Analyses of Character Emotions in Dramatic Works by Using EmoLex Unigrams

Mehmet Can Yavuz

Faculty of Engineering and Natural Science, Sabancı University, Tuzla
İstanbul, Türkiye
mehmetyavuz@sabanciuniv.edu

Abstract

In theatrical pieces, written language is the primary medium for establishing antagonisms. As one of the most important figures of renaissance, Shakespeare wrote characters which express themselves clearly. Thus, the emotional landscape of the plays can be revealed from the texts. It is important to analyze such landscapes for further demonstrating these structures.

We use word-emotion association lexicon with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive). By using this lexicon, the emotional state of each character is represented in 10 dimensional space and mapped onto a plane. This principle axes planes position each character relatively. Additionally, tempora-emotional evaluation of each play is graphed.

We conclude that the protagonist and the antagonist have different emotional states from the rest and these two emotionally oppose each other. Temporal-Emotional timeline of the plays are meaningful to have a better insight into the tragedies.

1 Introduction

Shakespeare's plays are one of the most important works of early modernity with their dramaturgy, strong and in-depth characters and poems, that are all still contemporary. Antagonistically most powerful works of Shakespeare, who wrote in three genres, are tragedies. Tragedies have strong antagonisms and written language is the primary medium for establishing antagonisms. Therefore, the characters in these plays express themselves through dialogues or monologues clearly. By using the emotional association of each word, from this perspective, it is possible to reveal the emotional landscape of the plays. Emotionally positioning the characters relative to each other is important for further understanding of the structure of the plays. It is also important to extract overall emotional variations throughout the play to have an insight about the tragedies. In this study, it is aimed to answer these two questions by using word-emotion association lexicon with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive).

These tasks are not only important for artificial literature (Lebrun, 2017; Yavuz, 2020), but also for the purpose of increasing artistic creativity through computerized analysis. The algorithmic study of literary works has given rise to objective criticism in literary theory, (Moretti, 2013). Accordingly, the discovery of new or previously theoretically laid out features of literary texts through mathematical experiments or statistics brought along conjunctions on literature, (Moretti, 2000; Yavuz, 2020). Art and criticism develop in parallel today as of yesterday. Computational linguistics enriched with fields such as advanced chatbots, conversational AI, or text style transfers, the fields give clues that artificial literature will also develop rapidly in the upcoming period. Whether it is a computer-assisted writing process or fully automated *writing machines*, these generative models need evaluation metrics. For this purpose, objective metrics should be developed for outstanding literary pieces such as Shakespeare, and for what makes these plays so great, if possible... Aim of this paper is in accordance with the previous works, (Yavuz, 2019; Yavuz, 2020), we would like to discuss objective

and quantifiable reasons behind the literariness of a drama.

In order to convey such analyses, each character in any play represented in 10 dimensional emotional space and mapped onto a plane. The emotional timeline of the plays are also revealed. We would like to overview the word-emotion association lexicon and dimension reduction algorithm. Then, the last section is left to discussions on the emotional states of the characters and temporal analyses of the emotions.

1.1 Related Works

Distant reading is a established approach/methodology in digital humanities (DH) and mostly deals with the quantitative analyses of literary and cultural studies (Clement, 2008;Crane, 2006). Inside DH, "Drametrics" is a sub-research field, specialized on quantitative analysis of the literary genre of drama (Romanska, 2015). Digital Shakespeare research and projects have gotten attention since the 2000s (Hirsch, 2014; Mueller, 2008). The dramatic structures in the form of antagonisms are revealed by topic modeling algorithms, (Yavuz, 2019). Exceptional characters, such as Deus-ex-Machina, is also detected in semantic space, (Yavuz, 2020). The graphs are also used to extract secondary antagonisms, (Yavuz,2020). Machine learning based text analyses are also carried out for genre classifications (Yavuz, 2019; Ardanuy, 2014; Hope, 2010; Schöch, 2016; Underwood, 2013; Yu, 2008). In literature, structural elements such as *dramatis persone* are also analyzed and applications are developed for further analyses (Dennerlein, 2015; Krautter, 2018; Schmidt, 2019, Trilcke, 2015; Wilhelm, 2013). In addition to dramatic structure works, there is literature that apply sentiment analyses for dramatic works, (Nalisnick, 2013; Schmidt, 2018; Schmidt, 2018a; Schmidt, 2018b).

