University of York
Department of Computer Science

A Review of two Academic papers including recommendations for one

With a discussion of the key characteristics of qualitative and quantitative research

Word Count: 2992

May 1, 2023

Abstract

This document provides a report on two papers: 'Predicting Electric Energy Use of a Low Energy House: A Machine Learning Approach'[1] and 'Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data'[2]. These will be referred to as paper 1 and paper 2 respectively. The hypothesis, research questions, methodology and methods used in each paper are discussed. Recommendations are made for an alternative approach to the research conducted in paper 1 and the document concludes with a discussion of key characteristics of qualitative and quantitative data.

1 Evaluation

1.1 Hypothesis for Paper 1

The research aims to establish a relationship between electrical energy demand from a low energy home, and environmental factors. The study then seeks to make predictions about energy use.

There is no given hypothesis, but one might imagine an experimental hypothesis such as:

> It is possible to predict demand in electrical supply to a low energy home by considering environmental factors using a machine learning approach.

1.2 Hypothesis for Paper 2

The research is aimed at using Big Data Analytics and visual representations to forecast trends in crime in three major U.S. cities. Again, no hypothesis is given, but a suitable one might be:

> It is possible to use Big Data Analytics and visualisations to forecast trends in crime.

1.3 Review of paper 1

Regarding design, this paper has room for improvement. The reader is unaware of the authors' philosophical worldview, methods are not in appropriate places and there are poorly placed tables.

Philosophically, it is possible the authors have a postpositivist worldview. Creswell and Creswell confirm that postpositivists believe that causes affect outcomes[3]. The causes in this research could be identified as environmental with the impact being on demand for electricity.

The research also attempts to predict electrical demand. Therefore, it could be argued that it is a pragmatic worldview since the work is concerned with the application of theories in the real world[4].

In terms of methods, the research combines environmental data from a secondary source[1] [5] with energy use recorded from smart meters. Support Vector Regression (which uses machine learning to create a line of best fit within a specific range to predict discrete values[6]) is then applied to identify if there is a link between these variables[1]. An artificial neural network with back-propagation is also used[1]. This type of system aims to learn from patterns identified in non-linear variables and make predictions[7].

The methodology here falls under the category of correlational studies. As Creswell explains, when we are relating variables, we are conducting correlational research[8]. Also, since the paper deals exclusively with the relationship between variables, we can conclude that it is quantitative research[3].

Regarding structure, a methodology is the philosophy behind, and the reasons why research has been conducted in a particular way, whilst a method should describe the actual procedures used in the research[3]. This paper does not follow these definitions with methods described in various locations and no methodology given (despite the presence of a methodology section). There are also misplaced tables. For example, the results of the principal component analysis (which is used to help simplify the variety of data collected in the study to make it easier to interpret[9]) should appear in the results section, not in the methodology section.

The methods work well though in confirming the hypothesis inferred above.

1.4 Review of paper 2

Paper 2 uses publicly available crime data from three major U.S cities [10] [11] [12] to create visualisations. It also uses models to confirm predictions about crime rates at certain times.

Overall, the structure of paper 2 is more organised. It is good to see a separate section for the literature review. Boote and Beile explain that a literature review is the foundation of good research[13]. However, again there is no discussion of philosophy or methodology. A plan for research should involve decisions that include philosophical assumptions, research and method designs[3]. These are absent in this paper.

There is also a problem with one of the citations. In explaining the prophet model, the authors cite a source[14] that does not appear to mention the prophet model at all. The description is actually similar to text found when performing a google search on the terms "Prophet model time series". The first result is from a website that has an almost identical definition for the prophet model[15]. The tutorial shown there was created after paper 2, but none of the references are shared. So where did the explanation of a prophet model actually come from? This casts doubt on some of the research.

Two other prediction models are also used (neural network model and the LSTM model) to confirm visualisations and hence, recommendations in terms of police deployment to combat crime.

Philosophically, the study appears to stem from a postpositivist worldview. Again, postpositivists believe that causes affect outcomes[3]. The causes here are times of day/seasons with the outcome being crime rate.

For methodology, this research would also fall under the category of a correlational study. The research is establishing whether there is a link between variables. This a defining factor of a correlational study[8] and again, we can conclude this is quantitative research[3].

