Iyer, S. et al. (Eds.) (2022). Proceedings of the 30th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Comparing Short-Term and Long-Term Online Courses Using the Kano Model and Neural Network Language Models

Daniel Moritz MARUTSCHKE[a*] **& Yugo HAYASHI**[b]
[a]*Ritsumeikan University, College of Global Liberal Arts, Osaka, Japan*
[b]*Ritsumeikan University, Department of Comprehensive Psychology, Osaka, Japan*
[*]moritz@fc.ritsumei.ac.jp

Abstract: Evaluating rich data from online courses poses challenges, even with established methodologies. A mixed methods approach can help in addressing research questions. One of the difficulties is to understand users' expectations before taking a course and their consumption experience after completing a course. The focus of this study is on the difference in short-term courses, which can be taken within a day, and long-term courses, which span several weeks or months. This paper compares two online courses for undergraduate students in Japan. The courses were hosted on the same system, one that was complete within 90 minutes (introduction to photography) and one that was conducted for a semester totaling 15 weeks (introduction to programming). Students were surveyed on the same 12 features related to online course satisfaction before and after each course. Textual comments were also gathered. The Kano model from customer satisfaction research was used to perform an ex-ante and ex-post comparative analysis for the 12 features of both short-term and long-term courses. A simple neural network was trained on freeform comments for both courses to create language models (Word2Vec) and compare the findings with the Kano model results.

Keywords: Online course, higher education, e-learning, kano model, machine learning, word2vec, language model

1. Introduction

Online courses and e-learning have gained popularity not least due to the current Covid-19 pandemic. While online-supported and digital education has uncovered shortcomings in many areas throughout 2020 and 2021, a broader challenge in research is the handling of large data collected from online courses. Most online courses that are in form of Massive Open Online Courses (MOOC) and students can enroll in a course and take lessons at their own pace. Famous universities offer courses for free with an option to purchase a certificate of completion, or companies provide fee-charging content for specific skills. Courses can last a whole semester or can be finished in one session.

As learning outcome can differ, depending student expectation and consumption experience and course duration, this research looked at two different courses provided on the same platform, one short-term introduction to photography class and one long-term introduction to programming class. Undergraduate students in Japan from different faculties (psychology for the short-term and liberal arts for the long-term) were surveyed before and after taking the online course. The expectations and consumption experience of both courses are evaluated by the Kano model from customer satisfaction research (Kano et al., 1984) and by creating an AI language model from open-ended questions.

The rest of the paper is structured as follows. Section 2 lists previous works regarding the Kano model in e-learning and language models using Word2Vec. Section 3 describes the research method using the Kano model and Word2Vec. Results and discussion is covered in Section 4. Limitations of this study are considered in Section 5. The paper concludes in Section 6.

2. Previous Work

2.1. Kano Model in E-Learning

While the Kano model is mostly used in business or customer satisfaction research, other previous research in e-learning include learning motivation and exploring multiple hypothesis that tie into student satisfaction (Sun et al., 2008). The Kano method was used to poll students on several e-learning factors in a previous study by Dominici and Palumbo (Dominici and Palumbo, 2013) to construct a theoretical framework of online course design (ex-ante questionnaire without implementation and testing the ex-post consumption experience). Wang et al. investigated a hybrid online/face-to-face course using blended learning (Wang et al., 2014).

2.2. Uses of Language Models Trained by Word2Vec

Large language models have gained popularity, but also garnered discussions about implicit biases in large models (Nissim et al., 2020; Zou and Schiebinger, 2018). This should be taken into account when using blackbox approaches, such as neural network training. Altszyler et al. investigate small text corpora for word associations and find that Latent Semantic Analysis (LSA) can provide superior results for low-frequency words compared with Word2Vec (Altszyler et al., 2017). They apply their algorithms to dream journals. A large corpus of abstracts and papers related to biomedical research was analyzed by Zhu et al. (Zhu et al., 2017). They conclude that increasingly large datasets can result in more relations in biomedical terms, but do not guarantee better precision. This ties into the strength of word associations, but limited semantic properties of Word2Vec. Kazama et al. propose an ingredient substitution method based on the vector arithmetic allowed by the Word2Vec algorithm (Kazama et al., 2018). One of their proposals is to allows a change in cuisine style from one cultural background to another, but also to adjust meals to a persons's individual dietary preferences.

Other limitations and strengths of Word2Vec were investigated by Di Gennaro et al. (Di Gennaro et al., 2021). Overfitting was mostly ruled out for the training process of large datasets and they list analogy models as a strength. Syntactic relationships were not able to be achieved by Word2Vec.

