

DISCOVERING NOVEL CUSTOMER FEATURES WITH RECURRENT NEURAL NETWORKS FOR PERSONALITY BASED FINANCIAL SERVICES

Anonymous authors

Paper under double-blind review

ABSTRACT

The micro-segmentation of customers in the finance sector is a non-trivial task and has been an atypical omission from recent scientific literature. Where traditional segmentation classifies customers based on coarse features such as demographics, micro-segmentation depicts more nuanced differences between individuals, bringing forth several advantages including the potential for improved personalization in financial services. AI and representation learning offer a unique opportunity to solve the problem of micro-segmentation. Although ubiquitous in many industries, the proliferation of AI in sensitive industries such as finance has become contingent on the imperatives of responsible AI. We had previously solved the micro-segmentation problem by extracting temporal features from the state space of a recurrent neural network (RNN). However, due to the inherent opacity of RNNs our solution lacked an explanation - one of the imperatives of responsible AI. In this study, we address this issue by extracting an explanation for and providing an interpretation of our temporal features. We investigate the state space of our RNN and through a linear regression model reconstruct the trajectories in the state space with high fidelity. We show that our linear regression coefficients have not only learned the rules used to create the RNN's output data but have also learned the relationships that were not directly evident in the raw data.

1 INTRODUCTION

Customer segmentation is an important field in banking and with customer bases growing banks are having to employ ever more advanced methods to maintain, if not improve, levels of personalization (Stefanel & Goyal, 2019). Customer segmentation has typically been achieved using demographics such as age, gender, location, etc. (Kalia, 2018). However, these features not only produce coarse segments, but also introduce the potential for discrimination, e.g., when using postal codes for credit rating (Barocas & Selbst, 2016). In contrast, micro-segmentation provides a more sophisticated, fine-grained classification that depicts nuanced differences between individuals, improves personalization, and promotes fairness. Despite these advantages and the fact that the need for such fine-grained segmentation has been highlighted (Krishnapuram & Mondal, 2017), the scientific community has been surprisingly quiet on the topic with only a few recent publications from, e.g., the health sector (Kuwayama et al., 2019; Nandapala et al., 2020) and none from the finance sector.

Artificial intelligence is fast becoming ubiquitous across multiple industries with representation learning an auspicious method for customer micro-segmentation (Maree & Omlin, 2021). Sensitive industries such as finance face legal and moral obligations towards the responsible implementation of AI (van der Burgt, 2019). The European Commission has published several guidelines surrounding responsible AI and scientific fundamentals have been consolidated in recent surveys on the topic (European-Commission, 2020; Arrieta et al., 2020). Explainability and interpretability are key elements in responsible AI (Goodman & Flaxman, 2017), which are generally not yet adequately addressed in applications of AI in finance (Cao, 2021). Our perspective on explainability in AI refers to a symbolic representation of a model, whereas interpretability refers to a human understanding of and reasoning about the functionality of the model. Explainability therefore neither guarantees nor implies interpretability. In this study, we address both the issues of explainability and interpretability. Our aim is to extract and facilitate the use of salient features in future financial ser-

vices; we have already shown the potential in predicting default rate and customer liquidity indices (Maree & Omlin, 2021). Our ultimate goal is the development of personalized financial services in which responsible customer micro-segmentation is key.

2 RELATED WORK

2.1 REPRESENTATION LEARNING USING RECURRENT NEURAL NETWORKS

In (Matz et al., 2016), the authors developed a model for predicting spending personality from aggregated financial transactions with the intent to investigate the causality between personality-aligned spending and happiness. They rated each of 59 spending categories according to its association with the Big-Five personality traits - extraversion, neuroticism, openness, conscientiousness, and agreeableness - which resulted in a set of 59×5 linear coefficients. We used these coefficients in a previous study to train a recurrent neural network (RNN) to predict customers' personality traits from their aggregated transactions (Maree & Omlin, 2021). We showed that the temporal features in the state space of the RNN had interesting properties: they formed smooth trajectories which formed hierarchical clusters along successive levels of dominance of the personality traits. The dominant personality trait is the one with the largest coefficient in the Big-Five model of personality traits (Maree & Omlin, 2021). We also showed that similarly salient features could not be extracted from the raw data otherwise. Spending patterns over time are either more consistent than transactions aggregated over a short time period, they may fluctuate, or they may change based on life circumstances. Modelling spending over time reinforces spending patterns and thus may lead to better features. Fluctuations or changes are also better represented by time series. The hierarchical clustering of the extracted features provided a means of micro-segmenting customers based on their financial behaviour. However, the responsible employment of this model demands an explanation and interpretation, which is what we address in this study. Other studies have employed representation learning using RNNs to encode spatial and temporal information contained in the two-dimensional trajectories of physical objects (Yao et al., 2017), as well as for a wide variety of applications in speech and video data (Li & Zhao, 2020).