Missing data has been dealt with by imputing random variables sampled from non-missing values. Viswanathan and Viswanathan confirm this as a valid approach which avoids having to disregard other useful information from those specific cases[16].

Overall, the models used have worked well and visualisations have been created that are easy to follow. Agreeing with the recommendation of increasing police presence during certain periods is obvious from these results.

2 Recommendations

2.1 Hypothesis analysis

The structure of research stems from the hypothesis and a hypothesis should be testable[17]. Therefore, the inferred hypothesis for this paper should be modified to explain to what degree it is 'possible' to predict the demand for electricity.

If we are re-writing the hypothesis though, the whole structure of the research can be changed. Zobel explains that hypotheses are regularly refined, both after initial testing and for scientific progress in general[17].

In pondering a new hypothesis, one might consider the following questions which have not been addressed by the existing research:

1. What environmental factors affect the use of which appliances?

2. Are consumers able to modify their energy usage times to mitigate against environmental factors?

3. Could this research help consumers save money?

4. Does the affluence of an area affect how the demand for electricity changes with environmental factors?

Conducting initial interviews with consumers would transform the study to mixed methods research - specifically an Exploratory Sequential Mixed Methods Design where the results from interviews determine the quantitative parameters to be investigated[3]. Might consumers be interested in saving money through possible subsidies from electricity companies similar to 'Time of Use' systems[18] and usage cutting schemes[19] in the U.K.? Interviews could confirm this.

In transforming the study into mixed methods research, it is important to ensure both qualitative research questions and a quantitative hypothesis are present[3].

Qualitative Research Questions:

1. How can the effect of an area's affluence on a lower cost electrical supply be explored?

2. How can the process of consumers using appliances at different times to mitigate against higher energy costs caused by environmental factors be explored?

Quantitative null Hypothesis:

There is no decrease in cost for consumers in changing their habits when a machine learning approach is used to predict and inform them about demand for electrical supply based on environmental factors.

2.2 Recommended starting point

After conducting an updated literature review, the qualitative research questions could be used to develop sub-questions that narrow the focus of the study whilst leaving the question open[3]. Miles and Huberman advise no more than a total of twelve research questions and sub-questions[20].

Tashakkori and Creswell also suggest that a specific mixed methods question be created that encompasses both the qualitative and the quantitative aspects of the research[21]. Therefore, in addition to the above, a mixed methods question could be:

How can machine learning predictions of electrical supply based on environmental factors explain the energy habits of consumers?

From here, we might begin to ask how consumers' habits could be modified.

2.3 Justification

A mixed methods question should ask about how (in this case) the qualitative research will feed into the quantitative research[3]. Tashakkori and Creswell explain that in a sequential design, a mixed methods question should seek an explanation as to how the qualitative

findings explain the qualitative results[21]. This is exactly what the above question hopes to achieve.

3 Proposed research design

The design for writing the paper loosely follows the model set out by Zobel[17], and is described within each section below.

Abstract

An abstract's purpose is to allow a reader to decide if this paper is of significance to them[17], so the abstract would outline the paper.

Introduction

The introduction provides the reader with background information about the research described[22] and should include a purpose statement. This is because a purpose statement is the "...central controlling idea in a study".[3]. The purpose statement for the original research is in the abstract. It would be adjusted to include the input of consumers from the qualitative phase of the research and be placed at the end of the introduction[3].

The introduction would also share the philosophical worldview. A description of this worldview would follow along with its influence on the research[3]. This would be a pragmatic worldview as the research would be looking at consequences (changing of habits of consumers) of actions (giving consumers knowledge of environmental factors). It is also concerned with the real world. These are key attributes of a pragmatic worldview[3].

Literature review

The original paper discusses previous research carried out in this field[1] and this would be a good place to start for a literature review. Given that the paper is now 4 years old though, there may be some more recent sources worthy of consideration.

Methodology

The proposed methodology is an Exploratory Sequential Mixed Methods Design. As Creswell and Creswell explain, it is important to provide the reader with a brief explanation of what this process entails along with its history which is how this section would commence[3].

A mixed methods approach hopes to reveal additional information beyond what would be revealed by a quantitative study alone[3]. Specifically, an Exploratory sequential mixed methods approach hopes to explore exactly which variables need to be investigated in the quantitative phase of the study[23].

However, as Creswell and Creswell go on to explain, there are challenges with this approach and the researcher will need to be careful when selecting which aspects of the qualitative study will feed into the quantitative phase[3]. Particular care should also be taken in choosing the samples for each phase[3].