3. Research Method

3.1. Online Course and Questionnaire Design

The online course used for this research was built with WordPress, one of the most popular blogging and content management systems (CMS), and the premium plugin LearnDash (WordPress, 2022; Learndash, 2022). WordPress has also been used for educational purposes (Scott, 2012; Quesenberry et al., 2014). The course design followed the same format in both courses, with same functionality, user-interface, and lesson-topic-quiz structure.

To perform an ex-ante and ex-post analysis, students were surveyed before and after taking the online course. The questionnaires were designed with three types of questions in mind—to gather statistical information about age and English proficiency, functional and dysfunctional questions for online course features to assign the Kano categories, and freeform textual comments. The functional and dysfunctional questions were asked in the same form before and after the online course was taken. 12 online course features were surveyed, detailed in Section 4. Three additional questions about what students liked, disliked, and wished for were asked ex-post. Questions were selected based on previous research and course specific insights (Selvi, 2010; Dominici and Palumbo, 2013).

3.2. Kano Model

The Kano model was first introduced by Kano et al in 1984 and is a tool to conduct customer satisfaction research (Kano et al., 1984). At the core is a questionnaire that asks (prospective) customers to answer functional and dysfunctional questions regarding a product's feature. The *functional* question asks customers how they feel if a feature is present (implemented) and the *dysfunctional* question asks customers how they feel if a feature is missing. Combining these answers results in one of six Kano categories. Participants have to rate a feature in five levels for both these questions—highly satisfied, as expected, neutral, can live with it, and highly dissatisfied.

The list below details the categories that the Kano model produces. As the original publication describing the categories is written in Japanese, the terminology has been adjusted. *Quality* has been replaced by *requirement* to reflect terminology from software engineering and the term *must-be* was replaced by *basic*. For this research, the term *customer* is replaced by *user*, *student*, or *participant*.
B (basic requirement), O (one-dimensional requirement), I (indifferent requirement), Q (questionable requirement), A (attractive requirement), R (reverse requirement).

To illustrate the meaning of the Kano model categories covered in the case of an e-learning platform or online course, the following examples are explained for each requirement. A working website, consistent URLs, and mobile-friendly content can be considered basic requirements (B). They do not increase the satisfaction if properly included, but would reduce the satisfaction if missing or implemented poorly. An attractive feature (A) could be a live note-taking function or engaging implementation of gamification. Over time, these features can become one-dimensional (O) or basic requirements (B). One-dimensional requirements (O) increase the user's satisfaction proportional to the degree of implementation. Factors that result in a good User Experience (UX) increase the satisfaction proportionally to the degree of implementation, as well as in the opposite direction. The Kano model includes states that cannot be captured with other methods in this field. Reverse requirements (R) have the opposite effect of one-dimensional requirements (O) or in some cases of attractive requirements (A), i.e., they decrease (proportionally) the satisfaction when implemented. An example for web-based applications would be pop-up dialogue boxes. This is one of the reasons for choosing the Kano over other commonly used ones, such as SERVQUAL, E-Learning Satisfaction (ELS), or e-SERVQUAL (Wang et al., 2007; Parasuraman et al., 1988, 2005). Dominici and Palumbo also detail two more reasons for this methodology: the first is crucial to this research and allows ex-ante and ex-post analysis, the second is that this model does not assume a linear relationship between the service performance and satisfaction (Dominici and Palumbo, 2013; Chaudha et al., 2011).

3.3. Language Model with Word2Vec

Machine learning algorithms to create language models can help analyze unstructured textual data by converting words into numerical vector values called word embeddings. By training Word2Vec on a text corpus, words can be compared in vector space. The algorithm allows to create a model that returns a numerically close word, given a context, or return a numerically close context, given a word. The former is called continuous bag-of-words (CBOW) and the latter a skip-gram model and used in this research. As words are represented as multidimensional vectors, the algorithms also allows to perform vector arithmetic (Gittens et al., 2017). Word embeddings can be added (context positive) to each other or subtracted from each other (context negative), which numerically shifts the predicted context or target word in the word-vector space. Stanza, a natural language processing (NLP) toolkit developed by the Stanford NLP group was used to tokenize and lemmatize the words (Qi et al., 2020). Stop-words were removed from the sentences and all words were converted into lowercase.

