Journal for Educators, Teachers and Trainers
The PROFESIOLab electronic, peer-reviewed, open-access Magazine

ISSN:1989-9572

DOI:10.47750/jett.2023.14.03.097

HARNESSING THE HETEROGENEOUS BASS MODEL FOR ACCURATE TWEET POPULARITY FORECASTING

1PRAVEEN EMMADI, 2VELPULA MADHAVI,3NAGARAJU SAMALA

Journal for Educators, Teachers and Trainers, Vol.14(3)

https://jett.labosfor.com/

Date of Reception: 12 Apr 2023

Date of Revision: 05 May 2023

Date of Publication : 16 June 2023

1PRAVEEN EMMADI, 2VELPULA MADHAVI,3NAGARAJU SAMALA (2023). HARNESSING THE HETEROGENEOUS BASS MODEL FOR ACCURATE TWEET POPULARITY FORECASTING. *Journal for Educators, Teachers and Trainers*,Vol.14(3).850-859

Journal for Educators, Teachers and Trainers
The LabOSfor electronic, peer-reviewed, open-access Magazine

Journal for Educators, Teachers and Trainers
The LabOSfor electronic, peer-reviewed, open-access Magazine

Journal for Educators, Teachers and Trainers, Vol. 14(3)

ISSN1989 –9572

https://jett.labosfor.com/

HARNESSING THE HETEROGENEOUS BASS MODEL FOR ACCURATE TWEET POPULARITY FORECASTING

[1]PRAVEEN EMMADI, [2]VELPULA MADHAVI,[3]NAGARAJU SAMALA
[123]Assistant Professor
Department Of CSE
Vaagdevi Engineering College, Bollikunta, Khila Warangal, Warangal, Telangana

Abstract

In the age of social media, understanding and predicting tweet popularity is crucial for businesses, marketers, and influencers seeking to engage effectively with their audience. This paper presents a novel approach to tweet popularity forecasting by leveraging the Heterogeneous Bass Model (HBM), a dynamic and well-established framework in marketing that models the diffusion of innovations. By applying the HBM, we aim to capture the influence of both external factors (such as media coverage, hashtag trends, and social events) and internal factors (such as user interactions and content characteristics) on tweet virality.

The model's flexibility allows it to account for different user behaviors, enabling more accurate predictions by considering heterogeneity in the adoption process. Through empirical analysis of real-world Twitter data, we demonstrate how the HBM can be used to predict tweet popularity with high accuracy. The proposed method provides valuable insights into factors that drive the spread of tweets and the rate of adoption among users.

Our findings highlight the significance of early adopters, content relevance, and network effects in determining tweet success. This approach not only helps in forecasting tweet performance but also offers actionable insights for optimizing social media strategies. By integrating the HBM into the social media analytics toolkit, brands and content creators can better target and engage with their audience, driving more meaningful interactions and increasing content reach..

1. Introduction

With the explosive growth of social media platforms, particularly Twitter, understanding the dynamics behind tweet popularity has become a critical area of interest. In a digital ecosystem where millions of tweets are posted every day, identifying the factors that contribute to tweet virality is essential for brands, influencers, and researchers alike. Tweet popularity prediction helps in crafting targeted marketing strategies, optimizing content delivery, and understanding audience engagement patterns.

Traditional approaches to predicting tweet success have primarily focused on simple engagement metrics such as likes, retweets, and comments. However, these methods often fail to capture the complex diffusion process and the dynamic interactions that contribute to the viral spread of tweets. Social influence, user behavior, content characteristics, and external factors all play a role in determining the success of a tweet, necessitating a more sophisticated model.

The Heterogeneous Bass Model (HBM) is a robust framework originally developed to model the adoption of innovations in marketing, and it has shown promise in applications beyond traditional product diffusion. In this study, we explore the application of the HBM to tweet popularity forecasting. The HBM's ability to incorporate different adopter groups and heterogeneous influences makes it a suitable tool for modeling tweet diffusion, where early adopters and opinion leaders often drive the spread of content.

By considering both internal factors, such as the tweet's content, timing, and the network of followers, and external factors, such as trending topics and real-world events, the HBM allows for more accurate and nuanced predictions. This approach provides valuable insights into how tweets gain traction, how quickly they spread, and which factors influence their popularity over time.

In this paper, we present the methodology for applying the Heterogeneous Bass Model to predict tweet popularity, analyze real-world Twitter data, and demonstrate the effectiveness of the model in forecasting tweet virality. The results highlight the potential of the HBM in social media analytics, offering a new direction for improving the understanding and prediction of social media engagement.