2.2 EXPLAINING RECURRENT NEURAL NETWORKS

Finding symbolic representations of AI models is a key challenge in responsible AI. In (Udrescu & Tegmark, 2020) the authors developed a symbolic regression algorithm that successfully extracted physics equations from neural networks. They managed to extract all 100 of the equations from the well known Feynman Lectures on Physics and 90% of more complicated equations, an improvement from 15% using state of the art software. This was an important study because it not only proved that deep neural networks are capable of learning complicated equations and coefficients, but that it is possible to extract symbolic knowledge from such networks. The authors in (Ming et al., 2017) presented a visual method to explain RNNs used in natural language processing problems. They clustered the activations in the state space and used word clouds to visualize correlations between node activations and words in the input sentences. Similarly, the authors in (Omlin & Giles, 1996) applied clustering in the state space of RNNs, but here the authors showed that *symbolic* representations could be extracted as opposed to visual explanations. Studies such as these prove that deep neural networks are indeed not inexplicable black box systems, but could be a means of discovering symbolic representations of complex relationships in data.

3 EXPERIMENTAL METHODOLOGY

Our input data for training our RNN were aggregated financial transactions of approximately 26,000 customers, while the output data were their personality traits (Matz et al., 2016). The transactions were aggregated annually across 97 transaction classes, such as groceries, transport, leisure, etc., over a period of six years. The personality traits were based on the Big-Five personality model and were calculated using a set of 59×5 published coefficients linking transaction classes to personality traits (De Raad, 2000; Matz et al., 2016). Our RNN consisted of three long short-term memory (LSTM) nodes (Hochreiter & Schmidhuber, 1997). After training and during prediction, we inspected the activations of the three hidden nodes for each of the six time-steps; each customer was

represented by a trajectory with six data-points in a three-dimensional space, where each dimension represented the activation of one LSTM node. These trajectories were our extracted features which may be used for micro-segmentation of customers. To provide an explanation for the model, we trained a linear regression model - an inherently transparent class of models (Arrieta et al., 2020) - to replicate the trajectories from each customer’s aggregated spending distribution. Further inspection of the trajectories’ behaviour in relation to the associated customers’ fuzzy grades of membership in the personality traits provided an interpretation of the extracted features. The results are discussed in the following section.

4 RESULTS

In Figure 1, we show the features that we extracted from our RNN. Figure 1a illustrates the clustering behaviour of the trajectories in the state space. Our empirical observations led us to hypothesise the existence of attractors for each of the 5 personality traits. Figure 1b shows two trajectories for the same customer where the inputs to the RNN were aggregated over two different time periods: one year and six years. The fact that there is little difference between these two trajectories is significant; it demonstrates that the duration of the time window does not affect customer classification. This was not the case when clustering the raw personality data, where customers frequently moved between different clusters for different time periods due to variations in spending with changing life circumstances. Although we did observe significant course changes for some customers’ trajectories, the vast majority of customers remained in their assigned clusters for the six-year period. This stability in customer micro-segmentation is key for personalized financial services, as financial advice has to be consistent.

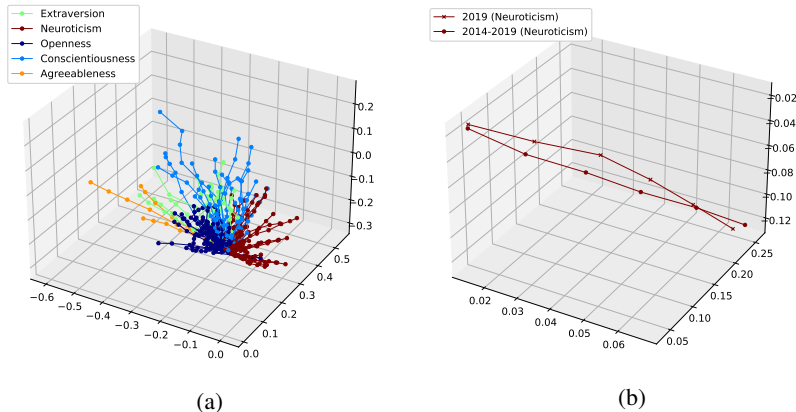


Figure 1: Trajectories in the 3-dimensional state space of a recurrent neural network trained to predict personality from aggregated transactions. While (a) shows the clustering of the trajectories of many customers according to the most dominant personality trait, (b) shows two trajectories for the same customer being identically classified over two different time periods: one year vs. six years.

