

Influential Article Review - Using an Online Approach to Help End-users to Overcome Information Surplus

Chad Bell

Vincent Clark

Ernest Mack

This paper examines commerce and technology. We present insights from a highly influential paper. Here are the highlights from this paper: Redundant online reviews often have a negative impact on the efficiency of consumers' decision-making in their online shopping. A feasible solution for business analytics is to select a review subset from the original review corpus for consumers, which is called review selection. This study aims to address the diversified review selection problem, and proposes an effective review selection approach called Simulated Annealing-Diversified Review Selection (SA-DRS) that considers the semantic relationship of review features and the content diversity of selected reviews simultaneously. SA-DRS first constructs a feature taxonomy by utilizing the Latent Dirichlet Allocation (LDA) topic model and the Word2vec model to measure the topic relation and word context relation. Based on the established feature taxonomy, the similarity between each pair of reviews is defined and the review quality is estimated as well. Finally, diversified, high-quality reviews are selected heuristically by SA-DRS in the spirit of the simulated annealing method, forming the selected review subset. Extensive experiments are conducted on real-world e-commerce platforms to demonstrate the effectiveness of SA-DRS compared to other extant review selection approaches. For our overseas readers, we then present the insights from this paper in Spanish, French, Portuguese, and German.

Keywords: Business analytics, Online reviews, Feature taxonomy, Diversified subset, Review selection, Simulated annealing-diversified review selection (SA-DRS), Ecommerce

SUMMARY

- In this section, extensive comparative experiments are conducted to evaluate the effectiveness of SA-DRS. The experiments are carried out using real online review data collected from two typical e-commerce platform Tmall with Chinese reviews and Amazon with English reviews.
- Feature taxonomy results. Online reviews used in the experiments are first structured by feature extraction and sentiment analysis. In this paper, feature extraction is implemented by the method proposed by Hu and Liu , and sentiment analysis is implemented by the technique proposed by Ding et al. .

- Feature taxonomy of mobile phone in the Tmall dataset. From the feature taxonomy above, we can find that the root feature is quality which is pretty general in online reviews. It has several second-level features, including speed, price, etc., which reveals more detailed information than quality.
- Effectiveness experiment on Tmall and Amazon datasets. Greedy-U & Greedy-Q: Tsaparas et al. proposed several greedy approaches for review selection. Greedy-U and Greedy-Q are selected for comparison. Greedy-U aims to cover as many features as possible, while Greedy-Q aims to select reviews with high quality.
- Diversity measure comparison with the Tmall dataset. In Fig. 3, we see that SA-DRS always performs best on this metric and it achieves the largest diversity values all the time. The results demonstrate that SA-DRS can select a high-quality and diversified subset, which will assist potential consumers to make purchase decisions. In contrast, Greedy-DRS and Greedy-Q perform worse than SA-DRS, because neither of them considers the quality and diversity in selecting review subsets.
- Opinion distance comparison with the Tmall dataset. From the results shown in Fig. 4, we find that IR-CRS and Greedy-QARS rank first and second on the opinion distance measure, since they are designed just to optimize the consistency between opinions of the review corpus and the review subset.
- Coverage comparison with the Tmall dataset. According to Fig. 5, Greedy-U performs better than the other approaches in terms of coverage, since it directly aims to optimize the coverage object. In a general way, a high-quality review often covers the great majority of important features of the product.
- Diversity measure comparison with the Amazon dataset. Coverage comparison with the Amazon dataset. In order to compare these six approaches more clearly, we also rank their performance on the three measures above.
- Ranking comparison of the six approaches with the Amazon dataset. According to Table 4, we find that SA-DRS has the highest rank among six approaches, which is the same as the result for the Tmall dataset.
- Efficiency comparison. From Fig. 9, we can see all the approaches are efficient in review selection. Although SA-DRS is not the best in terms of efficiency, its running time is still comparable parallel with the other benchmark approaches. The reason why SA-DRS needs to take additional time is that it takes into consideration more factors such as the similarity calculation between two reviews which is time-consuming.

HIGHLY INFLUENTIAL ARTICLE

We used the following article as a basis of our evaluation:

Jin, Z., Zhangwen, W., & Naichen, N. (2019). Helping consumers to overcome information overload with a diversified online review subset. *Frontiers of Business Research in China*, 13(1), 1–25.

This is the link to the publisher's website:

<https://fbr.springeropen.com/articles/10.1186/s11782-019-0062-1>

INTRODUCTION

With the rapid development of the Internet, more and more people are writing textual reviews on e-commerce platforms, leading to a proliferation of online reviews. Prior to purchasing products, consumers usually read online reviews from previous buyers, demonstrating their trust in such kinds of user-generated content (Archak et al. 2011; Chen and Xie 2008). Online reviews provide detailed information about products and help customers better understand product quality and functions (Sun 2012). Thus, the valuable information provided by online reviews plays an important role in the purchase decision-making of potential consumers (Dhar and Chang 2009).

Resource-matching theory explains that the balance between the cognitive resources available to process the information and the mental resources required for the task is of huge importance in decision-making. When current cognitive resources and mental resources required are matched, the decision-making processing is more efficient and effective; otherwise decision performance is low (Anand et al. 1989; Mantel and Kellaris 2003). Nevertheless, on a popular e-commerce website, the online reviews of a product can easily exceed hundreds or even thousands (Hu and Liu 2004; Park and Lee 2009), leading to information overload (Bawden and Robinson 2009). Due to the limitations in customers' browsing time and the small screens of smart phones widely used in consumers' online shopping (the mental resources), it is impossible for consumers to read all related reviews about a particular product (the cognitive resources). This motivated firms to design an IT artifact to balance mental and cognitive resources, by providing consumers with a subset of reviews (e.g., the top 5 or 10 results) that reflect various aspects of the original large-scale review collection.

In recent years, review selection has become an attractive business intelligence and analytics research direction that aims to address the information overload problem of online reviews. In particular, some researchers have attempted to select a comprehensive subset of reviews, which would fully cover all opinions appearing in a review corpus (Tsaparas et al. 2011). Moreover, Lappas et al. (2012) examined the consistency of opinion proportions between the selected review subset and the original review corpus. E-commerce platforms like Amazon rank reviews according to their helpfulness votes or product ratings to provide consumers with several high-ranked helpful reviews. However, most research and e-commerce platforms rarely consider the diversity of selected subsets. As such, diversified review subset selection has not been well addressed. Furthermore, previous research simply assumes that the extracted features are of the same semantic level, which cannot reflect the multi-level features of certain products.

To overcome the limitations of previous studies, this research focuses on addressing the diversified review subset selection problem with the constraint of high-quality for selected reviews, where the semantic hierarchy of features has been properly considered. In particular, we apply the LDA topic model and the Word2vec model to measure the topic relation and context relation between each pair of features and then construct a feature taxonomy. Based on the established feature taxonomy, we calculate the similarity between each pair of reviews and then propose a diversified review selection approach similar to the simulated annealing method. Through extensive experiments on real data, we demonstrate that the proposed approach can select a diversified review subset, and provide concise, high-quality information for consumers in their decision-making.

From the perspective of design science research, this work can be positioned in the "improvement" quadrant of the design science knowledge contribution framework (Gregor and Hevner 2013). The review selection problem is a known research field which has been recognized by existing studies in e-commerce. To further improve the solutions to the problem, this study develops a new formulation of the online review selection problem taking into consideration both the diversity and quality of selected review subsets, and proposes a heuristic approach to effectively solve the problem. In the proposed approach, the multi-level features of certain products are well reflected by the feature taxonomy constructed based on the topic relation and content relation between features. To achieve a rigorous design evaluation, experimental analyses on real-world data collected from Tmall and Amazon have been conducted to demonstrate the performance advantages of the proposed approach over other approaches. In this way, the contributions of this study can be summarized as follows:

1. A diversified review selection problem considering both diversity and quality of selected review subsets is formulated in this study. In order to address this problem, a heuristic diversified review subset selection approach that can achieve high diversity and high quality in the selected results is proposed.
2. To reflect the multi-level features of certain products, a feature taxonomy construction method taking into consideration both topic relation and content relation is explored, based on which the diversity of reviews is measured.

The rest of the paper is organized as follows. Section 2 discusses related work on review selection, review helpfulness and quality prediction, and ontology learning. Section 3 defines the proposed diversified

review subset selection problem and provides an example to illustrate it. Section 4 proposes the feature taxonomy construction method and the diversified review subset selection approach, i.e., SA-DRS. Section 5 presents the experimental results on real data. Section 6 concludes the entire work and highlights some future research directions.

CONCLUSION

To help consumers overcome the information overload problem in e-commerce, this study focuses on the research on review selection and proposes an approach which can select a diversified and high-quality review subset. The proposed approach first establishes a feature taxonomy based on the topic relation and word context relation. Then, the diversified review selection problem is transformed into a constraint optimization problem according to the constructed feature taxonomy and is solved by a simulated annealing-like method. Extensive real data experiments conducted in this paper demonstrate the effectiveness of the proposed approach compared to other review selection approaches. According to the experiment results, the proposed approach can not only select a diversified review subset, but also have a good performance in terms of coverage and consistency as well.

Although there exist prior studies focused on coverage and/or other measures to find a representative subset, this paper selects a review subset considering diversity and quality based on the constructed feature taxonomy which contributes to the literature of ontology learning, review selection, and consumer behavior studies. First, this paper designs a feature taxonomy method based on the topic relation and word context relation to reveal the hierarchical structure of key features in user generated content. Second, the subset considering diversity based on the feature taxonomy also contributes to the literature with respect to opinion analysis for user generated content. The experimental results reveal that the proposed approach has superiority in selecting a high coverage and consistent subset. Last, the new IT artifact of the proposed review selection approach could have different influences on the decision-making process of consumers when online shopping, which further motivates behavioral science in terms of online consumer behavior studies.

The results of this study can provide technical support for e-commerce platforms. It is very common that consumers cannot read the whole body of reviews of certain products. However, when they are keen to master all details of the candidate products and read high-quality reviews in a short time simultaneously, e-commerce platforms can show a diversified review subset selected by the proposed approach to help consumers improve their decision-making efficiency. When consumers read the selected subsets, they will quickly grasp the features of the product and the useful feedback of other consumers. Therefore, the implications of this work are twofold. For one thing, the application of the proposed approach can tailor the browsing time of consumers, enhance their satisfaction and help them make purchase decisions. For the other, the proposed approach can simplify the review corpus on e-commerce platforms and raise the readability of large-scale online reviews which are very common in the era of big data.

Future work can be focused on two aspects. One is to incorporate more factors into consideration when estimating the quality of reviews, such as the reputation of authors, the temporal information of reviews and the length of review. The other is to add other factors like the semantic words shared by features into the proposed approach when exploring the semantic relations of extracted features, which will help establish a more feasible feature taxonomy.