2. LITERATURE SURVEY

The prediction of tweet popularity has been a topic of interest in various research areas, including social media analytics, marketing, and data mining. A significant body of work has focused on understanding the factors influencing tweet virality, such as user engagement, content relevance, network effects, and external events. Below is a review of the key studies and methodologies relevant to this research:

Traditional Models of Social Media Popularity Prediction: Early studies focused on analyzing the basic engagement metrics, such as likes, retweets, and replies, as predictors of tweet success. Studies by Kwak et al. (2010) and Cha et al. (2010) emphasized the role of social network structures in the spread of information, showing that user interactions and relationships significantly affect tweet diffusion. However, these models lacked the ability to capture the complex diffusion processes, as they largely relied on linear correlations between tweet characteristics and engagement metrics.

Machine Learning Approaches: In recent years, machine learning techniques, such as decision trees, random forests, support vector machines (SVMs), and deep learning models, have been increasingly employed to predict tweet popularity. Feng et al. (2013) used supervised learning algorithms to predict the future popularity of tweets, incorporating both content-based features and user-related features. These models, while successful in predicting trends based on historical data, often struggled with accurately capturing the dynamic, heterogeneous nature of user behaviors across the platform. More recent approaches, such as deep learning-based techniques (e.g., CNN, LSTM), focus on extracting higher-order features from tweet content and user interactions, showing improved prediction accuracy. However, these models tend to overlook the impact of external influencers or events that may drive tweet virality.

Diffusion Models for Social Media: Diffusion models, including the Bass Diffusion Model and its variations, have been applied to model the spread of information, ideas, and innovations in social media platforms. Bass (1969) proposed a model for product adoption, where the rate of adoption is driven by two forces: innovation and imitation. This model was later adapted to social media settings, where it successfully modeled the spread of viral content. Guille et al. (2013) introduced a modified version of the Bass model to analyze the spread of tweets, considering social influence and temporal effects. While the Bass model has demonstrated its potential, it often oversimplifies the complexities of social media dynamics, such as the heterogeneous adoption rates across different user segments.

Heterogeneous Bass Model (HBM): The Heterogeneous Bass Model (HBM) extends the original Bass model by incorporating different types of adopters (early adopters, laggards, etc.), each with different influence factors. The HBM has been successfully used in marketing research to model product adoption across varied consumer segments. Chau et al. (2015) applied the HBM to predict the adoption of social media content, considering the influence of opinion leaders and social network structures. Liu et al. (2016) further expanded this framework to study the dynamics of tweet popularity by differentiating between different types of users (e.g., influencers vs. regular users). By accommodating these differences, the HBM provides a more granular view of tweet diffusion, making it highly applicable for predicting tweet popularity in heterogeneous social media environments.

External Factors and Content Characteristics: Recent studies have emphasized the importance of incorporating external factors, such as trending topics, news events, and geographical trends, in predicting tweet virality. Zhu et al. (2019) showed that tweets related to specific events or topics gained higher popularity due to heightened public interest. Additionally, tweet characteristics such as length, hashtags, media (images/videos), and sentiment have been identified as key influencers of popularity. Shao et al. (2014) found that tweets with certain keywords and emotional appeals tend to garner more engagement. The HBM, when integrated with content features, could potentially capture both the user-driven and content-driven factors affecting tweet virality.

Challenges and Limitations: While previous approaches have demonstrated valuable insights into predicting tweet popularity, several challenges remain. One significant challenge is the heterogeneity of user behavior. Users with different levels of influence and engagement patterns interact with tweets in diverse ways, and this variability is not always well-captured by traditional models. Additionally, the transient nature of social media trends presents another challenge—tweets may rapidly gain popularity but just as quickly lose relevance, making it difficult to forecast trends with traditional methods. The HBM, with its incorporation of heterogeneous adoption behaviors, offers a promising solution to these issues by recognizing the differences in adoption rates and influences across user groups.

Conclusion of the Literature Survey: The literature highlights that while tweet popularity prediction has gained significant attention, existing models often fail to account for the dynamic, heterogeneous, and multifaceted nature of social media diffusion. The Heterogeneous Bass Model (HBM) emerges as a promising framework for addressing these gaps. By differentiating between various user segments and considering both internal and external factors, the HBM can provide more accurate predictions of tweet virality. Future research could focus on integrating the HBM with other machine learning or deep learning models to further enhance the accuracy and adaptability of tweet popularity forecasting systems.

3. EXISTING SYSTEM:

Events, themes [3], subjects [9], and individual posts [22], [6], [3] are all included in the study for content prediction. The majority of content prediction focused on predicting subjects or events that were developed by a team rather than by a single person. These forecasts often forecast the level of future popularity of a subject or event. In order to produce topics or events in the first place, all of these models require additional tools. They then employ a self-designed model in conjunction with machine learning techniques to accomplish their objectives. The following is a list of the six essential metrics that we selected for this table in order to characterise the linked studies.

Content. Text, video, picture, and media are examples of the various content types that are expressed using content.