To explain our model, we fit a linear regression model to reproduce the trajectories from the RNN’s input data: the 97 aggregated transaction classes, or spending patterns. From our observations in Figure 1, we hypothesized that the lengths of the trajectories were not as important as their directions. We therefore simplified the trajectory vectors and represented them by the two angles which fully describe their directions in three-dimensional space. These angles were the outputs of our linear regression model, which fit the data with a coefficient of determination (R^2) of 0.78 for an unseen test set, while a more complicated polynomial regression model managed an only slightly better 0.79. Other methods such as ridge regression and decision tree regression could not perform better. Our 97 transaction classes mostly overlapped with those of the 59×5 published coefficients and due to aggregations such as "health and fitness" being expanded to "health" and "fitness", there were 61×5 non-zero coefficients for calculating our customers’ personality traits. The linear re-

gression model had 69×2 non-zero¹ coefficients with a strong correlation to the original non-zero coefficients. Due to this and the relatively high coefficient of determination, we conclude that the linear regression model matched the RNN with high fidelity.

Within each of the clusters in Figure 1, we observed hierarchical sub-clusters along the second, third, and fourth most dominant personality traits. This hierarchical sub-clustering is important because it provides a means of micro-segmenting customers which was not present in the raw data and could neither be replicated using feed-forward neural networks nor auto-encoders. Using our linear regression model, we created a two-dimensional plot of trajectory angles (Figure 2). In this figure, we illustrate the hierarchical clustering behaviour that we observed for the trajectories from the RNN, where (a) shows the clustering along the customers’ most dominant personality trait and (b) through (d) show the hierarchy of sub-clusters within the parent clusters. These clusters, like the trajectory clusters, were consistent in time, i.e., the new features retained the desirable properties of the features from the state space of our RNN.

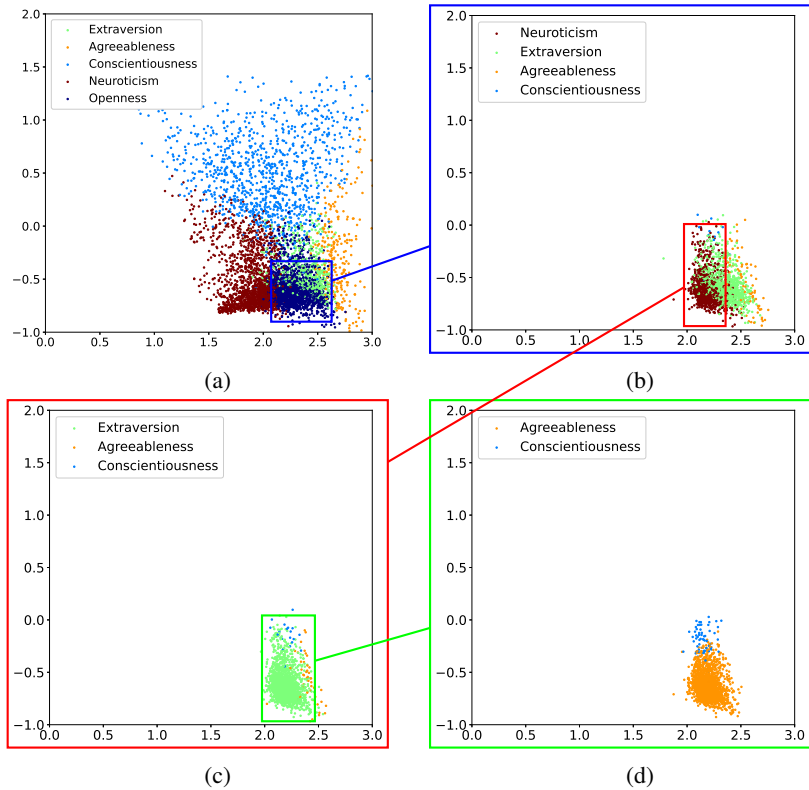


Figure 2: Hierarchical clustering for trajectory angles in 2-dimensional space. Each axis represents an angle (in radians) which describes the direction of the trajectories in 3-dimensional space and each data point represents a trajectory. We show all the levels of hierarchical clustering: (a) shows the highest level, while (b) through (d) show sub-clustering within each of the subsequent parent clusters.

