identified and summarized scenes associated with a TV show by selecting a few representative social media posts (5 posts) that were published during the timeline of the scenes (Andy et al., 2019). One of the challenges of this method is that some scenes are made up of several smaller sub-events and so selecting 5 representative posts might not capture all or most of what happened during the scenes. For example, Figure 1 shows an example of a scene from an episode of GoT; several sub-events happened in this scene e.g. two characters *Arya* and *Brienne of Tarth* spar while two other characters, *Sansa* (who is Arya's sister) and *Little Finger* watch.

To gain insights as to what transpired during an event from social media posts related to the event (and published while the event was airing), it is important to



Arya spars with Brienne

Figure 1: Example of an image and description of a scene in an episode of GoT (this was obtained from the Amazon prime video streaming service).

identify and delineate the posts that are indicative of what happened in each of the scenes of the event (i.e. *indicative posts*) from those that are not indicative (*non-indicative posts*). This is a challenging task partly because a significant number of social media posts associated with an event / TV show may not clearly describe what happened during the show, rather, they may express user reactions (e.g. memes) / opinions/thoughts/feelings of what happened during a scene. Below, we define *indicative* and *non-indicative* posts and show examples.

- **an indicative post:** is a social media post associated with an event/TV show - published during the timeline of a scene of the event, that describes part

of or all of what transpired during the scene

- a **non-indicative post**: is a social media post associated with an event/TV show - published during the timeline of a scene of the event, that is either unrelated to the scene or does not describe part of or all of the what transpired during the scene.

Examples of Indicative posts associated with a scene

- "woah this is sansa's first time actually seeing arya fight"
- "arya sparring with brianne" .
- "arya fighting one handed."

Examples of non-indicative posts associated with a scene

- "arya is awesome!"
- "brianne watch your step arya is a legend in bravos, a legend! "
- "Arya looked so much like Ned Stark during that training fight with Brienne. her clothes reminded me of Ned"

The examples above are indicative and non-indicative posts associated with the GoT scene in which the character Arya spars with *Brienne of Tarth*. From these examples: (a) the *indicative posts* capture parts of what transpired during this scene i.e. Arya sparred with *Brienne of Tarth*, Arya held her sword with one hand, and Sansa watched Arya and *Brienne of Tarth* sparring (b) the *non-indicative posts* capture the user reactions to this scene, however they do not clearly describe part of or all of what happened during the scene.

In this work, we aim to capture all or most of the sub-events associated with a scene. In order to do this, first we automatically identify the various scenes associated with the TV show, GoT, we label the *indicative posts* and *non-indicative posts* associated with each of these scenes, and train a machine learning model to predict the *indicative posts* from the *non-indicative posts* in each of these scenes. While it is possible that over the course of the TV show, users may publish social media posts about a scene after the scene occurred, in this work, we are focused on identifying indicative posts published during the timeline of a given scene.

This work makes the following contributions:

- We create a new data set made up of Twitter posts (tweets) associated with scenes of a TV show; we label these tweets as either indicative (*indicative post*) or not indicative (*non-indicative*) of what happened during a scene.
- We apply a machine learning model to this data set to show how *indicative posts* can be delineated from *non-indicative posts*

2. Related Work

Social media and online forum data has been used by prior works to understand how users communicate on these forums as it relates to loneliness (Guntuku et al., 2019b; Andy and others, 2021; Andy, 2021), depression



Jon receives a message from kings landing

Figure 2: Example of an image and description of a scene in an episode of GoT on Amazon prime. The duration of the scene (in minutes) is not shown.

(Guntuku et al., 2019a; Guntuku et al., 2017; De Choudhury et al., 2013), cancer (Yang et al., 2017; Yang et al., 2019b; Yang et al., 2019a; Andy et al., 2021a), substance use (Andy and Guntuku, 2020), and COVID-19 (Stokes et al., 2020; Andy et al., 2021b). Also, social media data has been used to predict patients' risk for certain health conditions such as cardiovascular disease (Andy et al., 2021c).

As mentioned in section 1, while some televised events are airing, some people interested in these events publish social media posts about these events in real-time. Some of these posts describe part of / all of what happened during the event. Using social media posts associated with events, prior works have summarized events (Andy et al., 2019; Goyal et al., 2019; Li and Zhang, 2021). In (Andy et al., 2019), tweets were collected around episodes of a popular TV show and the highlights of these episodes were identified and summarized. In (Goyal et al., 2019), a method is proposed to (a) detect events and scenes of the event and (b) produce abstract summaries and storylines that capture the diverse viewpoints of the events. (Li and Zhang, 2021) examines two approaches to using tweets to summarize events i.e. using terms related to events and using a graph convolutional network. Some tweets related to events make pronomial references to people/characters involved in the event; (Andy et al., 2020) developed an algorithm to resolve these pronomial mentions in tweets related to two events.