APPENDIX

FIGURE 1 THE FRAMEWORK OF THE REVIEW SELECTION APPROACH

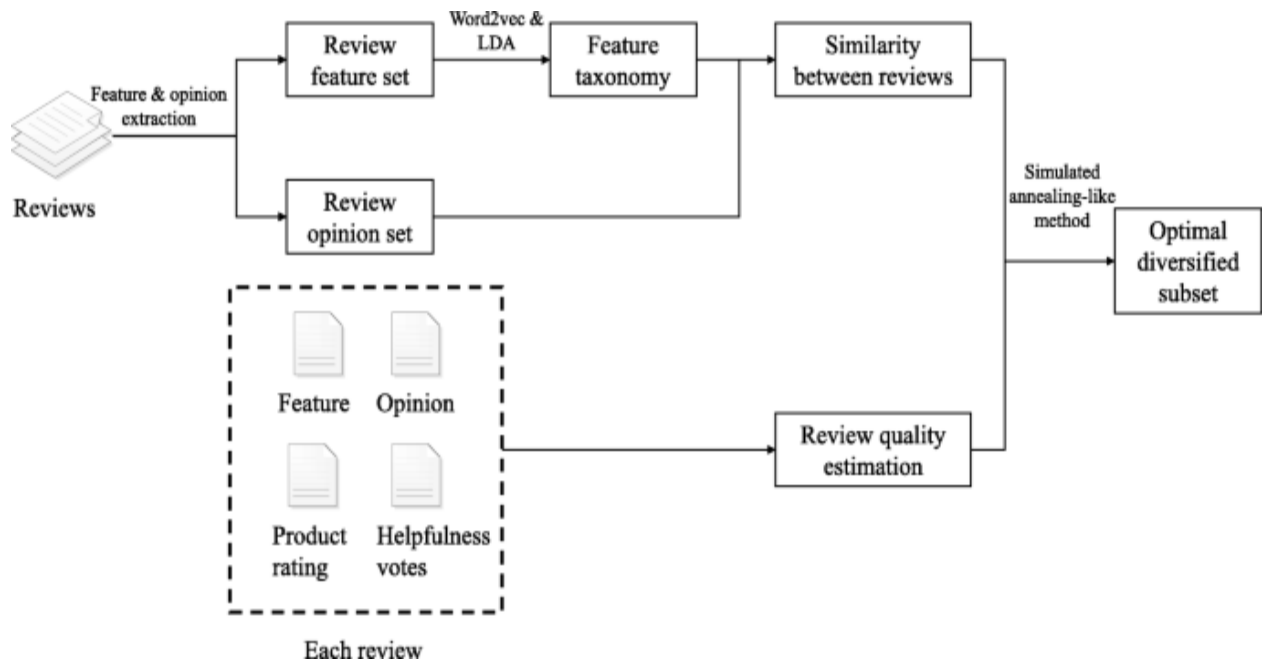


FIGURE 2
FEATURE TAXONOMY OF MOBILE PHONE IN THE TMALL DATASET

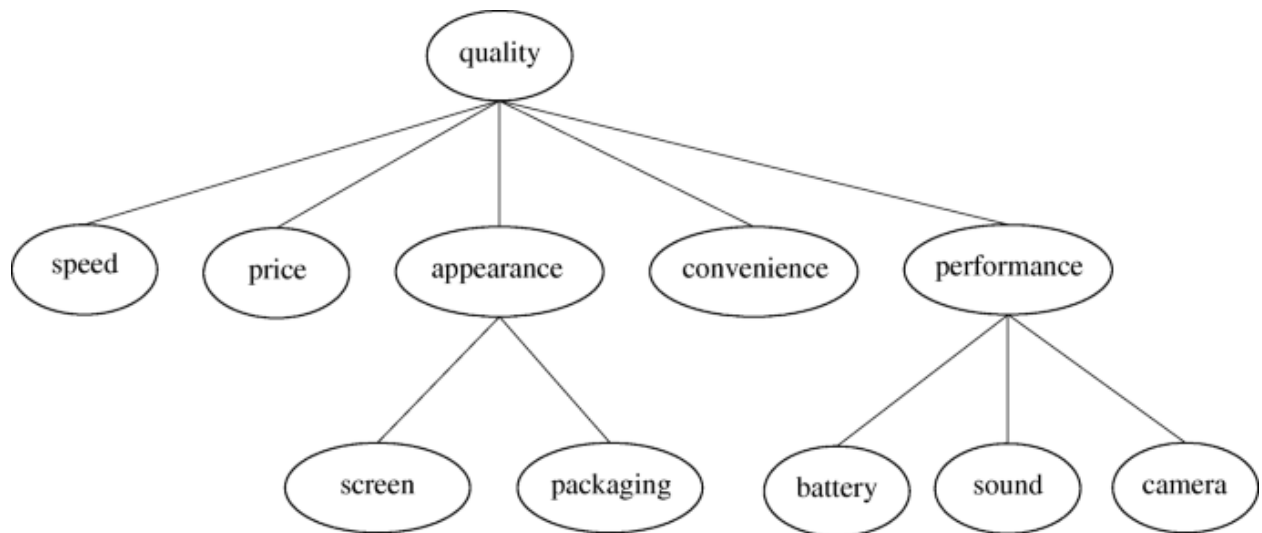


FIGURE 3

DIVERSITY MEASURE COMPARISON WITH THE TMALL DATASET

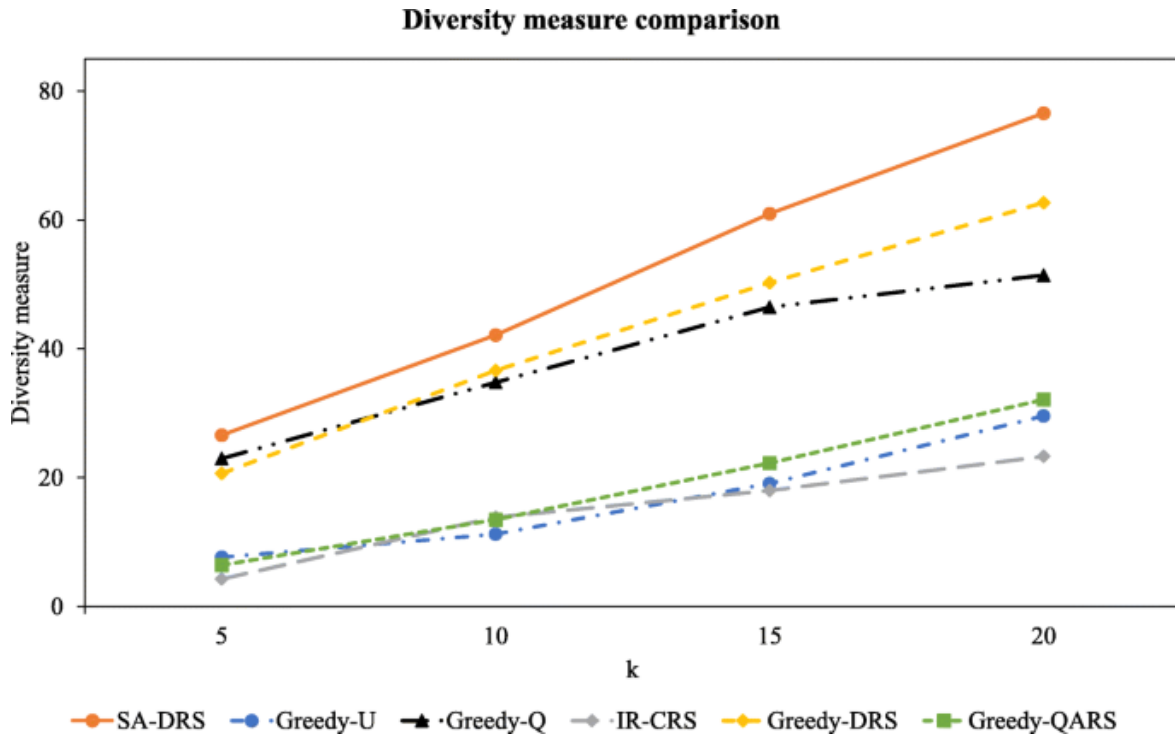


FIGURE 4
OPINION DISTANCE COMPARISON WITH THE TMALL DATASET

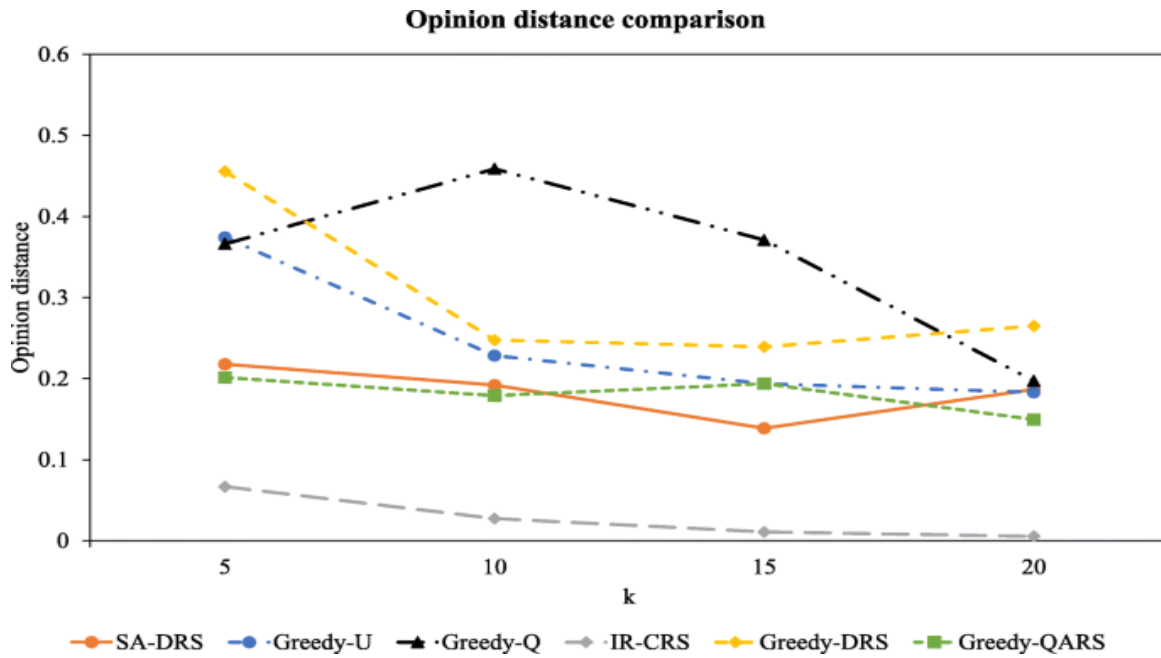


FIGURE 5

COVERAGE COMPARISON WITH THE TMALL DATASET

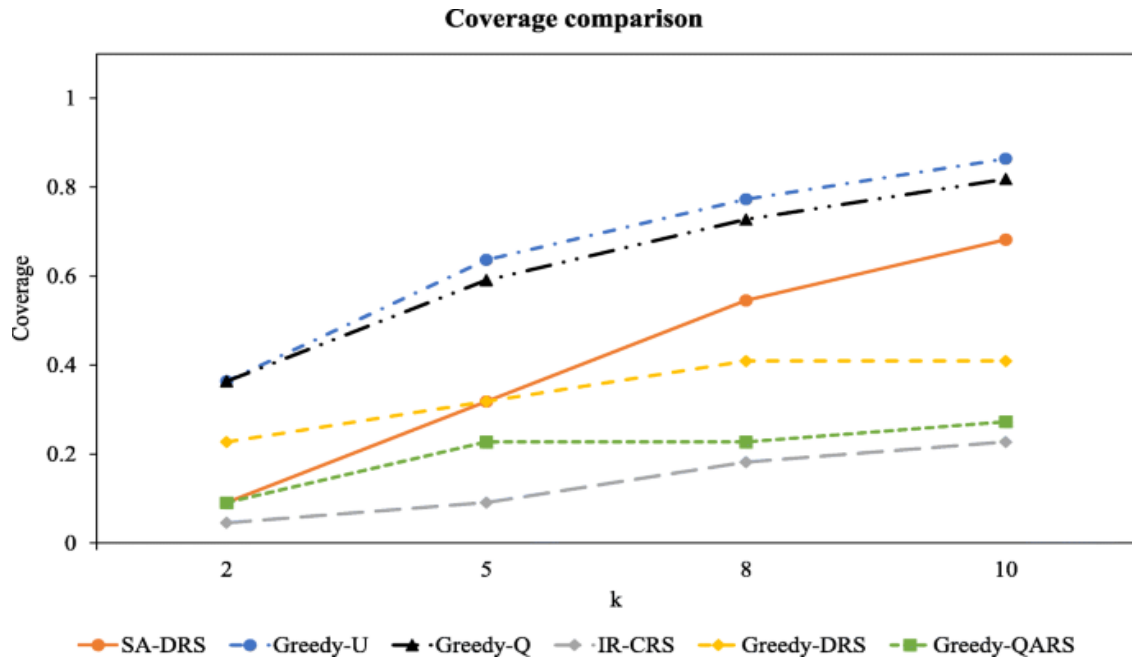


FIGURE 6
DIVERSITY MEASURE COMPARISON WITH THE AMAZON DATASET

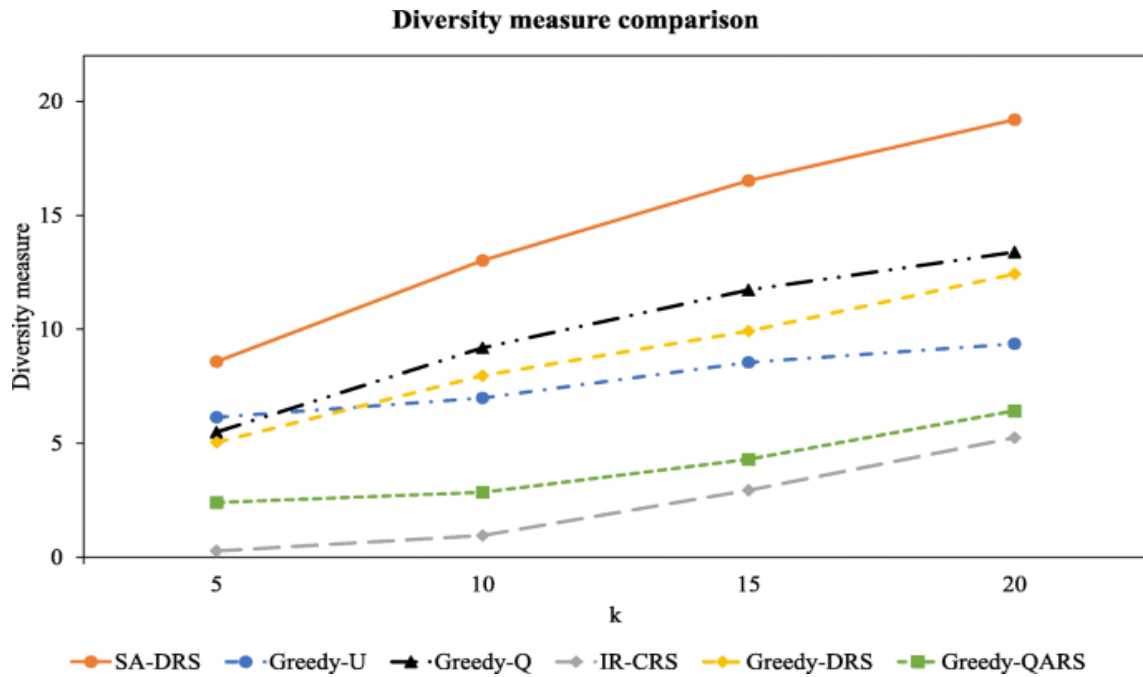


FIGURE 7

OPINION DISTANCE COMPARISON WITH THE AMAZON DATASET

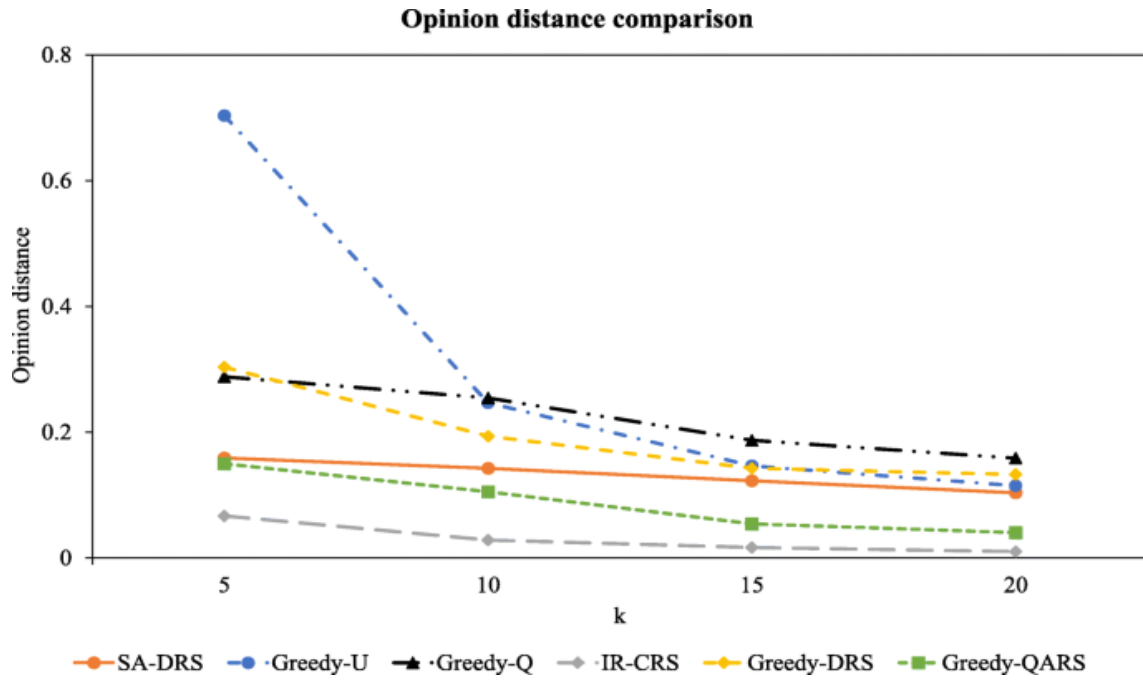


FIGURE 8
COVERAGE COMPARISON WITH THE AMAZON DATASET

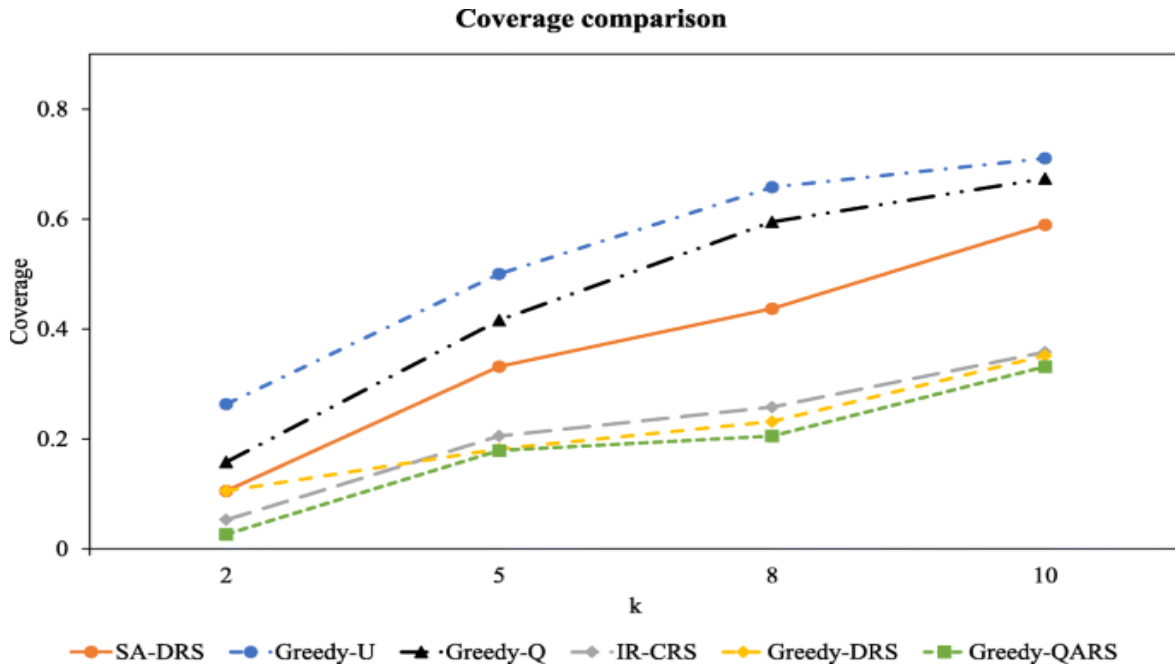


FIGURE 9

EFFICIENCY COMPARISON

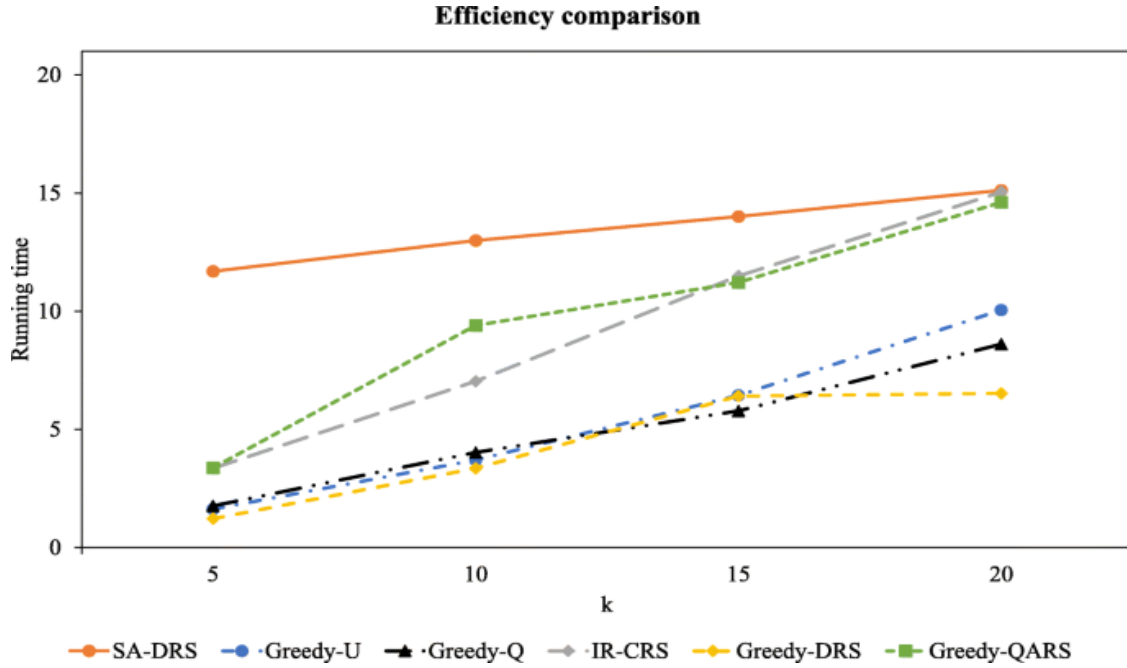


TABLE 1
OBJECTIVE FUNCTION VALUES OF REVIEW SUBSETS

S	Value	S	Value
$\{r_1, r_2, r_3\}$	6.2162	$\{r_2, r_3, r_4\}$	6.1111
$\{r_1, r_2, r_4\}$	5.8333	$\{r_2, r_3, r_5\}$	5.2500
$\{r_1, r_2, r_5\}$	4.9180	$\{r_2, r_3, r_6\}$	6.5217
$\{r_1, r_2, r_6\}$	4.8718	$\{r_2, r_4, r_5\}$	5.2778
$\{r_1, r_3, r_4\}$	6.6316	$\{r_2, r_4, r_6\}$	5.1923
$\{r_1, r_3, r_5\}$	5.6075	$\{r_2, r_5, r_6\}$	5.4839
$\{r_1, r_3, r_6\}$	5.5882	$\{r_3, r_4, r_5\}$	2.1591
$\{r_1, r_4, r_5\}$	5.2941	$\{r_3, r_4, r_6\}$	2.0769
$\{r_1, r_4, r_6\}$	4.2149	$\{r_3, r_5, r_6\}$	2.0648
$\{r_1, r_5, r_6\}$	4.3636	$\{r_4, r_5, r_6\}$	1.7510

TABLE 2
EXPERIMENT DATASETS

Platform	Category	Product	# of Average reviews
Tmall	Books	61	672.0
	Digital cameras	53	180.3
	Laptops	60	1110.3
	Mobile phones	61	2507.6
	Diapers	59	4745.2
	Shirts	70	596.1
Amazon	Random	89	306.5

TABLE 3
RANKING COMPARISON OF THE SIX APPROACHES WITH THE TMALL DATASET

Approach	Diversity rank	Opinion distance rank	Coverage rank	Average rank
SA-DRS	1	3	3	2.3
Greedy-U	6	4	1	3.7
Greedy-Q	3	6	2	3.7
IR-CRS	4	1	6	3.7
Greedy-DRS	2	5	4	3.7
Greedy-QARS	5	2	5	4.0

TABLE 4
RANKING COMPARISON OF THE SIX APPROACHES WITH THE AMAZON DATASET

Approach	Diversity rank	Opinion distance rank	Coverage rank	Average rank
SA-DRS	1	3	3	2.3
Greedy-U	4	4	1	3.0
Greedy-Q	2	6	2	3.3
Greedy-DRS	3	5	4	4.0
Greedy-QARS	5	2	5	4.0
IR-CRS	6	1	6	4.3

REFERENCES

- Anand, P., & Sternthal, B. (1989). Strategies for designing persuasive messages: Deductions from the resource matching hypothesis. In P. Cafferata & A. M. Tybout (Eds.), *Cognitive and affective responses to advertising* (pp. 135–159). Lexington: Lexington Books/D. C. Heath and Com.
- Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Social Science Electronic Publishing*, 57(8), 1485–1509.
- Bawden, D., & Robinson, L. (2009). The dark side of information: Overload, anxiety and other paradoxes and pathologies. *Journal of Information Science*, 35(2), 180–191.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(4–5), 993–1022.
- Chen, C. C., & Tseng, Y. D. (2011). Quality evaluation of product reviews using an information quality framework. *Decision Support Systems*, 50(4), 755–768.