The interpretation of the features in Figure 2 is that they mark the locations where the trajectories would penetrate the inside of a sphere in the state space of the RNN. We observed that the directions of the trajectories were consistent with the grades of the customers’ fuzzy memberships in each of the five personality traits, i.e., the output data of the RNN. The greater a customer’s fuzzy membership in the dominant personality trait, the quicker the trajectories converged towards the corresponding hypothesised attractor. The five attractors acted not only on the dominant personality trait, but also on succeeding lesser personality traits with succeeding lesser forces. We demonstrate this

¹Non-zero here refers to coefficients with values that are not insignificantly small compared to the mean value of all the coefficients.

in Figure 2 where the sub-clusters preserve the structure of their parent clusters: trajectories of lesser personality traits also converged to their respective attractors. Intuitively, people spend differently according to their dominant personality trait. Within a group of their peers, their lesser personality traits still differentiate them from each other. Thus the hierarchical clustering of trajectories and their labeling is the model interpretation.

Finally, Figure 3 shows the long-term (six years) and short-term (one year) trajectories of a single customer who changed their spending behaviour such that their dominant personality type changed in the last year. In this figure it is clear that, for the final year, both trajectories moved towards the same attractor (conscientiousness), with the neuroticism attractor no longer acting upon the long-term trajectory. We do not yet know the locations of the hypothesised attractors and though it is not required for an interpretation we nevertheless intend to locate them in future work by following methods described in, e.g., (Ceni et al., 2019; Maheswaranathan et al., 2019; Katz & Reggia, 2018).

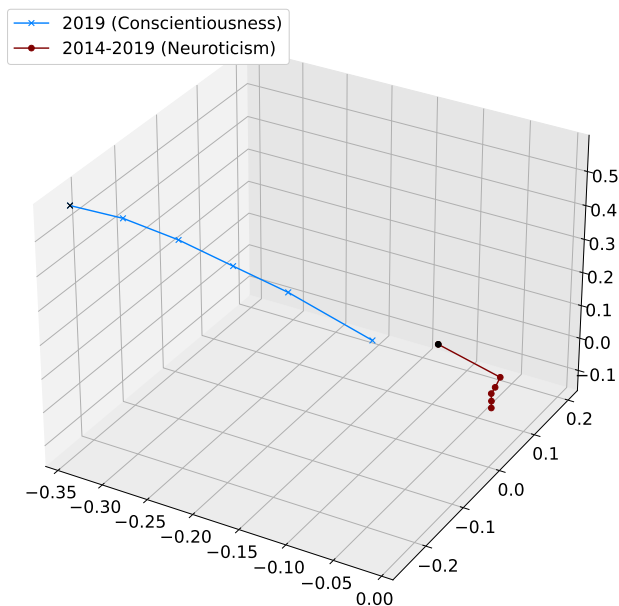


Figure 3: Two trajectories - long-term and short-term - for the same customer that seem to converge to a common attractor in the last year. In the long-term trajectory, the customer’s dominant personality trait is neuroticism as derived from annual aggregated spending patterns for the first five years. In the sixth year, the customer changed their spending pattern to conscientiousness, which is reflected in the directional change of the trajectory towards the conscientiousness attractor. The short-term trajectory corresponds to that same customers’ spending behaviour derived from bi-monthly aggregated spending for the last year.

5 CONCLUSIONS AND DIRECTIONS FOR FUTURE WORK

The financial sector is experiencing an increased demand in the level of personalization offered to its customers, which requires more nuanced segmentation techniques than the current offerings from traditional features such as demographics. Representation learning offers such an alternative technique for fine-grained segmentation, but it is plagued by the inherent opacity introduced by deep learning; explainability and interpretability are key in sensitive industries such as finance which must comply with regulations regarding the responsible use of AI. We proposed a solution by extracting temporal features from the state space of a RNN (Maree & Omlin, 2021). Our extracted features formed trajectories which were associated with the Big-Five personality traits. The dominant personality trajectories clustered and within each such cluster, we found a hierarchy of sub-clusters which corresponded to successively lesser personality traits. The clusters of feature trajectories corresponding to the dominant personalities provide a coarse customer segmentation,

while the hierarchy of trajectories associated with lesser personality traits offers the opportunity for micro-segmentation. We further observed the convergence of trajectories in the state space and we hypothesised the existence of attractors for each of the five personality traits. Finally, we provided an explanation for the trajectories through a high fidelity linear regression model which answers questions such as "why was Customer A classified in this way" by referring to their historic financial transactions. In future work, we intend to find the locations of the attractors in the state space which govern the directions of the customer trajectories, and to use our explainable features in the development of personal financial services such as personalized savings advice, advanced product recommendations, and wealth forecasters.

