Bridging the gap between human-gaze data and table summarisation

Jessica Amianto Barbato[1], Marco Cremaschi[1]

[1]University of Milan - Bicocca, Milan, Italy

Abstract

Understanding users' reading behaviour can facilitate and support the development of data lexicalisation models based on users' characteristics. A significant amount of data can be found online in tabular form, and several models have been developed to provide the user with summaries of the content of such tables. Nevertheless, studies analysing table reading patterns are almost entirely lacking in the literature, making it almost impossible to integrate findings on user reading behaviour into lexicalisation models. This work aims to suggest a new line of gaze-related research that can integrate insights about user behaviour and characteristics into data summarisation algorithms to provide textual content that meets the user's information needs. An overview of human-gaze studies applied to natural language will be presented to outline a study on human interaction with tables. Potential fields of application and challenges in applying these results to the field of table summarisation, namely the task of producing short summaries of tabular data, from a user-centred perspective will be discussed.

Keywords

table summarisation, eye-tracking studies, human-gaze data, data summarisation, lexicalisation, human intepretation

1. Introduction

When a user reads a text, their eyes make rapid and unconscious movements from one point to another. These movements are called *saccades*. Among several saccades, the eyes stop on portions of the text; it is only during these *fixations* that the content is retained and processed.

Text features and user characteristics drive these movements. For example, language determines reading direction; in English texts, most of the saccades have left-to-right movements. Users, however, may encounter unfamiliar words or syntactic constructions that violate the rules of the language they know [1][1]; these situations will lead the user to reread portions of the text, thus performing saccades backwards, from right-to-left. Also the duration of fixations is highly dependent on the text features; content words, especially adjectives and verbs, elicit longer fixations, while articles, conjunctions, prepositions, and pronouns and functional words are often skipped [1]. In addition, users tend to make different fixations patterns depending on their knowledge of the language [3] and the context.

Italian Workshop on Artificial Intelligence for Human Machine Interaction (AIxHMI 2022), December 02, 2022, Udine, Italy

jessica.amiantobarbato@unimib.it (J. Amianto Barbato); marco.cremaschi@unimib.it (M. Cremaschi)
0000-0002-9831-3210 (J. Amianto Barbato); 0000-0001-7840-6228 (M. Cremaschi)

[1]See the garden-path model [2] for further discussion.

Important previous works in cognitive psychology and linguistics investigate how a user reads a text, whether in print or on-screen [4, 5]. These studies have been integrated into many Natural Language Processing (NLP) and Natural Language Generation (NLG) approaches that rely on how the user reads, retains and processes a text [6, 7, 8, 9, 10]. These instances of prior research, however, analyse texts that commonly have the following characteristics: (i) they are sequential, such as Wikipedia pages [10, 6], and (ii) they have a context or, if they are decontextualised, their location and surrounding content provide a context for the interpretation [1, 11].

Currently, however, a vast amount of information is provided as structured data in tables. This increase can be linked to the uptake of the Open Data movement, whose purpose is to make a large number of tabular data sources freely available, addressing a wide range of domains, such as finance, mobility, tourism, sports, or cultural heritage [12]. The phenomenon can be sized by the number of available tables or the number of users who use Google Sheets of Excel:

- Web Tables: in 2008 were extracted 14.1 billion HTML tables and it was found out that 154 million are high-quality tables (1.1%);
- Web Tables: 233 million content tables in Common Crawl 2015 repository[2];
- Wikipedia Tables: the 2022 English snapshot of Wikipedia contains 2 803 424 tables from 21 149 260 articles [13];
- Spreadsheets: there are 750 million to 2 billion people in the world who use either Google Sheets or Excel[3].

Tables, datasets, databases, and infographics are non-linear content, and they are read differently from linear content (*i.e.*, linear text). For instance, a user may explore a table from top-to-bottom, from left-to-right or vice-versa. These reading behaviours are analysed in usability studies that aim to direct the user's reading patterns [14].