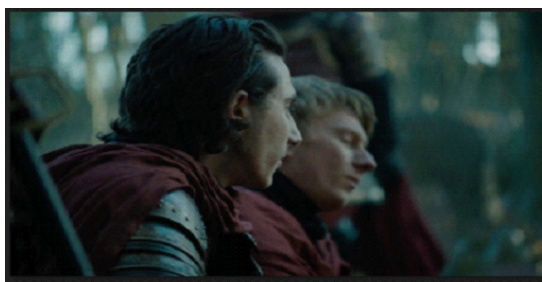
The goal of this work is different from prior works. The task in this work is as follows: given social media posts (tweets) associated with an episode of GoT and published while the show was airing, we identify the various scenes of the show and for each scene, we label the tweets that are *indicative* or *non-indicative* of what happened in each scene. We then train a machine learning model to predict *indicative posts* associated with each scene from *non-indicative posts*.

3. Data

The data used for this work is made up of tweets associated with an episode of GoT. Similar to prior works

Scene	No. of indicative posts	No. of non-indicative posts	Total No. of posts
1	11	197	208
2	67	835	902
3	68	274	342
4	86	654	740
5	63	914	977
6	112	1,359	1,471
7	45	877	922

Table 1: Information about *indicative* and *non-indicative posts* in each scene



Arya runs into a group of soldiers

Figure 3: Example of an image and description of a scene in an episode of GoT on Amazon prime. The duration of the scene (in minutes) is not shown.

(Andy et al., 2017; Andy et al., 2019; Andy et al., 2020), using the Twitter streaming API, we collected 5,562 tweets that contained "#got" while an episode of GoT was airing. The streaming service by Amazon Prime contains videos from each episode of GoT, the timestamps of the beginning and end of each scene, a brief description of each scene, and an image that is representative of each scene, as shown in Figures 2 and 3. For ground truth, we collected the timestamp of the beginning and end of each scene from the GoT episode on the Amazon prime video streaming service.

4. Scene detection and annotation

Using the method from (Andy et al., 2019) which predicted scenes of a TV show from social media posts related to the show by using the spikes in the tweet stream, we predict scenes of the GoT episode; for each predicted scene, we mark the start time and the end time. Similar to prior work (Bekoulis et al., 2019), a predicted scene is selected as correct if its start time is within the start time and end time of the scenes ground truth (from Amazon prime video streaming service). Our scene detection algorithm correctly predicted 7 out of 10 scenes in the GoT episode.

For each scene correctly detected by our method, we collected all the tweets that were published during the timeline of the scene. The New York Times (NYT) published a summary of each episode of GoT a day after

each episode aired. We showed the NYT summary of the GoT episode, the timestamp of each scene (from the Amazon stream service), and the tweets published during the timeline of the scene to 3 annotators who had watched the GoT episode. We asked the annotators to label the tweets associated with each scene as either an *indicative post* or *non-indicative post* (indicative and non-indicative posts are defined in section 1). Cohen’s kappa score was used to measure the agreement between the annotators and a Cohen’s kappa score of 0.75 was obtained. We selected an *indicative post* or *non-indicative post* if at least 2 of the 3 annotators agreed on the annotation.

Table 1 shows the number of *indicative posts* and *non-indicative posts* in each scene.

In this work, we conduct two sets of analyses. The first analysis aims to determine if the tweets (or some of the tweets) associated with each scene reflect what happened in these scenes. In the second task, we apply a machine learning model to predict *indicative posts* from *non-indicative posts* in each identified scene.

5. Compare posts in scenes

In this section, we carry out an analysis to determine if the tweets in the identified scenes capture what happened in the scene. To do this, we use latent dirichlet allocation (LDA) (Blei et al., 2003). LDA is a generative model which assumes that topics consist of a combination of words and documents (i.e. tweets in this work) consist of a mixture of topics. LDA has been used by several prior works to gain insights from social media and online forum data (Guntuku et al., 2019b; Andy and others, 2021; Andy et al., 2021a; Andy et al., 2021c).