- Chen, Y., & Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science*, 54(3), 477–491.
- Dhar, V., & Chang, E. A. (2009). Does chatter matter? The impact of user-generated content on music sales. *Journal of Interactive Marketing*, 23(4), 300–307.

- Ding, X., Liu, B., & Yu, P. S. (2008). A holistic lexicon-based approach to opinion mining. In Proceedings of the 2008 international conference on web search and data mining (pp. 231–240).
- Djuana, E., Xu, Y., Li, Y., & Cox, C. (2012). Personalization in tag ontology learning for recommendation making. In Proceedings of the 14th international conference on information integration and web-based applications and services (pp. 368–377).
- Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, 23(10), 1498–1512.
- Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, 37(3), 337–355.
- Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 168–177).
- Jin, J., Ji, P., & Yan, S. X. (2017). Comparison of series products from customer online concerns for competitive intelligence. *Journal of Ambient Intelligence & Humanized Computing*, 10(3), 937–952.
- Jindal, N., & Liu, B. (2008). Opinion spam and analysis. In Proceedings of the 2008 international conference on web search and data mining (pp. 219–230).
- Kim, S. M., Pantel, P., Chklovski, T., & Pennacchiotti, M. (2006). Automatically assessing review helpfulness. In Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (pp. 423–430).
- Korfatis, N., García-Bariocanal, E., & Sánchez-Alonso, S. (2012). Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications*, 11(3), 205–217.
- Lappas, T., Crovella, M., & Terzi, E. (2012). Selecting a characteristic set of reviews. In Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining (pp. 832–840).
- Lee, I. (2018). Usefulness, funniness, and coolness votes of viewers: An analysis of social shoppers' online reviews. *Industrial Management & Data Systems*, 118(4), 700–713.
- Liu, J., Cao, Y., Lin, C. Y., Huang, Y., & Zhou, M. (2007). Low-quality product review detection in opinion summarization. In Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (pp. 334–342).
- Liu, Y., Huang, X., An, A., & Yu, X. (2008). Modeling and predicting the helpfulness of online reviews. In Proceedings of the 8th international conference on data mining (pp. 443–452).
- Maedche, A., & Staab, S. (2001). Ontology learning for the semantic web. *Intelligent Systems IEEE*, 16(2), 72–79.
- Mantel, S. P., & Kellaris, J. J. (2003). Cognitive determinants of consumers' time perceptions: The impact of resources required and available. *Journal of Consumer Research*, 29(4), 531–538.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (Vol. 26, pp. 3111–3119). USA: Curran Associates, Inc.
- Park, D. H., & Lee, J. (2009). Ewom overload and its effect on consumer behavioral intention depending on consumer involvement. *Electronic Commerce Research & Applications*, 7(4), 386–398.