Although many studies consider *linear content*, studies that consider *tabular content* are absent. This work aims to suggest a new line of gaze-related research that can integrate insights about user behaviour and characteristics into tabular data summarisation algorithms to provide textual content that meets the user's information needs. The rest of the paper is organised as follows: Section 1.1 provides a comprehensive overview of state-of-the-art research on reading behaviours regarding linear text, highlighting the substantial lack of studies in the field of tabular data; Section 2 details the main motivations behind our research on tables; Section 3 introduces our proposed methodology to conduct such research, hinting at the possibility of leveraging eye-tracking techniques to acquire user data to integrate in Machine Learning (ML) algorithms; eventually, Section 4 delivers an outline of future directions towards a user-centric approach to tabular data summarisation.

1.1. State of the Art on Reading Behaviours

The analysis of on-screen reading behaviour on *linear texts* has been mainly conducted in *psychophysics* and *cognitive psychology* to understand how cognitive and lexical processes

[2] commoncrawl.org

[3] askwonder.com/research/number-google-sheets-users-worldwide-eoskdoxav

influence readers' eye movements. In particular, it is possible to infer which implicit mechanisms the user activates to retain and process the content through the patterns of fixations and saccades. In these contexts, it is essential to distinguish between two kinds of reading: i) in-depth reading, where the main objective is to understand the content of the text, ii) and cursory reading, which aims to capture the meaning of a text at a general level, also called skimming [1]. Considering in-depth reading, among the cognitive processes, *lexical access* is the most interesting for distinguishing users with or without background knowledge regarding the domain of the text under analysis. Lexical access means retrieving a word's meaning in memory. When a word is used often, its meaning is more available and accessible, instead, the meaning of an unknown word is difficult to retrieve. Lexical access[4] impacts the duration of fixations, which will be longer for unknown or unfamiliar words [1, 15]. These studies led to the realisation of language models, such as E-Z Reader [1, 11, 16], which have also found application in the field of NLG, in particular in the integration of the attention [7, 8, 6, 17, 9], namely the attribution of greater importance to certain parts of the text, similarly to what happens during reading, in ML models for the tasks of paraphrase generation and sentence compression [7].

Regarding *tables*, it is impossible to identify a systematic study of fixations patterns with similar objectives to those concerning the linear text. Tables arouse more interest in *User Experience (UX)* instead of *psychophysics* and *cognitive psychology*; [14] proposes a study investigating the influence of table design on the comprehension of the data it contains. In particular, it considers adding shading to highlight rows and columns selectively. [18], on the other hand, analyses the spatial dimension of tables to investigate whether cell size and spacing affect how users read, retain and process tabular data. The work presented in [5] is the first attempt to extend psycho-linguistic research on linear text to tables. The author uses eye-tracking techniques to detect the fixations patterns of users who have been asked to answer questions about the content of tables. The research focuses on tables that contain only numerical data. However, the experiment is conducted in a too-small setting, and the results are not generalisable to tables with non-numerical textual content. Thus, while there have currently been attempts to improve ML algorithms for linear text, systematic work moving in the same direction is still lacking for tables.

2. Motivations

Research related to the use of gaze-data has grown significantly in recent years, with effective applications in areas such as computer vision, decision learning, NLP, and NLG [7, 8, 19, 20]. The importance of the information that eye movements reveal about the user's cognitive state has been widely recognised in Artificial Intelligence (AI) studies, and their integration into ML algorithms has led to a marked improvement in performance [5].

In the context of NLP and NLG, the integration of user-data plays an important role [7, 20] as text comprehension is not only based on the characteristics of the text itself but also the characteristics of the users and the task they have to perform [8, 17, 19]. In fact, similar stimuli in different users generate different behaviours depending, for example, on the user's

[4] For a more detailed analysis of cognitive processes in reading behaviour, refer to [1]

[5] Please refer to [20] for an in-depth analysis of the state of the art on using gaze-data in AI.

information need, their interest in the task and the background knowledge they possess [21]. These user-data can be used to adapt attention in cognitive models [1, 11, 16] and to improve attentional mechanisms in pre-trained transformers such as BERT [6]. These integrations enable human-like results in tasks such as question answering, document retrieval and machine reading comprehension.

