Supplementary Material

Table S1

Previous studies on online movie reviews

Focus	Source	Object/Research problem	Methods	Database	Study Details
Predicting box office success	Baek et al. (2017)	Predicting box office success over time; result: online reviews more successful in late stage of a movie's release, as compared to social media	Regression	Yahoo	145 movies
	Hu et al. (2018)	Predicting box office success by combining first week movie reviews (number of reviews and sentiment) with other factors (basic movie information and external factors; e.g. special holiday)	Data mining technologies: M5 model trees, linear regression and support vector regression	IMDb	First week reviews of 1658 films released between 2009 and 2014
	Hur et al. (2016)	Predicting box office success with conventional predictors as well as sentiment of online reviews	Machine learning-based algorithms: Classification and regression tree, artificial neural network, and support vector regression	Naver	Reviews for 1200 films released in Korea (38,232 words)

(*table continues*)

Table S1 (*continued*)

Focus	Source	Object/Research problem	Methods	Database	Study Details
	J. H. Lee et al. (2017)	The influence of entropy of review text sentiments on online WOM and movie box office sales	Text-mining	Naver	256,998 reviews
	S. Lee and Choeh (2018)	Predicting box office success (length of review; and other numeric factors)	Regression	Naver	Reviews for 2090 movies
	Nagamma et al. (2015)	Predicting box office success with sentiment analysis and machine learning	Sentiment analysis with term frequency and inverse document frequency; prediction with support vecotor machines	IMDb	165 reviews
	X. Yu et al. (2012)	Predicting sales performance with review mining/sentiment analysis	Sentiment analysis with Sentiment PLSA; prediction with Autoregressive Sentiment-Aware model and Autoregressive Sentiment and Quality Aware model	IMDb	IMDB reviews and blog entries; 18,652 reviews and 45,046 blog entries

(*table continues*)

Table S1 (*continued*)

Focus	Source	Object/Research problem	Methods	Database	Study Details
Predicting sentiment	Amplayo and Song (2017)	Developing new method for generating summaries of multiple online reviews using a fine-grained sentiment extraction model for short texts	Sentiment classifier: three-level classification approach; aspect extractor: extended biterm topic model	Rotten Tomatoes, Naver, Douban	10,000 reviews each; max. character length of 140
	Bai (2011)	Proposing a model to capture the dependencies among words and provide a vocabulary that is adequate for the purpose of extracting sentiments	Heuristic search-enhanced Markov blanket model	IMDb	3400 reviews
	Liu et al. (2013)	Sentiment changes in online movie reviews; developing adaptive sentiment analysis models for online reviews	Adaptive latent semantic modeling techniques	IMDb	28,353 reviews for 20 drama films released in the US from 1 May 2006 to 1 September 2006
	Na et al. (2010)	Characteristics and differences in sentiment expression in movie review documents from four online opinion genres – blog postings, discussion board threads, user reviews, and critic reviews	Comparison based on statistical and linguistic aspects	IMDb	User reviews from IMDb; selection of 520 reviews

(*table continues*)

Table S1 (*continued*)

Focus	Source	Object/Research problem	Methods	Database	Study Details
	Pang and Lee (2005)	Developing new algorithm for multi-class text categorization	Meta-algorithm, based on a metric labeling formulation of the problem	-	Four corpora with movie reviews of the same author (1770, 902, 1307, 1027)
	Parkhe and Biswas (2016)	Sentiment analysis, identifying driving factors for movie reviews (result: movie, acting, plot)	Aspect-based sentiment analysis	IMDb	50,000 reviews from Stanford AI Lab
	Rani and Kumar (2019)	Deep learning model for sentiment analysis of online movie reviews in Hindi language	Convolutional neural network	Websites in Hindi language	7354 reviews
	Salvetti et al. (2004)	Polarity classification/sentiment analysis using supwervised machine learning mechanisms	Lexical filtering; two statistical classifiers (Naive Bayes and Markov Model)	IMDb	27,000 reviews from Cornell University
	Trivedi et al. (2018)	Comparing classifiers using natural language processing and various machine learning classifiers with Indian online movie reviews	Bayesian, naïve Bayesian and support vector machine	IMDb	2000 reviews of Indian movies released between 2000 and 2015

Table S1 (*continued*)