Word2Vec was trained on the dataset with a vector size of 60 to avoid overfitting (observed with smaller datasets and default vector size of 100). Ex-ante comments were primed with the word need, ex-post comments with like, dislike, and wish, respectively, to allow targeted context and clearer context shifts via word-vector arithmetic. For higher accuracy, negative sampling is used on contexts (Aggarwal, 2018). The model was trained using the Word2Vec implementation of Gensim for Python (Rehurek and Sojka, 2010) with the following parameters: Skip-Gram mode, learning rate $\alpha = 0.001$, window size 5 for short-term and 7 for long-term, vector size of 60, and minimum word-count of 1.

4. Results and Discussion

In the long-term online course, 36 undergraduate students participated, 18 male, 18 female, and 0 non-binary. The average age was 19.00 ($\sigma = 1.18$), ranging from 18 to 23. The course was conducted over a period of 15 weeks. Students could access new material on the online course each week after a preset date. For the short-term online course, 16 undergraduate students were surveyed, 5 male, 11 female, and 0 non-binary. The average age was 19.82 (σ = 1.70), ranging from 18 to 24. The experiment, including introduction, questionnaire conduct, and taking the online course was completed in 90 minutes. The following 12 online course features were surveyed before and after the online course (abbreviations corresponding to Figs. 1): User-friendly platform (Ufp) • Certificate of completion (Coc) • Download of course material (Docm) • Own profile and account page (Opaap) • Quizzes and exercises (Qae) • Interactive quizzes and exercises (Iqae) • Comment function (Cf) • Personal tutor (Pt) • User manual for the platform (Umftp) • Videos (V) • Photos/Graphics (PG) • Text (T). A satisfaction index (CS) and dissatisfaction index (CD) was calculated (Eq. (1) and Eq. (2)). Each feature has a pair of CS and CD values. C(k) denotes the frequency of each Kano category.

$$CS_f = \frac{C(A)+C(O)}{C(B)+C(O)+C(A)+C(I)} \tag{1}$$

$$CD_f = \frac{C(B)+C(O)}{C(B)+C(O)+C(A)+C(I)} \cdot -1 \tag{2}$$

$$C(k) = \sum_{i=1}^{n} [f_i = k], \tag{3}$$

where k ∈ {A,B,O,I}, and f is the feature variable of each of the 12 Kano features f ∈ {Ufp, Coc, Docm, Opaap, Qae, Iqae, Cf, Pt, Umftp, V, PG, T}. The two equations allow a cost-benefit tradeoff of gaining satisfaction and preventing dissatisfaction (Berger et al., 1993; Dominici and Palumbo, 2013).

Values for CS range from 0 to 1 and higher numbers indicating higher satisfaction, CD ranges from −1 to 0 and lower numbers indicating higher dissatisfaction. Values of CS and CD are plotted both for the short and long-term course in Fig. 1. The values are connected by a vector from ex-ante to ex-post. The length (magnitude) of the vectors is reflected by the line thickness. The two language models trained by Word2Vec resulted in the following list, word-embeddings (in bold) and context words (with similarity scores). Due to the small corpora sizes, words with a similarity scores larger than 0.2 are listed.