While the linear text has received the attention of various disciplines and its analysis has led to interesting results in NLG, table data have not attracted the same interest in the scientific community. The substantial absence of studies on table-reading patterns makes it very difficult to extend insights from the linear text approaches. For example, the user's exploration of tables is not strictly linked to the reading order. At the same time, while the content of a cell indeed follows the insights of linear text studies, the saccade that takes the eye from one cell to another may not necessarily follow the same rules. Similarly, UX studies show that the table's design may influence how it is read [14, 22]. The presence or absence of column headings and the uniformity of the type of data presented in the column are, for example, two essential characteristics that make the table reading different from that of linear text.

The studies in the literature that deal with supporting the user in analysing and understanding tables, as well as those that semantically interpret table data and provide summaries, are not user-centred. They do not consider the characteristics that determine reading behaviour and their influence on the user's eye movements. Like those on linear text, such studies refer to ML models that rely on the use of rules, templates and transformers [23] to do table summarisation; they could equally incorporate insights from user studies to improve the quality of results.

2.1. Use Case Scenario

As depicted, an enormous amount of data in table format is available online, and its use can be complicated for the user. For instance, suppose a web journalist needs to write an article on past editions of a film festival; the journalist will need to collect a modest amount of information to fulfil their information need. For example, they will need to gather information on the candidates, the winners in previous editions of the festival, the films in the competition and the jury. The journalist may start the research by using the sources they know best, such as the Internet Movie Database (IMDB) or may choose to consult the Wikipedia pages[6] related to the festival. In both cases, the journalist will most likely find the information in tabular format with numerical and textual data. The journalist will then have to process a large amount of data from which to extract statistics (*e.g.*, Leonardo DiCaprio received 7 Academy Award nominations but only won one) or maximum and minimum values (*e.g.*, the film that received the most Academy Award nominations was "All About Eva"[7]). In order to be able to derive this information independently, the user must scan each table in its entirety and then generate inferences from data that may be in areas of the table that are far apart from each other. This entails a considerable cognitive load. In this scenario, the user could benefit from the support of table summarisation tools that would help them to use the tabular data more effectively, for example by providing summaries that better contextualise, aggregate and summarise data in relation to their information needs and requirements.

[6] en.wikipedia.org/wiki/List_of_Academy_Award_records
[7] www.imdb.com/title/tt0042192/

3. Proposed Methodology

This paper proposes a new interdisciplinary line of research that considers psychophysics and cognitive psychology studies to analyse reading behaviour on tables to extend the results obtained from linear text to tabular data. The research will be conducted using a standard technique that is also common in the field of UX, eye-tracking. Eye-movement data is a highly efficient type of physiological data, as it allows a large amount of data to be collected at a low cost and in a short time [20]. Nowadays, an experiment based on eye-tracking can be conducted with dedicated high-resolution instruments, or with a wearable device or webcam [20, 24], with varying degrees of quality and accuracy. In general, however, the collection of real data on user behaviour is currently possible, as well as their integration into ML algorithms such as: i) additional information on the user's reading behaviour, and ii) filtering to exclude from the results items that did not capture the user's attention [20].

Furthermore, to understand which characteristics impact reading patterns and to what extent, it will be necessary to conduct investigations through questionnaires. In particular, we investigate the impacts of interest and prior domain knowledge (analysed as ease of lexical access) in reading tabular data. These characteristics are those that psychophysics and linguistics have recognised as relevant factors in determining eye movement on linear text [1]. The use of self-assessment tools, however, entails some critical issues, the main one being the self-assessment bias (described as the tendency to overestimate or underestimate one's characteristics and capabilities), which occurs below the level of consciousness and is a phenomenon that is difficult to capture, thus to mitigate. Particular attention should therefore be paid to the design of the instruments (*e.g.*, surveys, questionnaires) that will be used to collect data on user characteristics.