In our analysis in this section, we first used Happier-FunTokenizer¹ (which is a tokenization tool that can identify variations in word expressions and emoticons) to tokenize words in tweets. Then, LDA was used to extract LDA topics; specifically, using LDA, we compare the tweets published in each identified scene to the previous scene to identify the topics most associated with each scene. The intuition here is to determine if the topics associated with each scene (when compared to the previous scene) are representative of what happened during the scene. We varied the number of LDA

¹<https://github.com/dlatk/happierfuntokenizing/>

topics and obtained the most coherent topic themes with 10 LDA topics; hence, we used 10 LDA topics for the analysis in this section. We did not use the tweets published in scenes 1 and 3 (Table 1) because few tweets (< 500) were associated with these scenes. Tables 2, 3, 4, 5, 6 show the LDA topics most associated with each scene i.e. scenes 2,4,5,6, and 7 respectively. The effect sizes of the LDA topics were reported using Cohen’s D (which signifies the standardized difference between two means); only topics with a Cohen’s D equal to or above 0.10 were reported. One of the authors reviewed the topic themes associated with each scene and determined that they corresponded to what happened in the corresponding scene.

LDA Topics	Cohen’s D
arya, home, sansa, stark, list	0.226
bran, littlefinger, kill, dagger, raven	0.163
arya, winterfell, back, brianne, guards	0.119

Table 2: Results from LDA analysis for scene 2 compared to scene 4

LDA Topics	Cohen’s D
jon, dany, cave, tyrion, talking	0.312
knee, bend, north, king, dany	0.105

Table 3: Results from LDA analysis for scene 4 compared to scene 2

LDA Topics	Cohen’s D
arya, brianne, training, fighting, sparring	0.114

Table 4: Results from LDA analysis for scene 5 compared to scene 4

LDA Topics	Cohen’s D
dracarys, fire, queen, blood, mad	0.158
drogon, dany, lannisters, burning, westeros	0.103

Table 5: Results from LDA analysis for scene 6 compared to scene 5

LDA Topics	Cohen’s D
wow, tyrion, idiot, intense, end	0.103

Table 6: Results from LDA analysis for scene 7 compared to scene 6

6. Compare Indicative posts to Non-indicative posts

In this section, given each identified scene and the *indicative* and *non-indicative* posts associated with these scenes, we aim to train a machine learning model to predict *indicative posts* from *non-indicative posts*. Similar to prior works (Guntuku et al., 2019b; Andy et al., 2021c), we use an open vocabulary approach based on LDA to predict *indicative posts* from *non-indicative posts* for each scene. Specifically, similar to section 5, we extracted 10 LDA topics from tweets published in each scene. Using these LDA topics as features, we trained a logistic regression model (using 5-fold cross validation) to predict *indicative posts* from

non-indicative posts. We use area under the receiver operative curves (AUC) to measure the performance of the model. We did not apply our model to scenes 1 and 3 because these scenes had fewer tweets associated with them (i.e. < 500 tweets).

We observed that for each scene, our model was able to predict *indicative posts* from *non-indicative posts*, as shown in Table 7.

Scenes	AUC
2	0.852
4	0.854
5	0.925
6	0.753
7	0.848

Table 7: Results from predicting *indicative posts* from *non-indicative posts*

7. Discussion, Future Work, and Limitations,

In this work, using tweets associated with a TV show, we identified the scenes from the show and then applied a machine learning model to predict *indicative posts* from *non-indicative posts*. The results obtained from this work indicate that it is possible to delineate social media posts (associated with a TV show) that describe what happened during scenes of the show from other posts that are not descriptive of what happened. This work shows that despite there being more *non-indicative posts* associated with a scene compared to *indicative posts*, it is possible to train a machine learning model to delineate these posts.

The model applied to this data set can be used as a baseline for predicting *indicative posts* from *non-indicative posts*. This work focused on a TV show, however, potentially, the data collection and model described in this work can be extended to other events such as Presidential debates, sports events, political events, or events related to natural disasters.

In this work, we used the scene detection algorithm from (Andy et al., 2019), which identifies highlights of events. In the future, we aim to use a scene detection algorithm that identifies all the scenes (i.e. whether they are highlights or not).

This work has some limitations. Here we describe two:

- The data set used in this work is associated with a popular TV show and so the results obtained may vary with TV shows or events that have less social media posts associated with them
- During an on-going event or TV show, some people interested in specific scenes of the show may publish tweets associated with these scenes after the timeline of the scenes. In this work, we only focus on tweets published during the timeline of each scene. In the future, we aim to identify all the posts associated with a scene i.e. whether it was published during the timeline of the scene or afterwards.

8. Conclusion

In this work, using tweets associated with a TV show, we identify scenes of the TV show and show that the LDA topics associated with these scenes refer to what happened during the scenes. Also, we build a machine learning model to predict *indicative posts* from *non-indicative posts* in these scenes with high AUC's.