- Paul, D., Sarkar, S., Chelliah, M., Kalyan, C., & Sinai Nadkarni, P. P. (2017). Recommendation of high quality representative reviews in e-commerce. In Proceedings of the 11th ACM Conference on Recommender Systems (pp. 311–315).
- Suman, B., & Kumar, P. (2006). A survey of simulated annealing as a tool for single and multiobjective optimization. *Journal of the Operational Research Society*, 57(10), 1143–1160.
- Sun, M. (2012). How does the variance of product ratings matter? *Management Science*, 58(4), 696–707.

- Tang, J., Leung, H. F., Luo, Q., Chen, D., & Gong, J. (2009). Towards ontology learning from folksonomies. In Proceedings of the 21st international joint conference on artificial intelligence (pp. 2089–2094).
- Tian, N., Xu, Y., Li, Y., & Gabriella, P. (2015). Quality-aware review selection based on product feature taxonomy. In Proceedings of the 11th Asia information retrieval societies conference (pp. 68–80).
- Tian, N., Xu, Y., Li, Y., Abdel-Hafez, A., & Josang, A. (2014). Product feature taxonomy learning based on user reviews. In In WEBIST 2014 10th International Conference on Web Information Systems and Technologies (pp. 184–192).
- Tsaparas, P., Ntoulas, A., & Terzi, E. (2011). Selecting a comprehensive set of reviews. In Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining (pp. 168–176).
- Yu, W., Zhang, R., He, X., & Sha, C. (2013). Selecting a diversified set of reviews. In Ishikawa, Y. Li, J. Wang, W. Zhang, R. & Zhang W. (Eds.), Web technologies and applications. APWeb 2013. Lecture notes in computer science (Vol. 7808). Berlin: Springer.

TRANSLATED VERSION: SPANISH

Below is a rough translation of the insights presented above. This was done to give a general understanding of the ideas presented in the paper. Please excuse any grammatical mistakes and do not hold the original authors responsible for these mistakes.

VERSION TRADUCIDA: ESPAÑOL

A continuación se muestra una traducción aproximada de las ideas presentadas anteriormente. Esto se hizo para dar una comprensión general de las ideas presentadas en el documento. Por favor, disculpe cualquier error gramatical y no responsabilite a los autores originales de estos errores.

INTRODUCCIÓN

Con el rápido desarrollo de Internet, cada vez más personas están escribiendo reseñas textuales en plataformas de comercio electrónico, lo que conduce a una proliferación de revisiones en línea. Antes de comprar productos, los consumidores suelen leer opiniones en línea de compradores anteriores, demostrando su confianza en este tipo de contenido generado por el usuario (Archak et al. 2011; Chen y Xie 2008). Las revisiones en línea proporcionan información detallada sobre los productos y ayudan a los clientes a comprender mejor la calidad y las funciones del producto (Dom 2012). Por lo tanto, la valiosa información proporcionada por los exámenes en línea desempeña un papel importante en la toma de decisiones de compra de los consumidores potenciales (Dhar y Chang 2009).