Once human-gaze data have been collected and the impact of user characteristics on their reading patterns assessed, it will be necessary to compare the results obtained on tabular data with those on linear text to identify possible points of convergence. If the behaviour on tabular data is comparable to that on text, then the results of research on the latter can be extended, thus integrated, into tabular data summarisation approaches. At this point, an in-depth analysis of machine learning techniques employed, for example, for table summarisation, will be conducted to find possible room for improvement toward a more human-centred conception of NLP and NLG. Lastly, it is pointed out that, with the integration of reading patterns within ML procedures, it is not intended to obtain better results in absolute terms but rather closer to the user's behaviour.

3.1. Goals and Research Questions

The main research questions that motivate this work are:

1. RQ1: How does the user behave when reading a non-linear text?
2. RQ2: What is the impact of the user's characteristics, particularly his interest and knowledge of the table domain, on his reading pattern?
3. RQ3: Can the results of linear text studies be extended to table data?

Hence, the objectives of this research can be summarised as follows:

- Analysing, from an interdisciplinary perspective, eye movements when reading tables and deriving reading patterns (RQ1);
- Identify features that influence reading behaviour (RQ2);
- Compare the results obtained with findings from human-gaze research on linear texts to identify commonalities and differences (RQ3);
- Analysing the ML algorithms used in NLP and NLG, with particular attention to table summarisation, and identifying room for improvement from a purely user-centred perspective (RQ3).

4. Conclusions and Discussions

This work proposes a new line of interdisciplinary research to bridge the gap between studies on linear text and tabular data reading behaviour, focusing on user characteristics. It is well known that the reader's attitudes, the pattern of their fixations and the frequency and direction of their saccades are closely correlated with their linguistic knowledge of the text's domain, thus with the ease with which the text is processed. Only one study on table can be identified in the literature [5]. However, it has substantial limitations concerning the type of data presented and the analysis of the findings. The results of human-gaze research on linear text applied to question answering, data and document retrieval and text summarisation suggest how the research described in this paper can be used in different contexts. Realistic use cases include, for instance, the generation of customised reports in a business context and the semi-automated writing of articles for the Web. Nevertheless, the actual integration of human-gaze data within NLG and NLP algorithms, i.e. those processes that are used to produce summaries (or descriptions) of tabular data, which we anticipate as the ultimate purpose of this research still needs in-depth analysis and testing.

References

[1] E. D. Reichle, A. Pollatsek, D. L. Fisher, K. Rayner, Toward a model of eye movement control in reading., Psychological review 105 (1998) 125.

[2] L. Frazier, K. Rayner, Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences, Cognitive Psychology 14 (1982) 178–210.

[3] K. Rayner, L. Frazier, Selection mechanisms in reading lexically ambiguous words, J Exp Psychol Learn Mem Cogn 15 (1989) 178–210.

[4] S. Leckner, Presentation factors affecting reading behaviour in readers of newspaper media: an eye-tracking perspective, Visual Communication 11 (2012) 163–184.

[5] A. Xu, Eye tracking study on identifying and analyzing user behavior-eye movements, eye fixation duration and patterns-when processing numeric table data in paper or PDF format, Master's thesis, School of Information and Library Science of the University of North Carolina at Chapel Hill, 2000.

[6] S. Dong, J. Goldstein, G. H. Yang, Gazby: Gaze-based bert model to incorporate human attention in neural information retrieval, in: Proceedings of the 2022 ACM SIGIR In-

ternational Conference on Theory of Information Retrieval, ICTIR '22, Association for Computing Machinery, New York, NY, USA, 2022, p. 182–192.