The initial phase of the design would be a qualitative process. The results of this process would determine exactly which factors to focus on for the quantitative phase of the research. As Creswell and Creswell explain, this second phase of the design is used to tailor the third quantitative phase so that the needs of the individuals in the study can be better met[3].

The third phase would be the quantitative process.

Methods

A detailed analysis of the sources in the literature review will inform the questions that are to be asked of the consumers in interviews.

There are several factors that will influence how the sample for the qualitative phase will be determined. In considering the affluence of an area, data similar to that provided by the Office for National Statistics in the U.K. could be used[24]. This would help identify key regions for comparison. Care needs to be taken here though. It might be tempting to pick the most affluent area and the least affluent. However, since we are considering environmental factors, the regions should have similar environments to ensure a fair comparison.

Once the regions have been chosen, careful consideration of the specific housing type would need to be considered. According to the Office for National Statistics, in the U.K., the age of a property is the biggest factor affecting its energy efficiency[25]. The type of home would need to be considered as well. For example, semi-detached houses are more energy efficient than detached houses[26]. If possible, consumers from similar style and age of buildings should be chosen for the different regions in the sample.

For sample size, a power analysis should be conducted to ensure a large enough response from participants is acquired to make the results meaningful[3].

Once the participants have been identified, they should be interviewed. The questions to be asked would need to be written after the literature review was completed but would broadly seek to establish what energy saving mechanisms were already in place at the property and what the consumer's electrical habits were.

The raw data from interviews would need to be organised, read and coded in order to create groupings [3]. Then, the results from the interviews would feed into the quantitative phase of the research. This quantitative phase would follow similar methods from the original paper. However, following the results of the interviews, adjustments would be made. For example, it is envisaged that not all the environmental factors listed in the data source[5] would need to be considered. As part of data preparation, these could be removed. The different potential savings for each type of consumer could also be considered.

Following this, the machine learning algorithms would be applied to create the predictions necessary. Consumers Could then be told how they could modify their behaviour.

Results

Results from the qualitative research would be collated and used to explain the impact on the quantitative phase of the research. The grouped qualitative results would be displayed in charts and tables. From there, the results from the qualitative phase of the research could be shared, again, using charts and tables. An explanation of how the qualitative phase of the research has improved the results would be given[3].

Conclusions

This section would draw together the important findings from the results, state any limitations and discuss areas for future research[17].

4 Key characteristics of qualitative and quantitative research

4.1 Qualitative research

Qualitative research relies on data that is text and/or images[3]. Its ultimate goal is learning and should begin with asking questions[27]. These questions are essential to creating good research and they create its structure[17].

Qualitative research could include studying documents - the advantage of this is that it is typically quicker than collecting one's own data. However, the researcher would not gain the first-hand experience from working in the field which is the more usual method of conducting qualitative research[3]. This could include observing and/or interviewing participants and there would typically be multiple sources[3]. On the other hand, the process of sorting and analysing the data can be time-consuming.

Hatch is careful to point out that providing models for data analysis "...goes against the open-ended nature of qualitative work"[28], but goes on to present five:

- Typological
- Inductive
- Interpretive
- Political
- Polyvocal

Each of these analysis models seek to group the data using different approaches. However, they are given as examples by Hatch and he explains that they are to guide thinking about analysis and are flexible[28].

Finally, the write-up should be tailored accounting for one's audeience[27]. However, as Hatch explains, the writing should be very much part of the research and not something that happens afterwards[28].

4.2 Quantitative research

According to Creswell and Creswell, the two main types of quantitative research are survey design and experimental design and they stem from a postpositivist worldview[3]. Watson confirms that these are the main types of quantitative research and goes on to explain that it must be possible to measure the subject being studied in some way[29]. Creswell and Creswell further explain that the advantage of a survey design is that it can be used to look at key variables in a population whilst the advantage of an experimental design is that it will look at connections between variables[3].

A disadvantage of a survey design is that data would need to be collated before it can be analysed which can be time-consuming. Disadvantages of an experimental design is that it might be affected by conditions and assumptions often need to be made.

