

ABSTRACT

Students' procrastination negatively impacts academic performance. To address this challenge, student-to-student interaction can be used to promote engagement, such as group assignments. However, research on procrastination in group tasks versus individual tasks is limited. Especially in large online classes where the number of instructors is limited to monitor all student behavior. Recent studies analyzed Learning Management System (LMS) log data to identify large number of students behavior. Sequence analysis was used to group students based on quantified similarity between sequences conducted using sequence alignment technique. However, no research compared alignment-based dissimilarity measures to understand specific aspects of learning sequences. Because the dissimilarity measure affects student clustering results, a comparative study is needed to choose the best measure and gain appropriate insights from student clustering.

This study compared three measures, namely Optimal Matching, OMspell, and Hamming, to analyze the timing aspect in learning sequences. OM was chosen because it has been widely used in previous studies analyzing learning sequences. OMspell was chosen because it was used in previous studies that analyzed procrastination behavior. Hamming was selected based on previous study in sociological research that found Hamming is the most suitable method for observing timing aspect of sequence.

In this study, we found that Hamming with a customized substitution matrix is optimum for identify timing aspect in learning sequence (Pseudo $R^2 = 0.20$, Silhouette width = 0.34, 85% correct differentiation in the sequence simulation study). The study found no difference in number of student procrastination between individual and group tasks. Interestingly, students who did the reflection task in the group-based wiki showed an increasing access pattern compared to those who worked on individual tasks. A comparative study could serve as a guide for learning analytics researchers in selecting the most suitable measure to identify a specific aspect of the learning sequence, such as timing aspect in procrastination behavior.

Keywords: Sequence analysis, dissimilarity measures, sequence alignment, learning analytics, clustering, learning management system

INTISARI

Perilaku *procrastination* pada siswa berdampak negatif terhadap performa akademiknya. Untuk mengatasi tantangan ini, interaksi siswa-ke-siswa dapat digunakan untuk mendorong keterlibatan siswa, seperti dengan pemberian tugas kelompok. Namun, penelitian tentang perilaku *procrastination* antara siswa yang diberikan tugas kelompok dibandingkan dengan siswa yang diberikan tugas individu masih terbatas. Terutama pada kelas daring dengan jumlah peserta yang besar dan jumlah instruktur yang terbatas sehingga tidak dapat memantau seluruh perilaku siswa. Studi terbaru melakukan analisis data log pada *learning management system* (LMS) untuk mengidentifikasi perilaku siswa dalam jumlah besar. Analisis *sequences* digunakan untuk mengelompokkan siswa berdasarkan kemiripan antar *sequence* yang dilakukan dengan teknik *sequence alignment*. Namun, belum ada penelitian yang membandingkan ukuran ketidaksamaan berbasis *alignment* untuk memahami aspek spesifik dari *sequences* aktivitas pembelajaran. Karena ukuran ketidaksamaan mempengaruhi hasil pengelompokan siswa, maka diperlukan studi komparatif untuk memilih ukuran ketidaksamaan yang terbaik sehingga dapat memberikan informasi tentang siswa yang sesuai.

Penelitian ini membandingkan tiga ukuran ketidaksamaan yaitu Optimal Matching, OMspell, dan Hamming untuk menganalisis aspek *timing* dalam *sequences* aktivitas pembelajaran. OM dipilih karena telah banyak digunakan pada penelitian-penelitian yang menganalisis *sequence* aktivitas pembelajaran. OMspell dipilih karena pernah digunakan pada penelitian sebelumnya dengan kasus serupa, yaitu menganalisis perilaku *procrastination*. Hamming dipilih berdasarkan penelitian sebelumnya di bidang sosiologi yang menemukan bahwa Hamming adalah metode yang paling cocok untuk mengamati aspek *timing* dari suatu *sequence*.

Dalam penelitian ini, kami menemukan bahwa Hamming dengan *customized matrix* sebagai *substitution cost* adalah ukuran ketidaksamaan yang paling optimal untuk mengidentifikasi aspek *timing* dalam *sequence* aktivitas pembelajaran (Pseudo $R^2 = 0,20$, Silhouette Width = 0,34, dan 85% dapat membedakan *sequence* dengan benar pada *simulation study*). Penelitian ini tidak menemukan perbedaan jumlah perilaku *procrastination* siswa antara siswa yang mengerjakan tugas individu dan kelompok. Menariknya, siswa yang mengerjakan tugas refleksi berbasis kelompok di wiki menunjukkan pola akses yang meningkat dibandingkan siswa yang mengerjakan tugas



individu. Studi komparatif ini dapat menjadi panduan bagi peneliti di bidang *learning analytics* dalam memilih ukuran ketidaksamaan yang paling sesuai untuk mengidentifikasi aspek spesifik dari *sequence* aktivitas pembelajaran, seperti aspek *timing* dalam perilaku *procrastination*.

Keywords: Sequence analysis, dissimilarity measures, sequence alignment, learning analytics, clustering, learning management system