Focus	Source	Object/Research problem	Methods	Database	Study Details
	Yang and Yecies (2016)	Improving and testing big data analysis algorithms (problems of time constraints and memory costs)	Review crawling: asynchronous scraping crawler; content mining: Improved Apriori algorithm based on MapReduce	Douban	714,946 reviews
	Zhuang et al. (2006)	Proposing a multi-knowledge based movie review mining approach	Integrating WordNet, statistical analysis and movie knowledge	IMDb	1100 reviews
	Zin et al. (2017)	Testing performance of pre-processing strategies for sentiment analysis of online movie reviews	support vector machine (SVM) with linear and non-linear kernel	IMDb	2000 reviews from Pang & Lee (2005)
Other computational approaches	Bader et al. (2017)	Plutchik eight emotions, as well as positive and negative sentiment and their relation to movie genre	Calculating vector as emotional signature consisting of the Plutchik eight emotions	IMDb	717,498 reviews
	Otterbacher (2013)	Differences between reviews written by men vs women; comparing reviews with respect to writing style, content and metadata features	Logistic classification	IMDb	31,300 reviews (half male, half female reviewers)

Table S1 (*continued*)

Focus	Source	Object/Research problem	Methods	Database	Study Details
	Simmons et al. (2011)	Computer aided content analysis of movie reviews; result: Storyline is most important to moviegoers	CAINES (Content Analyzer and Information extraction System)	Yahoo	20,679 reviews
	Ullah et al. (2015)	Impact of emotions on helpfulness rating of movie reviews (result: emotional content considered more helpful)	Natural language processing; bag of words (positive/negative); coefficient correlation	IMDb	16,821
	Yang et al. (2020)	Proposing and testing algorithm for identifying opinion leadership on online movie evaluation sites	Cumulative distribution functions; Bayesian optimization-based imbalanced learning algorithm	Douban	17,786 reviews collected from 35 popular films covering a variety of genres

References for Supplementary Material

Amplayo, R. K., & Song, M. (2017). An adaptable fine-grained sentiment analysis for

summarization of multiple short online reviews. *Data & Knowledge Engineering*, *110*, 54–

67. https://doi.org/10.1016/j.datak.2017.03.009

Bader, N., Mokryn, O., & Lanir, J. (2017). Exploring emotions in online movie reviews for

online browsing. In G. A. Papadopoulos, T. Kuflik, F. Chen, C. Duarte, & W.-T. Fu (Eds.),

IUI'17: Companion of the 22nd International Conference on Intelligent User Interfaces:

March 13-16, 2017, Limassol, Cyprus (pp. 35–38). The Association for Computing

Machinery. https://doi.org/10.1145/3030024.3040982

Baek, H., Oh, S., Yang, H.-D., & Ahn, J. (2017). Electronic word-of-mouth, box office

revenue and social media. *Electronic Commerce Research and Applications*, *22*, 13–23.

https://doi.org/10.1016/j.elerap.2017.02.001

Bai, X. (2011). Predicting consumer sentiments from online text. *Decision Support Systems*,

50(4), 732–742. https://doi.org/10.1016/j.dss.2010.08.024

Hu, Y.-H., Shiau, W.-M., Shih, S.-P., & Chen, C.-J. (2018). Considering online consumer

reviews to predict movie box-office performance between the years 2009 and 2014 in the

US. *Electronic Library*, *36*(6), 1010–1026. https://doi.org/10.1108/EL-02-2018-0040

Hur, M., Kang, P., & Cho, S. (2016). Box-office forecasting based on sentiments of movie

reviews and Independent subspace method. *Information Sciences*, *372*, 608–624.

https://doi.org/10.1016/j.ins.2016.08.027

Lee, J. H., Jung, S. H., & Park, J. (2017). The role of entropy of review text sentiments on

online WOM and movie box office sales. *Electronic Commerce Research and*

Applications, *22*, 42–52. https://doi.org/10.1016/j.elerap.2017.03.001

Lee, S., & Choeh, J. Y. (2018). The interactive impact of online word-of-mouth and review helpfulness on box office revenue. *Management Decision*, *56*(4), 849–866. https://doi.org/10.1108/MD-06-2017-0561

Liu, Y., Yu, X., An, A., & Huang, X. (2013). Riding the tide of sentiment change: Sentiment analysis with evolving online reviews. *World Wide Web*, *16*(4), 477–496. https://doi.org/10.1007/s11280-012-0179-z

