論文題目

The retention problem in MOOC: Why do learners not complete their courses?

応用情報科学研究科応用情報科学専攻
高信頼情報科学コース・高信頼情報科学領域
IM19I501　Wang Lei

MOOC (Massive Open Online Course) used to attract a great deal of attention after its advent in 2012. It was expected to provide access of high quality, college level courses to anyone, anywhere, any time. Some researchers claimed that MOOC even had the potential to reform higher education.

However, it appeared that MOOC was not a salvation for the problems of higher education. MOOC had its own problems, too. The retention problem was one of those problems. Although so many learners enrolled in courses, only a very small proportion continued learning and completed these courses.

Many researchers explored features affecting the retention in MOOC. Some researchers provided some possible strategies to improve the retention rate. However, most of the researchers did not consider the learner's intention. A few researchers analyzed data of those who reported to intend to complete. However, the intention to complete was captured only in the pre-course survey. In order to get a better understanding of this retention problem in MOOC, we also need to consider the students' actual learning activities.

Our research objective in this study is to explore the patterns of retention in MOOC and identify features affecting retention in MOOC. To achieve this, we addressed the following research questions: (a) What are the retention rates of different subpopulations in MOOC? (b) Do course related features affect retention in MOOC? (c) Do demographic features affect retention in MOOC? (d) Do learners' activity related features affect retention in MOOC?

This study is based on statistical analysis and machine learning on two publicly available dataset from edX. For those who intended to complete a course, the certification rate ranges from 13% to 100%, with a mean of 32% and standard deviation of 19.67%. The features affecting certification rate include course duration, number of chapters, country, education, gender, region, total number of events, active days, number of videos played, proportion of chapters accessed, proportion of user's lifetime in course duration. Among these features, active days and proportion of chapters accessed have a strong correlation with certification.

Based on the exploration above, we suggest that MOOC should foster more

active participation. Instructors could work on getting a more appropriate course duration.

There are some limitations with this study, we list them below as well as provide suggestions for additional research:

(a) In this research we used secondary data. There are chances that we may interpret any feature in a slightly different way with the original intention.

(b) The datasets do not collect the information for those who intended to complete the course. We have to define this with our own conditions. The true cohort of intended to complete may not be well represented. However, in future work, this can be achieved in a few ways. One is combining pre-course survey and learners' learning activities to better capture the learners' intention to complete. We could also use an unsupervised machine learning algorithm to clustering the target cohort.

(c) There are some features that we cannot tell for sure whether they are significantly correlated to retention or not, these features would need more investigation in the future.

The findings of this research enable the MOOC instructors and platforms to narrow down the target of intervention. The resources can be used more efficiently. The features identified could serve as metrics for intervention.

Key words - MOOC, Massive Open Online Course, retention, intention to complete, Google Trends, Gartner's Hype Cycle

The retention problem in MOOC: Why do learners not complete their courses?

氏名　王磊

本調査の目的は、MOOC（Massive Open Online Course）の保持のパターンを調査し、保持に影響を与える因子を特定することである。edX から公開されている 2 つのデータセットの統計分析と機械学習に基づいている。データ分析を通じて、MOOC のすべての登録者の認証率は誤解を招く傾向があることで、学習者の意図に照らして、MOOC での保持を調査する必要があると言える。コースを修了する予定の人の場合、認定率は 13%～100％の範囲で、平均は 32％、標準偏差は 19.67%である。コース期間、イベントの総数、アクティブな日数、再生されたビデオの数、アクセスされたチャプターの割合、コース期間中のユーザーの存続期間の割合など、認定率に影響を与える因子がある。これらの因子の中で、アクティブな日数とアクセスされるチャプターの割合は、認定と強い相関関係があることが明らかとなった。

学習活動に関する想定に基づいて、コースを修了する予定の人を特定するのはあまり正確ではないかもしれない。後のコースでは、学習者の意図を特定するためのコース前調査を紹介している。学習者が報告した意図は、彼らの活動データと組み合わせることができる。これは、学習者の完了意向をよりよく理解するのに役立つと考えられる。本調査の結果により、MOOC のインストラクターとプラットフォームは介入の対象を絞り込むことで、リソースをより効率的に使用できる。さらに、識別された因子は介入の指標として役立つ可能性があることが明らかとなった。

キーワード：MOOC、Massive Open Online Course、保持、完了意向、Google トレンド、Gartner のハイプサイクル

提出年月日　令和 3 年 02 月 22 日

指導教員名　ダニー・フェルナンデス教授