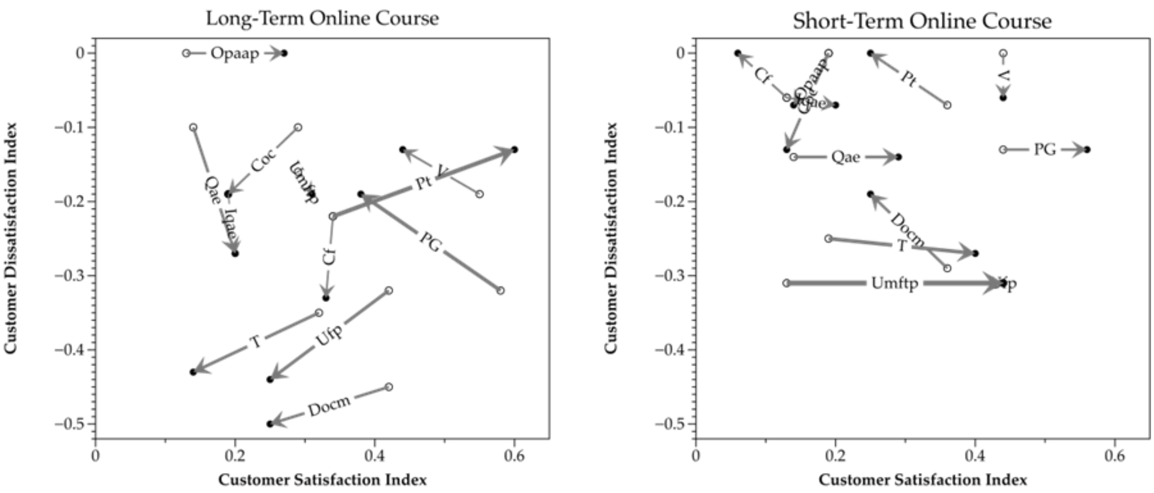
Short-term

- **need**: video (0.35), understand (0.29), test (0.26), system (0.23), content (0.22), teacher (0.21)
- **like**: professional (0.34), English (0.3), understand (0.28), photography (0.24), study (0.23), beautiful (0.21), restart (0.21), test (0.21)
- **dislike**: manual (0.39), complex (0.28), frustration (0.25), want (0.21)
- **wish**: explanation (0.37), professional (0.31), select (0.26), hint (0.24), understand (0.23), feature (0.22), video (0.2)
- **need − wish**: explain (0.22), weakness (0.22), system (0.22), content (0.21)
- **wish − need**: explanation (0.43), select (0.34), professional (0.27)

Long-term

- **need**: daily (0.41), assignment (0.31), platform (0.3), anonymously (0.29), concentrate (0.28), detailed (0.27), chat (0.25), motivative (0.23), textbook (0.23)
- **like**: professor (0.38), compare (0.38), presentation (0.37), work (0.34), clearly (0.31), lecture (0.3), discussion (0.27), platform (0.27), intuitive (0.26), motivative (0.23)
- **dislike**: participation (0.45), video (0.38), unmotivated (0.35), material (0.32), download (0.32), bad (0.3), vague (0.29), cheating (0.27), missing (0.26), concentrate (0.21)
- **wish**: time (0.38), download (0.36), miss (0.35), study (0.32), consult (0.31), find (0.31), anonymously (0.29), flexibility (0.27), q&a (0.26), concentrate (0.25), understanding (0.25), explain (0.24), motivative (0.24), detailed (0.24), data (0.23)
- **need − wish**: interaction (0.3), member (0.29), important (0.29), helpful (0.28), anonymously (0.25), contact (0.23)
- **wish − need**: time (0.41), post (0.34), question (0.33), learn (0.3), interesting (0.3), study (0.3),

proceed (0.3), description (0.28), data (0.27), example (0.26)



(a) Shift in long-term online course **(b) Shift in short-term online course**

Figure 1. Shifts of Satisfaction and Dissatisfaction Indices in online course of different duration.

From Fig. 1 and the context word list, it can be seen that the students from the short term online course struggled to simultaneously work through the course while figuring out how the site works. The feature *User manual for the platform* had the strongest increase in satisfaction index, while the language model for **dislike** lists *manual*, *complex* and *frustration* in combination with the word **wish** and *download*. Photos and graphics were regarded as positive in the freeform comments and a raise in user satisfaction index can also be observed. The Kano feature *Text* also experienced an increased satisfaction index. The authors suggest the ex-ante expectation to be more on multimedia, with an ex-post consumption experience that text helped in explaining concepts, supported by the word-vector arithmetic of **wish − need** and *explanation*. During the long-term online course, the feature *Personal tutor* experienced the largest increase in the satisfaction index (Fig. 1). Students felt the necessity to consult with the professor and tutor throughout the semester, supported by **like** and *professor*, *lecture*, and *discussion*. The largest decrease in satisfaction index could be observed for *Photos/Graphics*, *Text*, *User friendly platform*, and *Download of course material*. The first experienced a decrease in dissatisfaction index, whereas in the latter three, a decrease in CD could be observed. Features that simultaneously have a lower CS and CD indicate a drifting towards basic or one-dimensional requirements, that do not increase the satisfaction when implemented correctly, but suffer heightened dissatisfaction when missing. Some students wished for more detailed materials, as can be seen from **dislike** and *material*, *download* and *vague*.

5. Limitations

The content of the courses were different, with a bias towards visual content for the short-term introduction to photography course. Students that took part in this course were from a Japanese psychology college. Students from the long-term introduction to programming course were from a liberal arts college with a mix of Japanese and international (mostly Asian) students. A cultural component could also influence the results. The sample sizes, especially for the short-term online course is small and future work is planned to increase participants. However, while the absolute number of participants is in general terms statistically low, the homogeneous nature of Japanese students and the rarity of such data should also be taken into consideration. The limited availability of conducting comprehensive academic online courses should however be taken into account as well.