[7] E. Sood, S. Tannert, P. Mueller, A. Bulling, Improving natural language processing tasks with human gaze-guided neural attention, in: H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, H. Lin (Eds.), Advances in Neural Information Processing Systems, volume 33, Curran Associates, Inc., 2020, pp. 6327–6341.

[8] E. Sood, S. Tannert, D. Frassinelli, A. Bulling, N. T. Vu, Interpreting attention models with human visual attention in machine reading comprehension, in: Proceedings of the 24th Conference on Computational Natural Language Learning, Association for Computational Linguistics, Online, 2020, pp. 12–25.

[9] S. Wang, J. Zhang, C. Zong, Learning sentence representation with guidance of human attention, in: Proceedings of the 26th International Joint Conference on Artificial Intelligence, IJCAI'17, AAAI Press, 2017, p. 4137–4143.

[10] O. Eberle, S. Brandl, J. Pilot, A. Søgaard, Do transformer models show similar attention patterns to task-specific human gaze?, in: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 4295–4309.

[11] E. D. Reichle, T. Warren, K. McConnell, Using ez reader to model the effects of higher level language processing on eye movements during reading, Psychonomic bulletin & review 16 (2009) 1–21.

[12] S. Neumaier, J. Umbrich, J. X. Parreira, A. Polleres, Multi-level semantic labelling of numerical values, in: The Semantic Web – ISWC 2016, Springer International Publishing, Cham, 2016, pp. 428–445.

[13] M. Marzocchi, M. Cremaschi, R. Pozzi, R. Avogadro, M. Palmonari, Mammotab: a giant and comprehensive dataset for semantic table interpretation, Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab2022 (in press).

[14] M. Lee, T. Kent, C. M. Carswell, W. Seidelman, M. Sublette, Zebra-striping: Visual flow in grid-based graphic design, Proceedings of the Human Factors and Ergonomics Society Annual Meeting 58 (2014) 1318–1322.

[15] H. N. J. Ho, M.-J. Tsai, C.-Y. Wang, C.-C. Tsai, Prior knowledge and online inquiry-based science reading: Evidence from eye tracking, International journal of science and mathematics education 12 (2014) 525–554.

[16] E. D. Reichle, S. P. Liversedge, D. Drieghe, H. I. Blythe, H. S. Joseph, S. J. White, K. Rayner, Using e-z reader to examine the concurrent development of eye-movement control and reading skill, Developmental Review 33 (2013) 110–149.

[17] M. Hahn, F. Keller, Modeling human reading with neural attention, in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Austin, Texas, 2016, pp. 85–95.

[18] A. Shimojima, Y. Katagiri, An eye-tracking study of integrative spatial cognition over diagrammatic representations, in: International Conference on Spatial Cognition, Springer, 2010, pp. 262–278.

[19] J. Malmaud, R. Levy, Y. Berzak, Bridging information-seeking human gaze and machine reading comprehension, in: Proceedings of the 24th Conference on Computational Natural Language Learning, Association for Computational Linguistics, Online, 2020, pp. 142–152.

[20] R. Zhang, A. Saran, B. Liu, Y. Zhu, S. Guo, S. Niekum, D. Ballard, M. Hayhoe, Human gaze assisted artificial intelligence: A review, in: IJCAI: Proceedings of the Conference, volume 2020, NIH Public Access, 2020, p. 4951.
[21] M. Cremaschi, J. Amianto Barbato, A. Rula, M. Palmonari, R. Actis-Grosso, What really matters in a table? insights from a user study, in: IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, WI-IAT '22, in press.
[22] F. E. Gunawan, B. Soewito, N. Sekishita, Students' fixation on tables in powerpoint slides, in: 2019 IEEE International Conference on Engineering, Technology and Education (TALE), 2019, pp. 1–5.
[23] S. Jain, B. C. Wallace, Attention is not Explanation, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 3543–3556.
[24] A. Papoutsaki, Scalable webcam eye tracking by learning from user interactions, in: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA '15, Association for Computing Machinery, New York, NY, USA, 2015, p. 219–222.