Bryman sets out the principal components of qualitative research to include theory, hypothesis, selection of participants, survey or experiment, data collection, data analysis and findings[30]. Creswell and Creswell add that qualitative research may use research questions instead of a hypothesis but should not use both[3]. They further explain that quantitative research questions ask about relationships between variables whilst hypotheses make predictions about what can be expected as outcomes[3].

Independent and dependent variables should be named and an explanation given about how they are expected to relate to each other[3].

Fowler describes a variety of methods for sampling including area probability and random-digit-dialling[31]. Whichever method is chosen however, careful consideration will need to be given to the size of the sample. Many scientific journals expect a power analysis to be carried out if the results are to be judged as significant[3]. There should also be ethical considerations about who is selected and no individual should suffer as a direct result of taking part[31].

Details of the instruments used to conduct the surveys or experiments would need to be provided[3]. The reliability of the instruments is important and one of the main ways of proving reliability is through a test-retest process[30]. This process sees the instrument used with a test group and then used again with the same test group with broadly the same results expected[30].

Creswell and Creswell give a 6-step guide to explain how data can be analysed [3]. It includes giving information about numbers in the sample, a discussion of how to deal with response bias, giving information about central tendency and spread for all variables, identifying computer programs used to test statistics and finally presenting the findings in charts and tables[3].

The final stage of a quantitative design would be a discussion on findings, linking to other literature in the field[3].

References

[1] R. I. Rasel, N. Sultana, S. Akther, and A. Haroon, "Predicting electric energy use of a low energy house: A machine learning approach," in *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 2019, pp. 1–6. DOI: 10.1109/ECACE.2019.8679479.

[2] M. Feng, J. Zheng, J. Ren, A. Hussain, Y. Li X.and Xi, and Q. Liu, "Big data analytics and mining for effective visualization and trends forecasting of crime data," *IEEE Access*, vol. 7, no. 106111, 2019. DOI: 10.1109/access.2019.2930410.

[3] J. D. Creswell and J. W. Cresswell, *Research Design (5th ed.)* Thousand Oaks: SAGE Publications US, 2017.

[4] M. Q. Patton, *Qualitative research & evaluation methods: integrating theory and practice*. SAGE PUBLICATIONS, 2015.

[5] L. M. Candanedo, V. Feldheim, and D. Deramaix, *Appliances energy prediction data set*, https://archive.ics.uci.edu/ml/datasets/Appliances+energy+prediction, Accessed: 2023-04-23, 2017.

[6] M. Awad and R. Khanna, "Support vector regression," in *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*. Berkeley, CA: Apress, 2015, pp. 67–80, ISBN: 978-1-4302-5990-9. DOI: 10.1007/978-1-4302-5990-9_4. [Online]. Available: https://doi.org/10.1007/978-1-4302-5990-9_4.

[7] A. Goh, "Back-propagation neural networks for modeling complex systems," *Artificial Intelligence in Engineering*, vol. 9, no. 3, pp. 143–151, 1995, ISSN: 0954-1810. DOI: https://doi.org/10.1016/0954-1810(94)00011-S. [Online]. Available: https://www.sciencedirect.com/science/article/pii/095418109400011S.

[8] J. W. Creswell, *Educational research: Planning, conducting, and evaluating quantitative and qualitative research (Global ed.)* Pearson Education Limited, 2020.

[9] I. T. Jolliffe and J. Cadima, "Principal component analysis: A review and recent developments," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, 2016.

[10] City and County of San Francisco, *Police department incident reports: Historical 2003 to may 2018*, Accessed: 2023-04-26. [Online]. Available: https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-Historical-2003/tmnf-yvry.

[11] Chicago Police Department, *Crimes - 2001 to present*, Accessed: = 2023-04-26. [Online]. Available: `https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2`.

[12] Philadelphia Police Department, *Crime incidents*, Accessed: 2023-04-26, 2016. [Online]. Available: `https://opendataphilly.org/datasets/crime-incidents/`.

[13] D. N. Boote and P. Beile, "Scholars before researchers: On the centrality of the dissertation literature review in research preparation," *Educational Researcher*, vol. 34, no. 6, pp. 3–15, 2005. DOI: 10.3102/0013189X034006003. eprint: `https://doi.org/10.3102/0013189X034006003`. [Online]. Available: `https://doi.org/10.3102/0013189X034006003`.