9. Bibliographical References

- Andy, A. and Guntuku, S. (2020). Does social support expressed in post titles elicit comments in online substance use recovery forums? *arXiv preprint arXiv:2011.05103*.
- Andy, A. et al. (2021). Studying how individuals who express the feeling of loneliness in an online loneliness forum communicate in a nonloneliness forum: Observational study. *JMIR Formative Research*, 5(7):e28738.
- Andy, A., Dredze, M., Rwebangira, M., and Callison-Burch, C. (2017). Constructing an alias list for named entities during an event. In *Proceedings of the 3rd Workshop on Noisy User-generated Text*, pages 40–44.
- Andy, A., Wijaya, D. T., and Callison-Burch, C. (2019). Winter is here: Summarizing twitter streams related to pre-scheduled events. In *Proceedings of the Second Workshop on Storytelling*, pages 112–116.
- Andy, A., Callison-Burch, C., and Wijaya, D. T. (2020). Resolving pronouns in twitter streams: Context can help! In *Proceedings of the Third Workshop on Computational Models of Reference, Anaphora and Coreference*, pages 133–138.
- Andy, A., Andy, U., et al. (2021a). Understanding communication in an online cancer forum: Content analysis study. *JMIR cancer*, 7(3):e29555.
- Andy, A., Chu, B., Fathy, R., Bennett, B., Stokes, D., and Guntuku, S. C. (2021b). Understanding social support expressed in a covid-19 online forum. In *Proceedings of the 12th International Workshop on Health Text Mining and Information Analysis*, pages 19–27.
- Andy, A. U., Guntuku, S. C., Adusumalli, S., Asch, D. A., Groeneveld, P. W., Ungar, L. H., and Merchant, R. M. (2021c). Predicting cardiovascular risk using social media data: Performance evaluation of machine-learning models. *JMIR cardio*, 5(1):e24473.
- Andy, A. (2021). Understanding user communication around loneliness on online forums. *PloS one*, 16(9):e0257791.
- Bekoulis, G., Deleu, J., Demeester, T., and Develder, C. (2019). Sub-event detection from twitter streams as a sequence labeling problem. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 745–750.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022.
- De Choudhury, M., Gamon, M., Counts, S., and Horvitz, E. (2013). Predicting depression via social media. In *Seventh international AAAI conference on weblogs and social media*.
- Goyal, P., Kaushik, P., Gupta, P., Vashisth, D., Agarwal, S., and Goyal, N. (2019). Multilevel event detection, storyline generation, and summarization for tweet streams. *IEEE Transactions on Computational Social Systems*, 7(1):8–23.
- Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H., and Eichstaedt, J. C. (2017). Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences*, 18:43–49.
- Guntuku, S. C., Preotiuc-Pietro, D., Eichstaedt, J. C., and Ungar, L. H. (2019a). What twitter profile and posted images reveal about depression and anxiety. In *Proceedings of the international AAAI conference on web and social media*, volume 13, pages 236–246.
- Guntuku, S. C., Schneider, R., Pelullo, A., Young, J., Wong, V., Ungar, L., Polsky, D., Volpp, K. G., and Merchant, R. (2019b). Studying expressions of loneliness in individuals using twitter: an observational study. *BMJ open*, 9(11):e030355.
- Li, Q. and Zhang, Q. (2021). Twitter event summarization by exploiting semantic terms and graph network. In *Proceedings of the The Thirty-Third Annual Conference on Innovative Applications of Artificial Intelligence (IAAI-21)*.
- Panda, R., Kuanar, S. K., and Chowdhury, A. S. (2018). Nyström approximated temporally constrained multisimilarity spectral clustering approach for movie scene detection. *IEEE transactions on cybernetics*, 48(3):836–847.
- Shen, C., Liu, F., Weng, F., and Li, T. (2013). A participant-based approach for event summarization using twitter streams. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1152–1162.
- Stokes, D. C., Andy, A., Guntuku, S. C., Ungar, L. H., and Merchant, R. M. (2020). Public priorities and concerns regarding covid-19 in an online discussion forum: longitudinal topic modeling. *Journal of general internal medicine*, 35(7):2244–2247.
- Yang, D., Yao, Z., and Kraut, R. (2017). Self-disclosure and channel difference in online health support groups. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 11.
- Yang, D., Kraut, R. E., Smith, T., Mayfield, E., and Jurafsky, D. (2019a). Seekers, providers, welcomers, and storytellers: Modeling social roles in online health communities. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–14.
- Yang, D., Yao, Z., Seering, J., and Kraut, R. (2019b).

The channel matters: Self-disclosure, reciprocity and social support in online cancer support groups. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–15.