6. Conclusions

This study presents a combination of the Kano model with insights from language models to compare short-term and long-term online courses. The use of word-embeddings can be used to further quantify results obtained by the Kano method. The results show that needs differ depending on the format a course is offered. An ex-ante and ex-post analysis can deepen the understanding of expectation versus consumption experience. Short-term online courses could experience an increase in user satisfaction when including a short introduction video to explain how to navigate the platform. Long-term online courses benefit from a personal tutor and instructors can avoid dissatisfaction by continuously extending the online material, a well as investing in adequate web-hosting.

References

Aggarwal, C. C. (2018). Neural Networks and Deep Learning. Springer International Publishing AG.

Altszyler, E., Ribeiro, S., Sigman, M., and Fernandez Slezak, D. (2017). The interpretation of dream meaning: Resolving ambiguity using latent semantic analysis in a small corpus of text. Consciousness and Cognition, 56:178–187.

Berger, C., Blauth, R. E., and Boger, D. (1993). Kano's methods for understanding customer-defined quality. Center for Quality Management Journal, 2:3–36.

Chaudha, A., Jain, R., Singh, A. R., and Mishra, P. K. (2011). Integration of kano's model into quality function deployment (qfd). The International Journal of Advanced Manufacturing Technology, 53(5):689–698.

Di Gennaro, G., Buonanno, A., and Palmieri, F. A. N. (2021). Considerations about learning Word2Vec. The Journal of Supercomputing.

Dominici, G. and Palumbo, F. (2013). How to build an e-learning product: Factors for student/customer satisfaction. Business Horizons, 56(1):87–96.

Gittens, A., Achlioptas, D., and Mahoney, M. W. (2017). Skip-gram – Zipf + uniform = vector additivity. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 69–76, Vancouver, Canada. Association for Computational Linguistics.

Kano, N., Seraku, N., Takahashi, F., and ichi Tsuji, S. (1984). Attractive quality and must-be quality. Journal of The Japanese Society for Quality Control, 14(2):147–156.

Kazama, M., Sugimoto, M., Hosokawa, C., Matsushima, K., Varshney, L. R., and Ishikawa, Y. (2018). A neural network system for transformation of regional cuisine style. Frontiers in ICT, 5.

Learndash, (2022). Learndash.com

Nissim, M., van Noord, R., and van der Goot, R. (2020). Fair Is Better than Sensational: Man Is to Doctor as Woman Is to Doctor. Computational Linguistics, 46(2):487–497.

Parasuraman, A., Zeithaml, V. A., and Berry, L. (1988). Servqual: A multiple-item scale for measuring consumer perceptions of service quality. Journal of Retailing, 64:12–40.

Parasuraman, A., Zeithaml, V. A., and Malhotra, A. (2005). E-s-qual: A multiple-item scale for assessing electronic service quality. Journal of Service Research, 7(3):213–233.

Qi, P., Zhang, Y., Zhang, Y., Bolton, J., and Manning, C. D. (2020). Stanza: A Python natural language processing toolkit for many human languages. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations.

Quesenberry, K. A., Saewitz, D., and Kantrowitz, S. (2014). Blogging in the classroom: Using wordpress blogs with buddypress plugin as a learning tool. Journal of Advertising Education, 18(2):5–17.

Rehurek, R. and Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora. In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, pages 45–50, Valletta, Malta.

Scott, A. D. (2012). WordPress for Education. Packt Publishing.

Selvi, K. (2010). Motivating factors in online courses. Procedia - Social and Behavioral Sciences, 2(2):819–824. Innovation and Creativity in Education.

Sun, P.-C., Tsai, R. J., Finger, G., Chen, Y.-Y., and Yeh, D. (2008). What drives a successful e-learning? an empirical investigation of the critical factors influencing learner satisfaction. Computers & Education, 50(4):1183–1202.

Wang, Y.-S., Bauk, S., Scepanovic, S., and Kopp, M. (2014). Estimating students' satisfaction with web based learning system in blended learning environment. Education Research International, 2014.

Wang, Y.-S., Wang, H.-Y., and Shee, D. Y. (2007). Measuring e-learning systems success in an organizational context: Scale development and validation. Computers in Human Behavior, 23(4):1792–1808.

Wordpress, (2022). Wordpress.org

Zhu, Y., Yan, E., and Wang, F. (2017). Semantic relatedness and similarity of biomedical terms: examining the effects of recency, size, and section of biomedical publications on the performance of word2vec. BMC Medical Informatics and Decision Making, 17(1):95.

Zou, J. and Schiebinger, L. (2018). Ai can be sexist and racist—it's time to make it fair. Nature, 559(7714):324–326.