[14] A. AGRESTI, *AN INTRODUCTION TO CATEGORICAL DATA ANALYSIS*, D. J. Balding, N. A. C. Cressie, G. M. Fitzmaurice, G. H. Givens, H. Goldstein, G. Molenberghs, D. W. Scott, A. F. M. Smith, and R. S. Tsay, Eds. John Wiley and Sons, Inc, 2019.

[15] P. BANERJEE, *Tutorial: Time series forecasting with prophet*, Accessed: 2023-04-25, 2021. [Online]. Available: `https://www.kaggle.com/code/prashant111/tutorial-time-series-forecasting-with-prophet`.

[16] V. Viswanathan and S. Viswanathan, *R Data Analysis Cookbook*. Packt Publishing Ltd., 2015.

[17] J. Zobel, *Writing for Computer Science, 3 ed.* London: Springer Nature, 2015.

[18] Energy Saving Trust, *Time of use tariffs: All you need to know*, Accessed: 2023-04-26, 2022. [Online]. Available: `https://energysavingtrust.org.uk/time-use-tariffs-all-you-need-know/`.

[19] P. Lekarski, *Energy suppliers are paying households up to £100 to cut electricity use at peak times this winter – here's how it works and how to sign up*, Accessed: 2023-04-26, 2023. [Online]. Available: `https://www.moneysavingexpert.com/news/2022/10/octopus-ovo-energy-pay-to-cut-electricity-use/#:~:text=To%20earn%20the%20reward%2C%20you,the%20day%2C%20according%20to%20ovo.`.

[20] M. B. Miles and A. M. Huberman, *Qualitative Data Analysis, 2nd edition*. Sage Publications, 1994.

[21] A. Tashakkori and J. W. Creswell, "Editorial: Exploring the nature of research questions in mixed methods research," *Journal of Mixed Methods Research*, vol. 1, no. 3, pp. 207–211, 2007. DOI: 10.1177/1558689807302814. eprint: `https://doi.org/10.1177/1558689807302814`. [Online]. Available: `https://doi.org/10.1177/1558689807302814`.

[22] A. M. Wilkinson, *The Scientist's Handbook for Writing Papers and Dissertations*, M. P. Rottino, Ed. Prentice-Hall, Inc., 1991.

[23] W. A. Edmonds and T. D. Kennedy, *An Applied Guide to Research Designs: Quantitative, Qualitative, and Mixed Methods, Second ed.* Thousand Oaks, California: SAGE Publications, Inc, [Online]. Available: https://methods.sagepub.com/book/an-applied-guide-to-research-designs-2e, 2017.

[24] Office for National Statistics, *What are the regional differences in income and productivity?* Accessed: 2023-04-28, 2021. [Online]. Available: `https://www.ons.gov.uk/visualisations/dvc1370/`.

[25] Office for National Statistics, *Age of the property is the biggest single factor in energy efficiency of homes*, Accessed: 2023-04-28, 2022. [Online]. Available: `https://www.ons.gov.uk/peoplepopulationandcommunity/housing/articles/ageofthepropertyisthebiggestsinglefactorinenergyefficiencyofhomes/2021-11-01`.

[26] J. Gresham, *Are semi-detached houses warmer? (and cheaper to run?)* Accessed: 2023-04-28, 2021. [Online]. Available: `https://www.buildfanatic.com/semi-detached-warmer/#:~:text=Semi%2Ddetached%20houses%20are%20warmer,to%20maintain%20a%20comfortable%20temperature.`.

[27] G. B. Rossman and S. F. Rallis, *An Introduction to Qualitative Research: Learning in the Field, Fourth ed.* Thousand Oaks, California, [Online]. Available: https://methods.sagepub.com/book/an-introduction-to-qualitative-research-4e., 2017.

[28] H. J.A., *Doing Qualitative Research in Education Settings*. Albany: State University of New York Press, 2002.

[29] R. Watson, "Quantitative research," *Nursing standard : official newspaper of the Royal College of Nursing, 29(31)*, vol. 31, no. 29, 2015. DOI: `doi.org/10.7748/ns.29.31.44.e8681`.

[30] A. Bryman, *Quantitative data analysis with SPSS 12 and 13 : a guide for social scientists / Alan Bryman and Duncan Cramer.* eng. London: Routledge, 2005, ISBN: 0415340802.

[31] J. Floyd J. Fowler, *Survey Research Methods Fifth Edition*. SAGE Publications, Inc., 2014.