2 Methodology

Our approach is lexicon based. Lines uttered by each character treated as a document and represented with a tf-idf. Emotional weights are multiplied with the vectors and the summed up to have the final 10 dimensional emotion space and then reduced to a plane by linear dimension reduction. Overall emotional state of any character, thus, is represented in relative point to each other. Lexicon is also used to extract temporal emotional varia-

tions along the texts.

2.1 EmoLex Unigrams

Psychologists proposed many theories for classifying human emotions, (Dalgleish, 2000). Some emotions are considered basic, while others are considered as complex. The distinction can be between emotions that we can sense and perceive (instinctual), and emotions that can arrive after some thinking and reasoning (cognitive), (Zajonc,1984). There are also oppositions, (Lazarus, 1984). According to Plutchik, (Plutchick, 1985), the discussion may not be resolvable because there is no empirical basis. There is a high correlation between basic and instinctive emotions, similarly between complex and cognitive emotions. Many of the basic emotions are also instinctual.

There are theoretical studies on what the basic emotions are. Ekman argues about about the existence of 6 basic emotions: joy, sadness, anger, fear, disgust, and surprise, (Ekman, 1992). Plutchik includes two new emotions: trust and anticipation, (Plutchick, 1994). Plutchik displayed the emotions on a wheel. The distance from the center in the circle indicates intensity. Plutchik also state the basic emotions as opposing pairs: joy–sadness, anger–fear, trust–disgust, and anticipation–surprise. The opposing places and neighborhoods on the circle are formed accordingly.

Emolex is the dictionary of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and
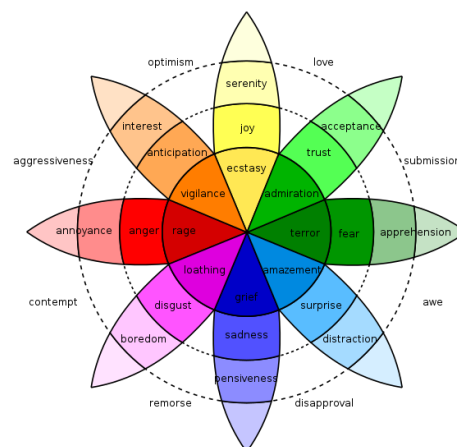

Figure 1: Eight basic emotions (*anger, fear, anticipation, trust, surprise, sadness, joy, and disgust*) on Plutchik's Wheel of Emotions

positive), (Mohammad, 2013). Thus, representing each word in 10 dimensional emotion space is effective as well as theoretically relevant. The dataset is unigram. Each single word represented in 10 dimensional feature space: 8 basic emotions and 2 sentiment labeling. Each emotion can be either 0 or 1. No intensity information is given. The words are crowd-sourced and manually labeled with Mechanical Turk. There are 14182 unigrams in the dictionary.

2.2 SVD (Golub, 1970)

Linear dimension reduction methods are for getting a n-dimensional plane over the hyperspace. For instance, if you have data cloud in 10 dimension, by mapping onto a plane, one can visualize such points. SVD is one of the relevant methods, it breaks any A matrix into three,

$$A = USV' \; which \qquad (1)$$
$$UU' = I \; and \; VV' = I \qquad (2)$$

S is a diagonal matrix that consists of r singular values. r is the rank of A. Truncated SVD is a reduced rank approximation. Only the most relevant dimensions are selected, these are the largest singular values. The dimensions of truncated SVD are $[uxk] * [kxk] * [kxv]$ Therefore A matrix is approximated by k dimensions, this is the dimension reduction. A descriptive subset of the data is called T, which is a dense summary of the matrix A,

$$T = US_k \qquad (3)$$

S_k denotes k largest singular values, which is the number of reduced features. Each feature is represented with a percentage of variance. Higher variance means more information gain.