Na, J.-C., Thet, T. T., & Khoo, C. S.G. (2010). Comparing sentiment expression in movie reviews from four online genres. *Online Information Review*, *34*(2), 317–338. https://doi.org/10.1108/14684521011037016

Nagamma, P., Pruthvi, H. R., Nisha, K. K., & Shwetha, N. H. (2015). An improved sentiment analysis of online movie reviews based on clustering for box-office prediction. In A. Swaroop (Ed.), *International Conference on Computing, Communication & Automation (ICCCA), 2015: 15 - 16 May 2015, Greater Noida, UP (India)* (pp. 933–937). IEEE. https://doi.org/10.1109/CCAA.2015.7148530

Otterbacher, J. (2013). Gender, writing and ranking in review forums: A case study of the IMDb. *Knowledge and Information Systems*, *35*(3), 645–664. https://doi.org/10.1007/s10115-012-0548-z

Pang, B., & Lee, L. (2005). Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In Association for Computational Linguistics (Ed.), *ACL '05: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics* (pp. 115–124). Association for Computational Linguistics. https://doi.org/10.3115/1219840.1219855

Parkhe, V., & Biswas, B. (2016). Sentiment analysis of movie reviews: Finding most important movie aspects using driving factors. *Soft Computing*, *20*(9), 3373–3379. https://doi.org/10.1007/s00500-015-1779-1

Rani, S., & Kumar, P. (2019). Deep learning based sentiment analysis using convolution neural network. *Arabian Journal for Science and Engineering*, *44*(4), 3305–3314. https://doi.org/10.1007/s13369-018-3500-z

Salvetti, F., Lewis, S., & Reichenbach, C. (2004). Automatic opinion polarity classification of movie reviews. *Colorado Research in Linguistics*, *17*, 1–15. https://doi.org/10.25810/atv1-v819

Simmons, L. L., Mukhopadhyay, S., Conlon, S., & Yang, J [Jun] (2011). A computer aided content analysis of online reviews. *Journal of Computer Information Systems*, *52*(1), 43–55. https://doi.org/10.1080/08874417.2011.11645521

Trivedi, S. K., Dey, S., & Kumar, A. (2018). Capturing user sentiments for online Indian movie reviews: A comparative analysis of different machine-learning models. *Electronic Library*, *36*(4), 677–695. https://doi.org/10.1108/EL-04-2017-0075

Ullah, R., Zeb, A., & Kim, W. (2015). The impact of emotions on the helpfulness of movie reviews. *Journal of Applied Research and Technology*, *13*(3), 359–363. https://doi.org/10.1016/j.jart.2015.02.001

Yang, J [Jie], & Yecies, B. (2016). Mining Chinese social media UGC: A big-data framework for analyzing Douban movie reviews. *Journal of Big Data*, *3*(1), 3. https://doi.org/10.1186/s40537-015-0037-9

Yang, J [Jie], Yecies, B., & Zhong, P. Y. (2020). Characteristics of Chinese online movie reviews and opinion leadership identification. *International Journal of Human-Computer Interaction*, *36*(3), 211–226. https://doi.org/10.1080/10447318.2019.1625570

Yu, X., Liu, Y., Huang, X., & An, A. (2012). Mining online reviews for predicting sales performance: A case study in the movie domain. *IEEE Transactions on Knowledge and Data Engineering*, *24*(4), 720–734. https://doi.org/10.1109/TKDE.2010.269

Zhuang, L., Jing, F., & Zhu, X.-Y. (2006). Movie review mining and summarization. In P. S.

Yu, V. Tsotras, E. Fox, & B. Liu (Eds.), *Cikm '06: Proceedings of the 15th ACM*

international conference on information and knowledge management: Cikm 2006

Workshops ; November 6 - 11, 2006, Arlington, VA, USA (pp. 43–50). ACM Press.

https://doi.org/10.1145/1183614.1183625

Zin, H. M., Mustapha, N., Murad, M. A. A., & Sharef, N. M. (2017). The effects of pre-

processing strategies in sentiment analysis of online movie reviews. In F. A. A. Nifa, C. K.

Lin, & A. Hussain (Eds.), *Conference collection: Vol. 1891, The 2nd International*

Conference on Applied Science and Technology 2017 (ICAST'17): Kedah, Malaysia, 3-5

April 2017 (20089). AIP Publishing. https://doi.org/10.1063/1.5005422