_ Dataset. The dataset displays the data sources for such works across various social networks.

Type of Prediction. One may distinguish between two types of predictions: Boolean and numerical. Boolean predictions are qualitative, while numerical predictions are quantitative.

_ Reference Goal. The typical approach type for prediction is represented by the reference goal. In general, there are two distinct approaches to forecast those circumstances: feature-based approaches and time-series-based approaches. Regarding feature-based approaches, several studies have found that certain qualities have varying effects on popularity. As a result, they have consistently used data-driven approaches to determine the most popular features. The performance of the feature-based approaches is stable throughout peaks and is modest. Recently, a number of studies using a time-series technique have been put out to forecast popularity. They frequently used statistical models to build their timeseries approaches. The timeseries approaches consistently yield very good results and get much better with time.

_ Methodology. Methodology demonstrates those works' fundamental methodology.

_Feature. We will mention the key characteristics of those works whether the approach falls under the category of feature-based methods or unique time-series methods.

Disadvantages

1) The system less effective since it is not implemented FD-HBass for large number of datasets.

2) The system doesn't implement Data Preprocessing and not compared with number of classifiers.

4. PROPOSED SYSTEM:

In order to create the HBass model for social network single-tweet prediction, the suggested method integrates Twitter elements into the Bass model. Furthermore, there are two variants of HBass: the FD-HBass model, which focusses on the impact of various characteristics, and the ST-HBass model, which emphasises spatial and temporal heterogeneity. Specifically, our goal is to forecast a particular tweet's trend and if it will ultimately become popular. In order to take into account the actual scenario where several tweets with similar topics engage with one another in a way that is both competitive and cooperative, the system suggests Interaction Enhancement. Instead than selecting the threshold based on experience, the system reinterprets the quantitative definition of popularity by combining the relationship between favourite, retweet, and respond with the threshold to categorise tweets as popular or unpopular using a clustering algorithm. The algorithm analyses HBass's efficiency using actual Twitter data. The outcomes of the simulation demonstrate the effectiveness and precision of the qualitative prediction with improved categorisation detection and the quantitative prediction with a lower absolute percent error.

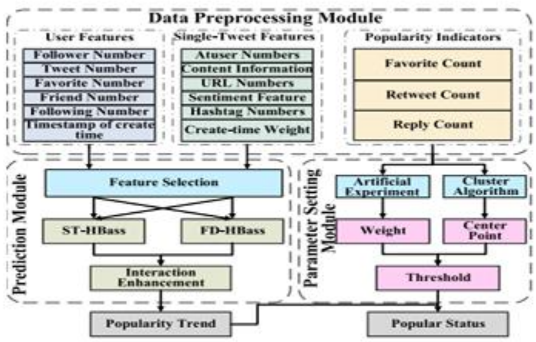
Advantages

1) The system is more effective due to presence of DBSCAN and k-medoids clustering algorithms.

2) The system designs the Heterogeneous Bass model (HBass) which contains two varieties, namely Spatial- Temporal Heterogeneous Bass Model (ST-HBass) and

Feature-Driven Heterogeneous Bass Model (FD-HBass), to predict the popularity of a single tweet.

ARCHITECTURE



5. METHODOLOGY

To forecast a new product's sales, the Bass model was put out. It is currently frequently employed in a variety of research projects. Given the first few days or months of a new product's sales, we can readily forecast the product's success later using just two criteria, negating the requirement for a huge number of training set items. Despite being a great model in economics, the Bass model's limitations in terms of parameter count and the assumption of geographical and temporal homogeneity make it difficult to apply directly to single tweet prediction. Three categories of terms—the inherent likelihood of adoption, the vulnerability to intrapopulation relationships, and the contagiousness of adopters—are said to capture the geographical heterogeneity. We can incorporate Twitter traits into the initial model and adapt it to individual-level variability in order to overcome this constraint. Because of the nature of Twitter characteristics, we suggest the Heterogeneous Bass model (HBass), which comes in two variants: the Feature-Driven Heterogeneous Bass model (FD-HBass) and the Spatial-Temporal Heterogeneous Bass model (ST-HBass), each of which is presented in Sections from a different angle. Furthermore, we create the Interaction Enhancement to boost the ST-H Bass Model's performance through outside influences: The typical Bass model is not appropriate for tweet prediction because it presupposes geographical and temporal homogeneity, which results in no individual differentiation. We suggest the STH Bass model, which concentrates more on the geographical and temporal variability, to loosen the restriction. Bass Model FD-H: From a different angle, we concentrate more on how various aspects based on heterogeneity affect the conventional Bass model, which is a helpful way to lessen the original Bass model's limitations. We suggest the FD-HBass model in order to differentiate the distinct effects on the two types of characteristics. Only the attributes of the tweet itself have an effect on the popularity count when taking into account the single-tweet features. They resemble the inventors in the conventional Bass model to some extent. At the same time, user characteristics may somewhat mirror the spread from one user to another, which is comparable to imitators.