3 Discussion

There are two analyzes we would like to discuss. The first is about whether we can get ideas about the plays based on the emotional state of the characters. For this purpose, each character is treated as a text and represented with Tf-Idf features. The weights of each word for a specific character multiplied with its 10-dimensional emotion vector. Right afterwards, SVD lets you extract principle axes and mapped onto the plane. This is the first analysis to be discussed. Secondly, the temporal dimensions of the plays are considered.

The opposing emotional pairs (given by Plutchik) were represented as a time series as positive / negative. The states are represented as a cumulative temporal sum and the emotional landscape of the tragedies are revealed.

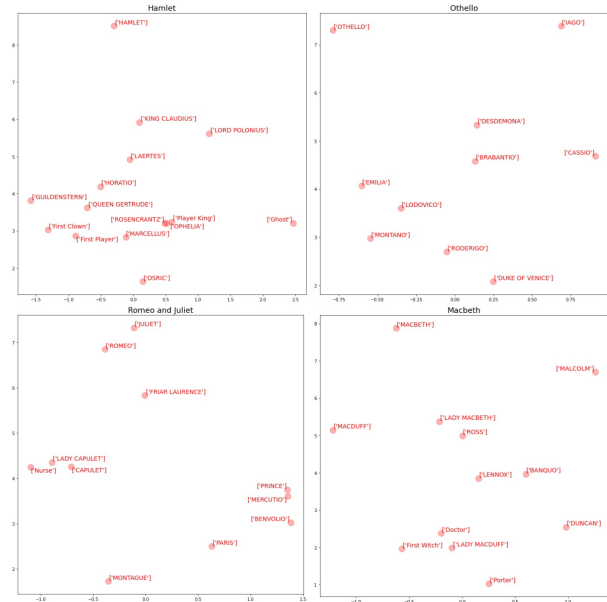## 3.1 Analyses of Character Emotions

Figure 2: Emotional position of each character is represented, Hamlet, Othello, Romeo and Juliet, Macbeth, respectively.

The weights of Tf-Idf features help to position character emotions. Each word contributes to the resulting emotion with their weights. Weights of more frequent terms affect the resulting emotional state more, while less frequent terms affect less. The 10-dimensional emotional space mapped onto an abstraction plane by linear dimension reduction. The dimensions of these planes correspond to abstract emotions or a mixture of the other 10 dimensions with certain proportions. The important thing in these graphics is the position of the characters relative to each other. The emotional positioning of Hamlet in the upper left, Othello in the upper right, Romeo and Juliet in the lower left, and Macbeth in the lower right.

In these graphics, the basic characteristics of the plays can be observed. The main characters or pairs of characters are emotionally different from the rest. The protagonist and the antagonist always have emotional contrast. For example, the Hamlet play is basically determined by the tension between the two people, Hamlet and King

Claudius. Although Hamlet is emotionally very different from other characters, King Claudius is emotionally close to the main character cluster. In all tragedies, there is a cluster of emotionally indifferent characters, we can call *the main cluster*. Characters like Lord Polonius and Laertes are also located around King Claudius with the main cluster. The Ghost character, like Hamlet, is different from all other characters and is in an opposite position to Hamlet. These observations follow the readings of the play. In Othello, Iago sets traps to harm Desdemona. Desdemona is also compatible with the main cluster. But Iago and Othello are positioned far apart and apart from the main cluster. In the Macbeth play, the two enemies, Macduff and Malcolm, are opposite and separately positioned. Lady Macbeth is emotionally compatible with the main cluster. In an interesting observation on Romeo and Juliet, the positioning of the clusters is placed in symmetry in accordance with Renaissance thought. It is known that the play is written symmetrically. There are three family positions in symmetry: Ruling house of Verona, House of Capulet, House of Montague.