La teoría de la coincidencia de recursos explica que el equilibrio entre los recursos cognitivos disponibles para procesar la información y los recursos mentales necesarios para la tarea es de gran importancia en la toma de decisiones. Cuando se emparejan los recursos cognitivos actuales y los recursos mentales necesarios, el procesamiento de la toma de decisiones es más eficiente y eficaz; de lo contrario el rendimiento de la decisión es bajo (Anand et al. 1989; Mantel y Kellaris 2003). Sin embargo, en un popular sitio web de comercio electrónico, las opiniones en línea de un producto pueden superar fácilmente cientos o incluso miles (Hu y Liu 2004; Park y Lee 2009), lo que lleva a una sobrecarga de información (Bawden y Robinson 2009). Debido a las limitaciones en el tiempo de navegación de los clientes y las pequeñas pantallas de los teléfonos inteligentes ampliamente utilizados en las compras en línea de los consumidores (los recursos mentales), es imposible para los consumidores leer todas las opiniones relacionadas sobre un producto en particular (los recursos cognitivos). Esto motivó a las empresas a diseñar un artefacto de TI para equilibrar los recursos mentales y cognitivos, proporcionando a los consumidores un subconjunto de opiniones (por ejemplo, los 5 o 10 mejores resultados) que reflejan varios aspectos de la colección de revisiones original a gran escala.

En los últimos años, la selección de revisiones se ha convertido en una atractiva dirección de investigación de inteligencia empresarial y análisis que tiene como objetivo abordar el problema de sobrecarga de información de las revisiones en línea. En particular, algunos investigadores han intentado seleccionar un subconjunto completo de revisiones, que cubriría plenamente todas las opiniones que aparecen en un corpus de revisión (Tsaparas et al. 2011). Además, Lappas et al. (2012) examinaron la coherencia de las proporciones de opinión entre el subconjunto de revisión seleccionado y el corpus de revisión original. Plataformas de comercio electrónico como Amazon clasifican opiniones de acuerdo con sus votos de utilidad o calificaciones de productos para proporcionar a los consumidores con varias revisiones útiles de alto rango. Sin embargo, la mayoría de las plataformas de investigación y comercio electrónico rara vez consideran la diversidad de subconjuntos seleccionados. Como tal, la selección diversificada de subconjuntos de reseñas no ha sido bien abordada. Además, investigaciones anteriores simplemente asumen que las características extraídas son del mismo nivel semántico, que no puede reflejar las características multinivel de ciertos productos.

Para superar las limitaciones de estudios anteriores, esta investigación se centra en abordar el problema de selección diversificada de subconjuntos de revisión con la restricción de alta calidad para revisiones seleccionadas, donde la jerarquía semántica de características se ha considerado correctamente. En particular, aplicamos el modelo de tema LDA y el modelo de Word2vec para medir la relación de tema y la relación de contexto entre cada par de entidades y, a continuación, construimos una taxonomía de entidades. Basándonos en la taxonomía de características establecida, calculamos la similitud entre cada par de revisiones y luego proponemos un enfoque de selección de revisión diversificado similar al método de recocido simulado. A través de extensos experimentos con datos reales, demostramos que el enfoque propuesto puede seleccionar un subconjunto de opiniones diversificado y proporcionar información concisa y de alta calidad a los consumidores en la toma de decisiones.

Desde la perspectiva de la investigación científica del diseño, este trabajo se puede posicionar en el cuadrante de "mejora" del marco de contribución al conocimiento de la ciencia del diseño (Gregor y Hevner 2013). El problema de selección de revisión es un campo de investigación conocido que ha sido reconocido por los estudios existentes en comercio electrónico. Para mejorar aún más las soluciones al problema, este estudio desarrolla una nueva formulación del problema de selección de la revisión en línea teniendo en cuenta tanto la diversidad como la calidad de los subconjuntos de revisión seleccionados, y propone un enfoque heurístico para resolver eficazmente el problema. En el enfoque propuesto, las características de varios niveles de ciertos productos se reflejan bien en la taxonomía de características construida sobre la base de la relación temática y la relación de contenido entre entidades. Para lograr una evaluación rigurosa del diseño, se han realizado análisis experimentales sobre los datos del mundo real recopilados de Tmall y Amazon para demostrar las ventajas de rendimiento del enfoque propuesto sobre otros enfoques. De esta manera, las contribuciones de este estudio se pueden resumir de la siguiente manera:

1. En este estudio se formula un problema de selección de revisión diversificado teniendo en cuenta tanto la diversidad como la calidad de los subconjuntos de opiniones seleccionados. Para abordar este problema, se propone un enfoque heurístico de selección de subconjuntos de reseñas diversificadas que puede lograr una alta diversidad y alta calidad en los resultados seleccionados.
2. Para reflejar las características de varios niveles de ciertos productos, se explora un método de construcción de taxonomía de características teniendo en cuenta tanto la relación temática como la relación de contenido, en función de la cual se mide la diversidad de revisiones.

El resto del documento se organiza de la siguiente manera. En la Sección 2 se analiza el trabajo relacionado sobre la selección de revisiones, la revisión de la utilidad y la predicción de calidad, y el aprendizaje ontológico. La Sección 3 define el problema de selección de subconjuntos de revisión diversificado propuesto y proporciona un ejemplo para ilustrarlo. La Sección 4 propone el método de construcción de taxonomía de características y el enfoque de selección diversificada de subconjuntos de revisión, es decir, SA-DRS. La Sección 5 presenta los resultados experimentales sobre datos reales. La Sección 6 concluye todo el trabajo y destaca algunas direcciones de investigación futuras.

CONCLUSIÓN

Para ayudar a los consumidores a superar el problema de sobrecarga de información en el comercio electrónico, este estudio se centra en la investigación sobre la selección de revisiones y propone un enfoque que puede seleccionar un subconjunto de revisión diversificado y de alta calidad. El enfoque propuesto establece primero una taxonomía de entidades basada en la relación temática y la relación de contexto de palabras. A continuación, el problema de selección de revisión diversificada se transforma en un problema de optimización de restricciones de acuerdo con la taxonomía de características construidas y se resuelve mediante un método simulado de recocido similar. Los extensos experimentos de datos reales realizados en este documento demuestran la eficacia del enfoque propuesto en comparación con otros enfoques de selección de revisiones. Según los resultados del experimento, el enfoque propuesto no sólo puede seleccionar un subconjunto de revisión diversificado, sino también tener un buen rendimiento en términos de cobertura y coherencia también.

Aunque existen estudios previos centrados en la cobertura y/u otras medidas para encontrar un subconjunto representativo, este documento selecciona un subconjunto de revisión teniendo en cuenta la diversidad y la calidad basadas en la taxonomía de características construidas que contribuye a la literatura del aprendizaje ontología, la selección de revisiones y los estudios de comportamiento del consumidor. En primer lugar, este documento diseña un método de taxonomía de características basado en la relación de tema y la relación de contexto de palabras para revelar la estructura jerárquica de las características clave en el contenido generado por el usuario. En segundo lugar, el subconjunto que considera la diversidad basada en la taxonomía de características también contribuye a la literatura con respecto al análisis de opinión para el contenido generado por el usuario. Los resultados experimentales revelan que el enfoque propuesto tiene superioridad en la selección de una alta cobertura y subconjunto consistente. Por último, el nuevo artefacto de TI del enfoque de selección de revisión propuesto podría tener diferentes influencias en el proceso de toma de decisiones de los consumidores cuando compra en línea, lo que motiva aún más la ciencia del comportamiento en términos de estudios de comportamiento del consumidor en línea.

Los resultados de este estudio pueden proporcionar soporte técnico para plataformas de comercio electrónico. Es muy común que los consumidores no puedan leer todo el conjunto de comentarios de ciertos productos. Sin embargo, cuando están interesados en dominar todos los detalles de los productos candidatos y leer opiniones de alta calidad en poco tiempo simultáneamente, las plataformas de comercio electrónico pueden mostrar un subconjunto de opiniones diversificado seleccionado por el enfoque propuesto para ayudar a los consumidores a mejorar su eficiencia en la toma de decisiones. Cuando los consumidores leen los subconjuntos seleccionados, comprenderán rápidamente las características del producto y los comentarios útiles de otros consumidores. Por lo tanto, las implicaciones de este trabajo son dobles. Por un lado, la aplicación del enfoque propuesto puede adaptar el tiempo de navegación de los consumidores, mejorar su satisfacción y ayudarles a tomar decisiones de compra. Por otro, el enfoque propuesto puede simplificar el corpus de revisión en las plataformas de comercio electrónico y aumentar la legibilidad de las revisiones en línea a gran escala que son muy comunes en la era del big data.

El trabajo futuro puede centrarse en dos aspectos. Uno es incorporar más factores en consideración a la hora de estimar la calidad de las revisiones, como la reputación de los autores, la información temporal de las revisiones y la duración de la revisión. El otro es añadir otros factores como las palabras semánticas compartidas por las características en el enfoque propuesto al explorar las relaciones semánticas de las características extraídas, lo que ayudará a establecer una taxonomía de características más factible.

TRANSLATED VERSION: FRENCH

Below is a rough translation of the insights presented above. This was done to give a general understanding of the ideas presented in the paper. Please excuse any grammatical mistakes and do not hold the original authors responsible for these mistakes.

VERSION TRADUITE: FRANÇAIS

Voici une traduction approximative des idées présentées ci-dessus. Cela a été fait pour donner une compréhension générale des idées présentées dans le document. Veuillez excuser toutes les erreurs grammaticales et ne pas tenir les auteurs originaux responsables de ces erreurs.

INTRODUCTION

Avec le développement rapide de l'Internet, de plus en plus de gens écrivent des critiques textuelles sur les plateformes de commerce électronique, conduisant à une prolifération d'avis en ligne. Avant d'acheter des produits, les consommateurs lisent habituellement les commentaires en ligne des acheteurs précédents, démontrant leur confiance dans ce genre de contenu généré par les utilisateurs (Archak et coll., 2011; Chen et Xie, 2008). Les revues en ligne fournissent des informations détaillées sur les produits et aident les clients à mieux comprendre la qualité et les fonctions du produit (Dim 2012). Ainsi, les renseignements précieux fournis par les examens en ligne jouent un rôle important dans la prise de décisions d'achat des consommateurs potentiels (Dhar et Chang, 2009).