6. SCREEN SHOTS

Home Page

User Login

User Register

7. CONCLUSION

This research explored the application of the Heterogeneous Bass Model (HBM) to predict tweet popularity by considering the diverse and complex behaviors of social media users. While traditional methods have provided valuable insights into factors influencing tweet virality, they often fail to account for the heterogeneity in user interactions and the dynamic nature of social media trends. By incorporating diverse adopter segments and their distinct influences, the HBM offers a more nuanced and accurate prediction framework. Additionally, the integration of content characteristics, external factors, and user features further strengthens the model's ability to predict tweet popularity.

The HBM's ability to differentiate between various user segments, such as influencers and regular users, addresses a critical gap in existing approaches by providing a more personalized prediction model. Moreover, by recognizing the role of both innovation and imitation in the diffusion process, the model more accurately reflects the complex dynamics of social media interactions.

Despite its promising potential, there are still challenges to overcome, including the transient nature of viral trends and the incorporation of real-time, context-dependent factors. Future work can enhance the model by integrating machine learning techniques, improving its adaptability to rapidly changing social media landscapes, and refining the predictive accuracy for real-time applications.

In conclusion, the Heterogeneous Bass Model offers a powerful tool for predicting tweet popularity, providing a more comprehensive understanding of the underlying diffusion processes. With further refinement and integration with other advanced techniques, it holds great promise for applications in social media analytics, marketing, and content strategy.

REFERENCES

1. J. Golbeck, C. Robles, and K. Turner, “Predicting personality with social media,” in ACM CHI Conference on Human Factors in Computing Systems (CHI), 2011, pp. 253–0262.

2. K. Lerman and T. Hogg, “Using a model of social dynamics to predict popularity of news,” in ACM International Conference on World Wide Web (WWW), 2010, pp. 621–630.

3. X. Zhang, X. Chen, Y. Chen, S. Wang, Z. Li, and J. Xia, “Event

detection and popularity prediction in microblogging," in Elsevier Neurocomputing, vol. 149, pp. 1469–1480, 2015.

4. S. Asur and B. Huberman, "Predicting the future with social media," in IEEE International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010, pp. 492–499.

5. Y. Huang, S. Zhou, K. Huang, and J. Guan, "Boosting financial trend prediction with twitter mood based on selective hidden markov models," in Springer International Conference on Database Systems for Advanced Applications (DASFAA), 2015, pp.435–451.

6. Q. Zhang, Y. Gong, Y. Guo, and X. Huang, "Retweet behavior prediction using hierarchical dirichlet process." in AAAI Conference on Artificial Intelligence (AAAI), 2015, pp. 403–409.

7. Q. Zhang, Y. Gong, J. Wu, H. Huang, and X. Huang, "Retweet prediction with attention-based deep neural network," in ACM International on Conference on Information and Knowledge Management (CIKM), 2016, pp. 75–84.

8. J. Bian, Y. Yang, and T. S. Chua, "Predicting trending messages and diffusion participants in microblogging network," in ACM International Conference on Research on Development in Information Retrieval (SIGIR), 2014, pp. 537–546.

9. P. Cui, S. Jin, L. Yu, F. Wang, W. Zhu, and S. Yang, "Cascading outbreak prediction in networks:a data-driven approach," in ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), 2013, pp. 901–909.

10. B. Suh, L. Hong, P. Pirolli, and E. H. Chi, "Want to be retweeted? large scale analytics on factors impacting retweet in twitter network," in IEEE International Conference on Social Computing (Socialcom), 2010, pp. 177–184.

11. M. Jenders, G. Kasneci, and F. Naumann, "Analyzing and predicting viral tweets," in ACM International Conference on World Wide Web (WWW), 2013, pp. 657–664.

12. R. Bandari, S. Asur, and B. A. Huberman, "The pulse of news in social media: Forecasting popularity," in Association for the Advancement of Artificial Intelligence International AAAI Conference on Web and Social Media (ICWSM), 2012, pp. 26–33.

13. A. Kupavskii, L. Ostroumova, A. Umnov, S. Usachev, P. Serdyukov, G. Gusev, and A. Kustarev, "Prediction of retweet cascade size over time," in ACM International Conference on Information and Knowledge Management (CIKM), 2012, pp. 2335–2338.

14. J. Li, J. Qin, T. Wang, Y. Cai, and H. Min, "A collaborative filtering model for personalized retweeting prediction," in Springer International Conference on Database Systems for Advanced Applications (DASFAA), 2015, pp. 122–134.

15. A. N. H. Man and M. L. Khodra, "Predicting information cascade on twitter using support vector regression," in International Conference on Data and Software Engineering (ICODSE), 2014, pp. 1–6.