ACKNOWLEDGMENTS

This work was partially funded by The Norwegian Research Foundation, project number 311465. We are grateful for fruitful discussions with Joe Gladstone on the topic of personality traits and the determination of their corresponding coefficients.

REPRODUCIBILITY STATEMENT

We believe that we have made our experimental methodology sufficiently clear for the work to be reproduced. We have used standard models in Tensorflow to create the recurrent neural network and scikit-learn for the linear regression model. However, due to privacy regulations and the sensitive nature of our data, we will not be able to make our data available. We trust that this will be understandable.

REFERENCES

- Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, 58(1):82–115, 2020.
- Solon Barocas and Andrew D. Selbst. Big data’s disparate impact. *California Law Review*, 104(671):671–732, 2016.
- Longbing Cao. Ai in finance: Challenges, techniques and opportunities. *Banking & Insurance eJournal*, 2021.
- Andrea Ceni, Peter Ashwin, and Lorenzo Livi. Interpreting recurrent neural networks behaviour via excitable network attractors. *Cognitive Computation*, 12(2):330–356, 2019.
- B. De Raad. The big five personality factors: The psycholexical approach to personality. *Hogrefe & Huber Publishers*, 2000.
- European-Commission. On artificial intelligence - a european approach to excellence and trust (whitepaper). Technical report, European Commission, Brussels, Belgium, 2020.
- Bryce Goodman and Seth Flaxman. European union regulations on algorithmic decision-making and a right to explanation. *AI Magazine*, 38(3):50–57, 2017.
- Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.
- Prateek Kalia. Product category vs demographics: Comparison of past and future purchase intentions of e-shoppers. *International Journal of E-Adoption (IJEa)*, 10(2):20–37, 2018.
- Garrett E. Katz and James A. Reggia. Using directional fibers to locate fixed points of recurrent neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 29(8):3636–3646, 2018.
- Raghu Krishnapuram and Anirban Mondal. Upcoming research challenges in the financial services industry: a technology perspective. *IDRBT Journal of Banking Technology*, 1(1):66–84, 2017.

- Kenji Kuwayama, Hajime Miyaguchi, Yuko T. Iwata, Tatsuyuki Kanamori, Kenji Tsujikawa, Tadashi Yamamuro, Hiroki Segawa, and Hiroyuki Inoue. Strong evidence of drug-facilitated crimes by hair analysis using lc–ms/ms after micro-segmentation. *Forensic Toxicology*, 37(1): 480–487, 2019.
- S. Li and Handong Zhao. A survey on representation learning for user modeling. *International Joint Conferences on Artificial Intelligence Organization (IJCAI)*, pp. 4997–5003, 2020.
- Niru Maheswaranathan, Alex H. Williams, Matthew D. Golub, S. Ganguli, and David Sussillo. Reverse engineering recurrent networks for sentiment classification reveals line attractor dynamics. *Advances in neural information processing systems*, 32:15696–15705, 2019.
- Charl Maree and Christian W. Omlin. Clustering in recurrent neural networks for micro-segmentation using spending personality. *arXiv*, 2109.09425, 2021.
- Sandra Matz, Joe Gladstone, and David Stillwell. Money buys happiness when spending fits our personality. *Psychological science*, 27, 04 2016.
- Yao Ming, Shaozu Cao, Ruixiang Zhang, Zhen Li, Yuanzhe Chen, Yangqiu Song, and Huamin Qu. Understanding hidden memories of recurrent neural networks. *IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 13–24, 2017.
- E.Y.L Nandapala, K.P.N Jayasena, and R.M.K.T Rathnayaka. Behavior segmentation based micro-segmentation approach for health insurance industry. *2nd International Conference on Advancements in Computing (ICAC)*, 1(1):333–338, 2020.
- Christian W. Omlin and Lee Giles. Extraction of rules from discrete-time recurrent neural networks. *Neural Networks*, 9(1):41–53, 1996.
- Matteo Stefanel and Udayan Goyal. Artificial intelligence & financial services: Cutting through the noise. Technical report, APIS partners, London, England, 2019.
- Silviu-Marian Udrescu and Max Tegmark. Ai feynman: a physics-inspired method for symbolic regression. *arXiv*, 1905.11481, 2020.
- Joost van der Burgt. General principles for the use of artificial intelligence in the financial sector. Technical report, De Nederlandsche Bank, Amsterdam, The Netherlands, 2019.
- Di Yao, Chao Zhang, Zhihua Zhu, Jianhui Huang, and Jingping Bi. Trajectory clustering via deep representation learning. *International Joint Conference on Neural Networks (IJCNN)*, pp. 3880–3887, 2017.