La théorie de l'appariement des ressources explique que l'équilibre entre les ressources cognitives disponibles pour traiter l'information et les ressources mentales requises pour la tâche est d'une importance énorme dans la prise de décision. Lorsque les ressources cognitives et mentales actuelles requises sont jumelées, le traitement décisionnel est plus efficace et efficient; autrement, le rendement des décisions est faible (Anand et coll., 1989; Mantel et Kellaris, 2003). Néanmoins, sur un site web populaire de commerce électronique, les critiques en ligne d'un produit peuvent facilement dépasser des centaines, voire des milliers (Hu et Liu, 2004; Park et Lee, 2009), ce qui a mené à une surcharge d'information (Bawden et Robinson, 2009). En raison des limites du temps de navigation des clients et des petits écrans de téléphones intelligents largement utilisés dans les achats en ligne des consommateurs (les ressources mentales), il est impossible pour les consommateurs de lire toutes les critiques connexes sur un produit particulier (les ressources cognitives). Cela a incité les entreprises à concevoir un artefact de l'it pour équilibrer les ressources mentales et cognitives, en fournissant aux consommateurs un sous-ensemble d'examen (p. Ex., les 5 ou 10 premiers résultats) qui reflètent divers aspects de la collection originale d'examen à grande échelle.

Au cours des dernières années, la sélection des examens est devenue une orientation attrayante en matière de recherche en intelligence d'affaires et en analyse qui vise à résoudre le problème de la surcharge d'information des examens en ligne. En particulier, certains chercheurs ont tenté de choisir un sous-ensemble complet d'examen, qui couvrirait entièrement toutes les opinions figurant dans un corpus d'examen (Tsaparas et coll., 2011). De plus, Lappas et coll. (2012) ont examiné la cohérence des proportions d'opinion entre le sous-ensemble d'examen sélectionné et le corpus d'examen original. Les plateformes de commerce électronique comme Amazon classent les avis en fonction de leurs votes utiles ou des évaluations de produits pour fournir aux consommateurs plusieurs avis utiles de haut rang. Toutefois, la plupart des plateformes de recherche et de commerce électronique considèrent rarement la diversité des sous-ensembles sélectionnés. Par conséquent, la sélection diversifiée des sous-ensembles d'examen n'a pas été bien abordée. En outre, des recherches antérieures supposent simplement que les caractéristiques extraites sont du même niveau sémantique, ce qui ne peut pas refléter les caractéristiques à plusieurs niveaux de certains produits.

Pour surmonter les limites des études antérieures, cette recherche se concentre sur la résolution du problème de sélection des sous-ensembles d'examen diversifié avec la contrainte de haute qualité pour certains examens, où la hiérarchie sémantique des caractéristiques a été correctement prise en compte. En particulier, nous appliquons le modèle de sujet LDA et le modèle Word2vec pour mesurer la relation de sujet et la relation contextuelle entre chaque paire de fonctionnalités, puis construisons une taxonomie de fonctionnalités. Sur la base de la taxonomie des fonctionnalités établie, nous calculons la similitude entre chaque paire d'examen, puis proposons une approche diversifiée de sélection des examens similaire à la

méthode d'annealage simulée. Grâce à des expériences approfondies sur des données réelles, nous démontrons que l'approche proposée peut sélectionner un sous-ensemble d'examen diversifié et fournir des renseignements concis et de haute qualité aux consommateurs dans leur prise de décision.

Du point de vue de la recherche scientifique de conception, ce travail peut être placé dans le quadrant « d'amélioration » du cadre de contribution de connaissance de science de conception (Gregor et Hevner 2013). Le problème de la sélection des revues est un domaine de recherche connu qui a été reconnu par les études existantes dans le commerce électronique. Afin d'améliorer davantage les solutions au problème, cette étude élabore une nouvelle formulation du problème de sélection des revues en ligne en tenant compte à la fois de la diversité et de la qualité des sous-ensembles d'examen sélectionnés, et propose une approche heuristique pour résoudre efficacement le problème. Dans l'approche proposée, les caractéristiques à plusieurs niveaux de certains produits sont bien reflétées par la taxonomie des fonctionnalités construite en fonction de la relation thématique et de la relation de contenu entre les caractéristiques. Afin d'obtenir une évaluation rigoureuse de la conception, des analyses expérimentales sur des données réelles recueillies auprès de Tmall et d'Amazon ont été menées pour démontrer les avantages de performance de l'approche proposée par rapport à d'autres approches. De cette façon, les contributions de cette étude peuvent être résumées comme suit :

1. Un problème diversifié de sélection des revues, compte tenu de la diversité et de la qualité des sous-ensembles d'examen sélectionnés, est formulé dans cette étude. Afin de résoudre ce problème, une approche heuristique de sélection diversifiée des sous-ensembles d'examen qui peut atteindre une grande diversité et une grande qualité dans les résultats sélectionnés est proposée.
2. Afin de refléter les caractéristiques à plusieurs niveaux de certains produits, une méthode de construction de taxonomie en tenant compte à la fois de la relation thématique et de la relation de contenu est explorée, en fonction de laquelle la diversité des examens est mesurée.

Le reste du document est organisé comme suit. La section 2 traite des travaux connexes sur la sélection de l'examen, l'examen de l'utilité et de la prévision de la qualité, et l'apprentissage de l'ontologie. L'article 3 définit le problème proposé de sélection des sous-ensembles d'examen diversifié et fournit un exemple pour l'illustrer. La section 4 propose la méthode de construction de la taxonomie des caractéristiques et l'approche diversifiée de sélection des sous-ensembles d'examen, c'est-à-dire sa-DRS. La section 5 présente les résultats expérimentaux sur des données réelles. La section 6 conclut l'ensemble des travaux et met en évidence certaines orientations de recherche futures.

CONCLUSION

Afin d'aider les consommateurs à surmonter le problème de la surcharge d'information dans le commerce électronique, cette étude met l'accent sur la recherche sur la sélection des revues et propose une approche qui peut choisir un sous-ensemble d'examen diversifié et de haute qualité. L'approche proposée établit d'abord une taxonomie des fonctionnalités basée sur la relation de sujet et la relation de contexte de mot. Ensuite, le problème de sélection des revues diversifiées se transforme en un problème d'optimisation des contraintes en fonction de la taxonomie des fonctionnalités construites et est résolu par une méthode d'annealage simulée. De vastes expériences de données réelles menées dans le présent document démontrent l'efficacité de l'approche proposée par rapport à d'autres approches de sélection des revues. Selon les résultats de l'expérience, l'approche proposée peut non seulement sélectionner un sous-ensemble d'examen diversifié, mais aussi avoir une bonne performance en termes de couverture et d'uniformité.

Bien qu'il existe des études antérieures axées sur la couverture et/ou d'autres mesures visant à trouver un sous-ensemble représentatif, cet article sélectionne un sous-ensemble d'examen s'appuyant sur la diversité et la qualité en fonction de la taxonomie des caractéristiques construites qui contribue à la littérature de l'apprentissage de l'ontologie, de la sélection des revues et des études sur le comportement des consommateurs. Tout d'abord, cet article conçoit une méthode de taxonomie des fonctionnalités basée sur la relation de sujet et la relation de contexte de mot pour révéler la structure hiérarchique des fonctionnalités clés dans le contenu généré par l'utilisateur. Deuxièmement, le sous-ensemble tenant compte de la diversité fondée sur la taxonomie des fonctionnalités contribue également à la littérature en

ce qui concerne l'analyse d'opinion pour le contenu généré par l'utilisateur. Les résultats expérimentaux révèlent que l'approche proposée a la supériorité dans le choix d'une couverture élevée et d'un sous-ensemble cohérent. Enfin, le nouvel artefact informatique de l'approche proposée de sélection de l'examen pourrait avoir des influences différentes sur le processus décisionnel des consommateurs lors des achats en ligne, ce qui motive davantage la science comportementale en termes d'études sur le comportement des consommateurs en ligne.

Les résultats de cette étude peuvent fournir un soutien technique aux plateformes de commerce électronique. Il est très courant que les consommateurs ne puissent pas lire l'ensemble des examens de certains produits. Toutefois, lorsqu'elles souhaitent maîtriser tous les détails des produits candidats et lire des revues de haute qualité en peu de temps simultanément, les plateformes de commerce électronique peuvent présenter un sous-ensemble d'examen diversifié sélectionné par l'approche proposée pour aider les consommateurs à améliorer leur efficacité décisionnelle. Lorsque les consommateurs liront les sous-ensembles sélectionnés, ils saisiront rapidement les caractéristiques du produit et les commentaires utiles des autres consommateurs. Par conséquent, les implications de ce travail sont doubles. D'une part, l'application de l'approche proposée peut adapter le temps de navigation des consommateurs, améliorer leur satisfaction et les aider à prendre des décisions d'achat. D'autre part, l'approche proposée peut simplifier le corpus d'examen sur les plateformes de commerce électronique et augmenter la lisibilité des revues en ligne à grande échelle qui sont très courantes à l'ère du Big Data.

Les travaux futurs peuvent être axés sur deux aspects. La première est d'intégrer davantage de facteurs en considération lors de l'estimation de la qualité des examens, tels que la réputation des auteurs, l'information temporelle des revues et la durée de l'examen. L'autre est d'ajouter d'autres facteurs comme les mots sémantiques partagés par les caractéristiques dans l'approche proposée lors de l'exploration des relations sémantiques des caractéristiques extraites, ce qui aidera à établir une taxonomie caractéristique plus réalisable.

TRANSLATED VERSION: GERMAN

Below is a rough translation of the insights presented above. This was done to give a general understanding of the ideas presented in the paper. Please excuse any grammatical mistakes and do not hold the original authors responsible for these mistakes.

ÜBERSETZTE VERSION: DEUTSCH

Hier ist eine ungefähre Übersetzung der oben vorgestellten Ideen. Dies wurde getan, um ein allgemeines Verständnis der in dem Dokument vorgestellten Ideen zu vermitteln. Bitte entschuldigen Sie alle grammatikalischen Fehler und machen Sie die ursprünglichen Autoren nicht für diese Fehler verantwortlich.

EINLEITUNG

Mit der rasanten Entwicklung des Internets schreiben immer mehr Menschen Textrezensionen auf E-Commerce-Plattformen, was zu einer Vielzahl von Online-Bewertungen führt. Vor dem Kauf von Produkten lesen Verbraucher in der Regel Online-Bewertungen von früheren Käufern und zeigen ihr Vertrauen in solche nutzergenerierten Inhalte (Archak et al. 2011; Chen und Xie 2008). Online-Bewertungen bieten detaillierte Informationen zu Produkten und helfen Kunden, die Produktqualität und -funktionen besser zu verstehen (Sun 2012). Daher spielen die wertvollen Informationen, die durch Online-Bewertungen bereitgestellt werden, eine wichtige Rolle bei der Kaufentscheidung potenzieller Verbraucher (Dhar und Chang 2009).