The graphs show that the emotional positioning of tragedies is compatible with the readings of the play. What we mean is the protagonists and antagonists are clearly observable. Distances or orientations, or rather relative positions, are significant. The main characters that experience basic tensions could be demonstrated. In the play of Romeo and Juliet, the affinities are observed and there is symmetrical positioning of the families.

3.2 Temporal-Emotional Evaluation of the Tragedies

Temporal-emotional evaluation of the tragedies are drawn. We can assume that each play is a temporal series, and we summed up the emotional state in each timestep, the cumulative emotional curves are calculated. Therefore, the emotional directions are determined. Emotions are positioned in contrast as (Negative-Positive), (Fear-Anger), (Anticipation-Surprise), (Sadness-Joy), (Disgust-Trust). For each timestep, or the word, the contribution is the expected emotion,

$$E[e] = e * p(e) \qquad (4)$$

which $p(e)$ is the occurrence probability of the emotion in the lexicon and e is the Bernoulli random variable, either 0 or 1, either has the emotion

or not. The cumulative sum of a emotional contrast pair,

$$C(\{e_1, e_2\}, T) = \sum_{t=0}^{T} \left(E[e_2, t] - E[e_1, t] \right) \qquad (5)$$

which $\{e_1, e_2\}$ are the random variables for emotional contrast pairs. The cumulative total for each pair is specified for 5 curves. Cumulative sums of (Negative-Positive), (Sadness-Joy), (Disgust-Trust) pairs for all four plays constantly increases. (Fear-Anger), (Anticipation-Surprise) are more neutral. All in all emotionally, the temporal word distributions for tragedies are similar.

4 Conclusion

The traces of the tensions between characters are observable from the emotional aspect. As we show, the emotional positions of the protagonist, the antagonist and the main cluster gives much insight about the greatness such pieces. Any great tragedy needs emotional contrast between the main characters and there is always *the main cluster*. The temporal-emotional characteristics of the plays are also important and very much similar to each other. There are constantly increasing emotions as well as neutrals. Each play grows towards positive, joyful and trusty emotional state. This might be the reason behind a followable play. Positive feelings should accumulate.

Either it is computer-assisted or fully automated *writing machine*, artificial literature needs emotional aspect. The emotional aspect of literary works should be conditional of such generative models. The common acceptance on this early
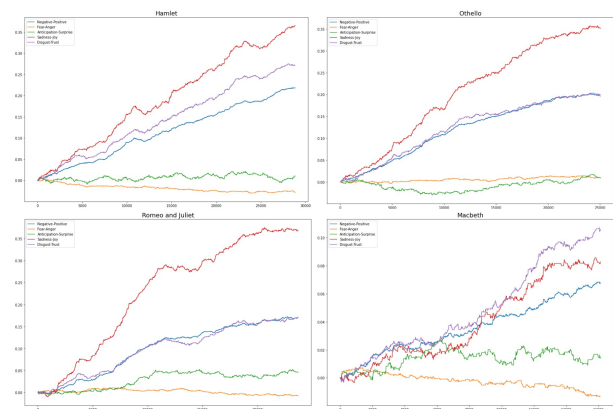

Figure 3: Emotional landscape of each character is represented, Hamlet, Othello, Romeo and Juliet, Macbeth, respectively.

modern author is his greatness as a tragedy writer. The theatrical pieces by Shakespeare are in dialog form, each character express themselves clearly. Therefore, it is shown that any dramatic antagonism is also emotional. Any artificial dramatic work should have a similar emotional resonance with such tragedies. With this analyses, we try to further develop evaluation metrics for artificial literature. A baseline metric to emotionally evaluate such theatrical forms.