Die Theorie des Ressourcenabgleichs erklärt, dass das Gleichgewicht zwischen den kognitiven Ressourcen, die für die Verarbeitung der Informationen zur Verfügung stehen, und den mentalen Ressourcen, die für die Aufgabe erforderlich sind, von enormer Bedeutung für die Entscheidungsfindung

ist. Wenn die aktuellen kognitiven Ressourcen und mentalen Ressourcen aufeinander abgestimmt sind, ist die Entscheidungsfindung effizienter und effektiver; andernfalls ist die Entscheidungsleistung gering (Anand et al. 1989; Mantel und Kellaris 2003). Dennoch, auf einer beliebigen E-Commerce-Website, die Online-Bewertungen eines Produkts kann leicht Hunderte oder sogar Tausende überschreiten (Hu und Liu 2004; Park und Lee 2009), was zu Informationsüberlastung führt (Bawden und Robinson 2009). Aufgrund der Einschränkungen in der Browser-Zeit der Kunden und die kleinen Bildschirme von Smartphones weit verbreitet in den Online-Shopping der Verbraucher (die geistigen Ressourcen) verwendet, ist es unmöglich für Verbraucher, alle damit verbundenen Bewertungen über ein bestimmtes Produkt (die kognitiven Ressourcen) zu lesen. Dies motivierte Unternehmen, ein IT-Artefakt zu entwerfen, um geistige und kognitive Ressourcen auszugleichen, indem sie den Verbrauchern eine Teilmenge von Bewertungen (z. B. Die Top 5 oder 10 Ergebnisse) zur Verfügung stellen, die verschiedene Aspekte der ursprünglichen groß angelegten Review-Sammlung widerspiegeln.

In den letzten Jahren hat sich die Überprüfungsauswahl zu einer attraktiven Forschungsrichtung für Business Intelligence und Analytics entwickelt, die darauf abzielt, das Problem der Informationsüberlastung von Online-Bewertungen anzugehen. Insbesondere haben einige Forscher versucht, eine umfassende Teilmenge von Bewertungen auszuwählen, die alle Meinungen, die in einem Review Corpus erscheinen, vollständig abdecken würde (Tsaparas et al. 2011). Darüber hinaus untersuchten Lappas et al. (2012) die Kohärenz der Meinungsproportionen zwischen der ausgewählten Bewertungsteilmenge und dem ursprünglichen Bewertungskorpus. E-Commerce-Plattformen wie Amazon bewerten Bewertungen nach ihren Hilfsstimmen oder Produktbewertungen, um Verbrauchern mehrere hochrangige hilfreiche Bewertungen zu bieten. Die meisten Forschungs- und E-Commerce-Plattformen berücksichtigen jedoch selten die Vielfalt ausgewählter Teilmengen. Daher wurde die Auswahl diversifizierter Überprüfungsteilmengen nicht gut berücksichtigt. Darüber hinaus gehen frühere Forschungen einfach davon aus, dass die extrahierten Merkmale den gleichen semantischen Grad aufweisen, was die mehrstufigen Eigenschaften bestimmter Produkte nicht widerspiegeln kann.

Um die Grenzen früherer Studien zu überwinden, konzentriert sich diese Forschung auf die Lösung des Diversifizierten Auswahlproblems für die Teilmenge der Überprüfung mit der Einschränkung von hoher Qualität für ausgewählte Bewertungen, bei denen die semantische Hierarchie der Merkmale richtig berücksichtigt wurde. Insbesondere wenden wir das LDA-Themenmodell und das Word2vec-Modell an, um die Themenbeziehung und Kontextbeziehung zwischen den einzelnen Features-Paaren zu messen und dann eine Feature-Taxonomie zu erstellen. Basierend auf der etablierten Feature-Taxonomie berechnen wir die Ähnlichkeit zwischen den einzelnen Überprüfungen und schlagen dann einen diversifizierten Bewertungsauswahlansatz ähnlich der simulierten Glühmethode vor. Durch umfangreiche Experimente zu realen Daten zeigen wir, dass der vorgeschlagene Ansatz eine diversifizierte Bewertungsuntergruppe auswählen und den Verbrauchern präzise und qualitativ hochwertige Informationen bei ihrer Entscheidungsfindung liefern kann.

Aus der Perspektive der designwissenschaftlichen Forschung kann diese Arbeit im "Verbesserungs"-Quadranten des designwissenschaftlichen Wissensbeitragsrahmens positioniert werden (Gregor und Hevner 2013). Das Problem der Überprüfungsauswahl ist ein bekanntes Forschungsfeld, das durch bestehende Studien im Bereich E-Commerce anerkannt wurde. Um die Lösung des Problems weiter zu verbessern, entwickelt diese Studie eine neue Formulierung des Online-Bewertungsproblems unter Berücksichtigung sowohl der Vielfalt als auch der Qualität ausgewählter Review-Teilgruppen und schlägt einen heuristischen Ansatz zur wirksamen Lösung des Problems vor. Im vorgeschlagenen Ansatz spiegeln sich die mehrstufigen Merkmale bestimmter Produkte gut in der Feature-Taxonomie wider, die auf der Themenbeziehung und der Inhaltlichkeit zwischen Features basiert. Um eine strenge Entwurfsbewertung zu erreichen, wurden experimentelle Analysen zu realen Daten durchgeführt, die von Tmall und Amazon gesammelt wurden, um die Leistungsvorteile des vorgeschlagenen Ansatzes gegenüber anderen Ansätzen aufzuzeigen. Auf diese Weise lassen sich die Beiträge dieser Studie wie folgt zusammenfassen:

1. In dieser Studie wird ein diversifiziertes Problem bei der Auswahl von Überprüfungen formuliert, das sowohl die Vielfalt als auch die Qualität ausgewählter Bewertungsteilmengen berücksichtigt.

- Um dieses Problem anzugehen, wird ein heuristisch diversifizierter Bewertungsansatz vorgeschlagen, der eine hohe Vielfalt und Qualität in den ausgewählten Ergebnissen erreichen kann.
2. Um die mehrstufigen Merkmale bestimmter Produkte widerzuspiegeln, wird eine Feature-Taxonomie-Konstruktionsmethode untersucht, die sowohl die Themenbeziehung als auch die Inhaltlichkeit berücksichtigt, basierend auf deren Grundlage die Vielfalt der Bewertungen gemessen wird.

Der Rest des Papiers ist wie folgt organisiert. In Abschnitt 2 werden die damit zusammenhängenden Arbeiten zur Überprüfungsauswahl, zur Überprüfung der Hilfs- und Qualitätsvorhersage sowie zum Lernen in der Ontologie erläutert. Abschnitt 3 definiert das vorgeschlagene Problem der Auswahl diversifizierter Überprüfungen und enthält ein Beispiel, um es zu veranschaulichen. In Abschnitt 4 werden die Konstruktionsmethode für die Feature-Taxonomie und der Ansatz für die Auswahl diversifizierter Überprüfungen, d. H. SA-DRS, vorgeschlagen. In Abschnitt 5 werden die experimentellen Ergebnisse zu realen Daten dargestellt. Abschnitt 6 schließt die gesamte Arbeit ab und zeigt einige zukünftige Forschungsrichtungen auf.

SCHLUSSFOLGERUNG

Um den Verbrauchern zu helfen, das Problem der Informationsüberlastung im elektronischen Geschäftsverkehr zu überwinden, konzentriert sich diese Studie auf die Forschung zur Auswahl von Überprüfungen und schlägt einen Ansatz vor, der eine diversifizierte und qualitativ hochwertige Bewertungsuntergruppe auswählen kann. Der vorgeschlagene Ansatz legt zunächst eine Feature-Taxonomie auf der Grundlage der Themenbeziehung und der Wortkontextbeziehung fest. Anschließend wird das diversifizierte Überprüfungsauswahlproblem entsprechend der konstruierten Feature-Taxonomie in ein Problem der Einschränkungsoptimierung umgewandelt und durch eine simulierte glühende Methode gelöst. Umfangreiche reale Datenexperimente, die in diesem Papier durchgeführt wurden, zeigen die Wirksamkeit des vorgeschlagenen Ansatzes im Vergleich zu anderen Bewertungsauswahlansätzen. Nach den Versuchsergebnissen kann der vorgeschlagene Ansatz nicht nur eine diversifizierte Bewertungsuntergruppe auswählen, sondern auch eine gute Leistung in Bezug auf Abdeckung und Konsistenz aufweisen.

Obwohl es frühere Studien gibt, die sich auf die Abdeckung und/oder andere Maßnahmen zur Suche nach einer repräsentativen Teilmenge konzentrieren, wählt dieses Dokument eine Review-Teilmenge aus, die Vielfalt und Qualität auf der Grundlage der konstruierten Feature-Taxonomie berücksichtigt, die zur Literatur des ontologischen Lernens, der Überprüfungsauswahl und von Verbraucherverhaltensstudien beiträgt. Zunächst entwirft dieses Dokument eine Feature-Taxonomiemethode, die auf der Themenbeziehung und der Wortkontextbeziehung basiert, um die hierarchische Struktur der wichtigsten Features in benutzergenerierten Inhalten aufzuzeigen. Zweitens trägt die Teilmenge, die die Diversität auf der Grundlage der Feature-Taxonomie berücksichtigt, auch zur Literatur in Bezug auf die Meinungsanalyse für nutzergenerierte Inhalte bei. Die experimentellen Ergebnisse zeigen, dass der vorgeschlagene Ansatz bei der Auswahl einer hohen Abdeckung und einer konsistenten Teilmenge eine Überlegenheit aufweist. Schließlich könnte das neue IT-Artefakt des vorgeschlagenen Bewertungsauswahlansatzes unterschiedliche Einflüsse auf den Entscheidungsprozess der Verbraucher beim Online-Shopping haben, was die Verhaltenswissenschaft in Bezug auf Online-Verbraucherverhaltensstudien weiter motiviert.

Die Ergebnisse dieser Studie können technische Unterstützung für E-Commerce-Plattformen bieten. Es ist sehr häufig, dass Verbraucher nicht den ganzen Körper der Bewertungen bestimmter Produkte lesen können. Wenn sie jedoch alle Details der Kandidatenprodukte beherrschen und gleichzeitig qualitativ hochwertige Bewertungen lesen möchten, können E-Commerce-Plattformen eine diversifizierte Bewertungsuntergruppe aufweisen, die durch den vorgeschlagenen Ansatz ausgewählt wurde, um den Verbrauchern zu helfen, ihre Entscheidungseffizienz zu verbessern. Wenn Verbraucher die ausgewählten Teilmengen lesen, werden sie schnell die Funktionen des Produkts und das nützliche Feedback anderer Verbraucher verstehen. Daher sind die Auswirkungen dieser Arbeit zweierlei. Zum einen kann die Anwendung des vorgeschlagenen Ansatzes die Browserzeit der Verbraucher anpassen, ihre Zufriedenheit

steigern und ihnen helfen, Kaufentscheidungen zu treffen. Zum anderen kann der vorgeschlagene Ansatz den Überprüfungskorpus auf E-Commerce-Plattformen vereinfachen und die Lesbarkeit von groß angelegten Online-Bewertungen erhöhen, die im Zeitraum von Big Data sehr häufig sind.

Die künftige Arbeit kann sich auf zwei Aspekte konzentrieren. Eine besteht darin, bei der Schätzung der Qualität von Bewertungen mehr Faktoren zu berücksichtigen, wie z. B. Den Ruf von Autoren, die zeitlichen Informationen von Bewertungen und die Dauer der Überprüfung. Die andere besteht darin, andere Faktoren wie die semantischen Wörter, die von Features geteilt werden, in den vorgeschlagenen Ansatz einzuschließen, wenn die semantischen Beziehungen extrahierter Features untersucht werden, was dazu beitragen wird, eine praktikablere Feature-Taxonomie zu etablieren.

TRANSLATED VERSION: PORTUGUESE

Below is a rough translation of the insights presented above. This was done to give a general understanding of the ideas presented in the paper. Please excuse any grammatical mistakes and do not hold the original authors responsible for these mistakes.

VERSÃO TRADUZIDA: PORTUGUÊS

Aqui está uma tradução aproximada das ideias acima apresentadas. Isto foi feito para dar uma compreensão geral das ideias apresentadas no documento. Por favor, desculpe todos os erros gramaticais e não responsabilize os autores originais responsáveis por estes erros.

INTRODUÇÃO

Com o rápido desenvolvimento da Internet, cada vez mais pessoas estão a escrever comentários textuais em plataformas de e-commerce, levando a uma proliferação de críticas online. Antes de comprar produtos, os consumidores geralmente lêem comentários online de compradores anteriores, demonstrando a sua confiança em tais tipos de conteúdo gerado pelo utilizador (Archak et al. 2011; Chen e Xie 2008). As análises online fornecem informações detalhadas sobre produtos e ajudam os clientes a compreender melhor a qualidade e funções do produto (Sun 2012). Assim, as informações valiosas fornecidas pelas avaliações online desempenham um papel importante na tomada de decisões de compra de potenciais consumidores (Dhar e Chang 2009).

A teoria da correspondência de recursos explica que o equilíbrio entre os recursos cognitivos disponíveis para processar a informação e os recursos mentais necessários para a tarefa é de enorme importância na tomada de decisões. Quando os recursos cognitivos atuais e os recursos mentais necessários são compatíveis, o processamento de decisões é mais eficiente e eficaz; caso contrário, o desempenho da decisão é baixo (Anand et al. 1989; Mantel e Kellaris 2003). No entanto, num popular site de e-commerce, as avaliações online de um produto podem facilmente ultrapassar centenas ou mesmo milhares (Hu e Liu 2004; Park e Lee 2009), levando à sobrecarga de informação (Bawden e Robinson 2009). Devido às limitações no tempo de navegação dos clientes e aos pequenos ecrãs de smartphones amplamente utilizados nas compras online dos consumidores (os recursos mentais), é impossível para os consumidores lerem todas as avaliações relacionadas sobre um determinado produto (os recursos cognitivos). Isto motivou as empresas a conceberem um artefacto de TI para equilibrar os recursos mentais e cognitivos, fornecendo aos consumidores um subconjunto de avaliações (por exemplo, os 5 ou 10 melhores resultados) que refletem vários aspetos da coleção original de revisão em larga escala.

Nos últimos anos, a seleção de revisão tornou-se uma direção de pesquisa de inteligência empresarial e análise atraente que visa resolver o problema da sobrecarga de informação de avaliações online. Em particular, alguns investigadores tentaram selecionar um subconjunto abrangente de revisões, que abrangeria plenamente todos os pareceres que aparecem num corpus de revisão (Tsaparas et al. 2011). Além disso, Iappas et al. (2012) examinaram a consistência das proporções de opinião entre o subconjunto de

revisão selecionado e o corpus de revisão original. Plataformas de e-commerce como a Amazon classificam as avaliações de acordo com os seus votos de utilidade ou classificações de produtos para fornecer aos consumidores várias avaliações úteis de alto nível. No entanto, a maioria das plataformas de pesquisa e e-commerce raramente consideram a diversidade de subconjuntos selecionados. Como tal, a seleção diversificada de subconjuntos de revisão não foi bem abordada. Além disso, estudos anteriores simplesmente assumem que as características extraídas são do mesmo nível semântico, o que não pode refletir as características a vários níveis de determinados produtos.

Para superar as limitações de estudos anteriores, esta investigação centra-se em abordar o problema diversificado de seleção de subconjuntos de revisão com a restrição de alta qualidade para avaliações selecionadas, onde a hierarquia semântica das características foi devidamente considerada. Em particular, aplicamos o modelo de tópico LDA e o modelo Word2vec para medir a relação de tópico e a relação de contexto entre cada par de funcionalidades e, em seguida, construir uma taxonomia de recurso. Com base na taxonomia de recurso estabelecida, calculamos a semelhança entre cada par de avaliações e, em seguida, propomos uma abordagem de seleção de revisão diversificada semelhante ao método de annealing simulado. Através de experiências extensivas sobre dados reais, demonstramos que a abordagem proposta pode selecionar um subconjunto de revisão diversificado e fornecer informações concisas e de alta qualidade para os consumidores na sua tomada de decisão.

Do ponto de vista da investigação científica do design, este trabalho pode ser posicionado no quadrante de "melhoria" do quadro de contribuição para o conhecimento da ciência do design (Gregor e Hevner 2013). O problema da seleção de revisão é um campo de investigação conhecido que foi reconhecido pelos estudos existentes no comércio eletrónico. Para melhorar ainda mais as soluções para o problema, este estudo desenvolve uma nova formulação do problema de seleção de revisão online, tendo em conta tanto a diversidade como a qualidade dos subconjuntos de revisão selecionados, e propõe uma abordagem heurística para resolver eficazmente o problema. Na abordagem proposta, as características a vários níveis de certos produtos refletem-se bem na taxonomia de recurso construída com base na relação de tópicos e na relação de conteúdo entre as características. Para obter uma avaliação rigorosa do design, foram realizadas análises experimentais sobre dados do mundo real recolhidos da Tmall e da Amazon para demonstrar as vantagens de desempenho da abordagem proposta sobre outras abordagens. Desta forma, as contribuições deste estudo podem ser resumidas da seguinte forma:

1. Neste estudo é formulado um problema diversificado de seleção de revisão, tendo em conta a diversidade e a qualidade dos subconjuntos de revisão selecionados. Para resolver este problema, é proposta uma abordagem de seleção de subconjuntos de revisão diversificada heurística que pode alcançar uma elevada diversidade e elevada qualidade nos resultados selecionados.
2. Para refletir as características multi-nível de determinados produtos, é explorado um método de construção de taxonomia de recurso, tendo em conta tanto a relação de tópicos como a relação de conteúdo, com base na qual a diversidade de revisões é medida.

O resto do jornal é organizado da seguinte forma. A secção 2 discute trabalhos relacionados na seleção de revisão, revisão da utilidade e previsão de qualidade e aprendizagem de ontologia. A secção 3 define o problema de seleção de subconjuntos de revisão diversificado proposto e constitui um exemplo para o ilustrar. A secção 4 propõe o método de construção da taxonomia de recurso e a abordagem diversificada de seleção de subconjuntos de revisão, ou seja, SA-DRS. A secção 5 apresenta os resultados experimentais em dados reais. A secção 6 conclui todo o trabalho e destaca algumas futuras direções de investigação.

CONCLUSÃO

Para ajudar os consumidores a superar o problema da sobrecarga de informação no comércio eletrónico, este estudo centra-se na investigação sobre a seleção de revisão e propõe uma abordagem que pode selecionar um subconjunto de revisão diversificado e de alta qualidade. A abordagem proposta estabelece primeiro uma taxonomia de recurso baseada na relação de tópico e no contexto da palavra. Em seguida, o problema diversificado de seleção de revisão é transformado em um problema de otimização de restrição de acordo com a taxonomia de recurso construído e é resolvido por um método simulado de annealing.

Experiências de dados reais extensivas realizadas neste artigo demonstram a eficácia da abordagem proposta em comparação com outras abordagens de seleção de revisão. De acordo com os resultados da experiência, a abordagem proposta não só pode selecionar um subconjunto de revisão diversificado, como também ter um bom desempenho em termos de cobertura e consistência.

Embora existam estudos anteriores focados na cobertura e/ou outras medidas para encontrar um subconjunto representativo, este artigo seleciona um subconjunto de revisão considerando a diversidade e a qualidade com base na taxonomia de características construídas que contribui para a literatura de aprendizagem ontologia, seleção de revisão e estudos de comportamento do consumidor. Em primeiro lugar, este artigo desenha um método de taxonomia de recurso baseado na relação de relação de tópico e contexto de palavras para revelar a estrutura hierárquica das principais funcionalidades no conteúdo gerado pelo utilizador. Em segundo lugar, o subconjunto que considera a diversidade baseada na taxonomia de recurso também contribui para a literatura no que diz respeito à análise de opinião para conteúdos gerados pelo utilizador. Os resultados experimentais revelam que a abordagem proposta tem superioridade na seleção de um subconjunto de alta cobertura e consistente. Por último, o novo artefacto de TI da abordagem de seleção de revisão proposta poderia ter influências diferentes no processo de tomada de decisão dos consumidores nas compras online, o que motiva ainda mais a ciência comportamental em termos de estudos de comportamento dos consumidores online.

Os resultados deste estudo podem fornecer suporte técnico para plataformas de e-commerce. É muito comum que os consumidores não possam ler todo o conjunto das revisões de determinados produtos. No entanto, quando estão interessados em dominar todos os detalhes dos produtos candidatos e ler avaliações de alta qualidade em simultâneo, as plataformas de e-commerce podem mostrar um subconjunto de revisão diversificado selecionado pela abordagem proposta para ajudar os consumidores a melhorar a sua eficiência na tomada de decisões. Quando os consumidores lerem os subconjuntos selecionados, compreenderão rapidamente as características do produto e o feedback útil de outros consumidores. Por conseguinte, as implicações deste trabalho são duplas. Por um lado, a aplicação da abordagem proposta pode adaptar o tempo de navegação dos consumidores, aumentar a sua satisfação e ajudá-los a tomar decisões de compra. Por outro lado, a abordagem proposta pode simplificar o corpus de revisão nas plataformas de comércio eletrónico e aumentar a legibilidade de revisões online em larga escala que são muito comuns na era dos big data.

O trabalho futuro pode centrar-se em dois aspetos. Uma delas é incorporar mais fatores em consideração na estimativa da qualidade dos comentários, como a reputação dos autores, a informação temporal das revisões e a duração da revisão. O outro é adicionar outros fatores, como as palavras semânticas partilhadas pelas características na abordagem proposta ao explorar as relações semânticas das características extraídas, o que ajudará a estabelecer uma taxonomia de recurso mais viável.