On the way to Artificial Literature (ALit), there needs more criteria and more complex tools to analyze literariness of such pieces (literariness, 2020). As we shown so far, emotion aspect of the plays are very crucial at establishing antagonisms.

References

Ardanuy, M. C., & Sporleder, C. (2014, April). Structure-based clustering of novels. *In Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLFL)* (pp. 31-39).

Clement, T., Steger, S., Unsworth, J. and Uszkalo, K. (2008). *How not to read a million books*. Available online at http://people.brandeis.edu/ unsworth/hownot2read.html

Crane, G. (2006). *What do you do with a million books?* D-Lib Magazine. Available online at http://www.dlib.org/dlib/march06/crane/03crane.html

Dalgleish, T., Power, M. (Eds.). (2000). *Handbook of cognition and emotion.* John Wiley Sons.

Dennerlein, K. (2015). Measuring the average population densities of plays. A case study of Andreas Gryphius, Christian Weise and Gotthold Ephraim Lessing. *Semicerchio. Rivista di poesia comparata* LIII: 80–88.

Ekman, P. (1992). An argument for basic emotions. *Cognition emotion*, 6(3-4), 169-200.

Golub, G. H.; Reinsch, C. (1970). "Singular value decomposition and least squares solutions". *Numerische Mathematik*. 14 (5): 403–420. doi:10.1007/BF02163027. MR 1553974.

Gao, J., Galley, M., Li, L. (2018, June). Neural approaches to conversational AI. In *The 41st International ACM SIGIR Conference on Research Development in Information Retrieval* (pp. 1371-1374).

Hirsch, B., & Craig, H. (2014). "Mingled Yarn": The State of Computing in Shakespeare 2.0. In T. Bishop, & A. Huang (Eds.), *The Shakespearean International Yearbook* (Vol. 14: Special Section, Digital Shakespeares, pp. 3-35). United Kingdom: Ashgate Publishing Limited.

Hope, J., & Witmore, M. (2010). The Hundredth Psalm to the Tune of "Green Sleeves": Digital Approaches to Shakespeare's Language of Genre. *Shakespeare Quarterly*, 61(3), 357-390. Retrieved from http://www.jstor.org/stable/40985589

Krautter, B. (2018). Quantitative microanalysis? Different methods of digital drama analysis in comparison. *Book of Abstracts, DH 2018*. Mexico-City, Mexico, pp. 225-228.

Lazarus, R. S. (1984). On the primacy of cognition. American Psychologist, 39(2), 124–129. https://doi.org/10.1037/0003-066X.39.2.124

Le, Q., Mikolov, T. (2014, January). Distributed representations of sentences and documents. In *International conference on machine learning* (pp. 1188-1196).

Lebrun, T. (2017, May). Who Is the Artificial Author?. In *Canadian Conference on Artificial Intelligence* (pp. 411-415). Springer, Cham.

literariness. Oxford Reference. Retrieved 8 Jun. 2020, from https://www.oxfordreference.com/view/10.1093/oi/authority.20110803100108912.

Mohammad, S. M., Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3), 436-465.

Moretti, F. (2013). *Distant reading*. Verso Books.

Moretti, F. (2000). Conjectures on world literature. *New left review*, 54-68.

Mueller, Martin. (2008). Digital Shakespeare, or towards a literary informatics. *Shakespeare*. 4. 284-301. 10.1080/17450910802295179.

Nalisnick, E. T., Baird, H. S. (2013, August). Character-to-character sentiment analysis in Shakespeare's plays. In extitProceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (pp. 479-483).

Plutchik, R. (1985) On emotion: The chicken-and-egg problem revisited. Motiv Emot 9, 197–200 . https://doi.org/10.1007/BF00991576

Plutchik, R. (1994) The psychology and biology of emotion. HarperCollins College Publishers.

Ramsay, S. (2011). *Reading Machines: Toward an Algorithmic Criticism.* University of Illinois Press.

Romanska, M. (2015). Drametrics: what dramaturgs should learn from mathematicians. In Romanska, M. (ed.), *The Routledge Companion to Dramaturgy*. Routledge, pp. 472-481.

Schmidt, T., Burghardt, M., Dennerlein, K. & Wolff, C. (2019). Katharsis – A Tool for Computational Drametrics. *In Book of Abstracts, DH 2019*.

Schmidt, T., Burghardt, M. (2018, August). An evaluation of lexicon-based sentiment analysis techniques for the plays of gotthold ephraim lessing. In *Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature* (pp. 139-149).

Schmidt, T. and Burghardt, M. (2018b). Toward a Tool for Sentiment Analysis for German Historic Plays. In Piotrowski, M. (ed.), COMHUM 2018: *Book of Abstracts for the Workshop on Computational Methods in the Humanities 2018.* Lausanne, Switzerland: Laboratoire laussannois d'informatique et statistique textuelle. pp. 46-48.

Schmidt, T., Burghardt, M. and Dennerlein, K. (2018a). Sentiment Annotation of Historic German Plays: An Empirical Study on Annotation Behavior. Sandra Kübler, Heike Zinsmeister (eds.), extitProceedings of the Workshop on Annotation in Digital Humanities (annDH 2018) Sofia, Bulgaria. pp. 47-52.

Schmidt, T., Burghardt, M. and Dennerlein, K. (2018b). Kann man denn auch nicht lachend sehr ernsthaft sein?" – Zum Einsatz von Sentiment Analyse-Verfahren für die quantitative Untersuchung von Lessings Dramen. In *Book of Abstracts*, DHd 2018 , Cologne, Germany.

Schöch, Christof. (2016). Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama. *Digital Humanities Quarterly*. http://doi.org/10.5281/zenodo.166356

Trilcke, P., Fischer, F. and Kampkaspar, D. (2015). Digital Network Analysis of Dramatic Texts. *Book of Abstracts, DH 2015.* Sidney, Australia

Underwood, T., Black, M.L., Auvil, L., & Capitanu, B. (2013). Mapping mutable genres in structurally complex volumes. 2013 *IEEE International Conference on Big Data*, 95-103.

Wilhelm, T., Burghardt, M., and Wolff, C. (2013). "To See or Not to See" - An Interactive Tool for the Visualization and Analysis of Shakespeare Plays. In R. Franken-Wendelstorf, E. Lindinger, and J. Sieck (Eds.), *Kultur und Informatik: Visual Worlds & Interactive Spaces.* Glückstadt: Verlag Werner Hülsbusch, pp. 175–185.

Yavuz, M. C.: (2019, November) Analyses of Literary Texts by Using Statistical Inference Methods. In: *Proceedings of the Sixth Italian Conference on Computational Linguistics*, CLiC-it'19. Bari, Italy., CEUR-WS.org, online http://ceur-ws.org/Vol-2481/paper75.pdf.

Yavuz, M. C.: (2020, March) Analyses of Characters in Dramatic Works by Using Document Embeddings. In: Proceedings of the Workshop on Digital Humanities and Natural Language Processing (DHandNLP 2020) co-located with International Conference on the Computational Processing of Portuguese (PROPOR 2020), Évora, Portugal., CEUR-WS.org, online http://ceur-ws.org/Vol-2481/paper75.pdf.

Yavuz, M. C. (2020, November) "Analyses of Character Networks in Dramatic Works by Using Graphs" *The 7th International Conference on Behavioural and Social Computing)*, Bournemouth UK.

Yu, B. (2008). An evaluation of text classification methods for literary study. *Literary and Linguistic Computing* 23(3): 327-343.

Xanthos, A., Pante, I., Rochat, Y and Grandjean, M. (2016). Visualising the dynamics of character networks. *Book of Abstracts, DH 2016.* Kraków, Poland, pp. 417-419.

Zajonc, R. B. (1984). On the primacy of affect. American Psychologist, 39(2), (1984). 117–123.