COMPLEX NETWORKS OF HUMAN-WEB INTERACTIONS

Peter Geczy, Noriaki Izumi, Shotaro Akaho, Koiti Hasida
The National Institute of Advanced Industrial Science and Technology
Tokyo and Tsukuba, Japan

Abstract
We analyzed human browsing behavior on a large-scale web-based system. Human-web interaction sequences were temporally segmented into blocks encompassing elemental and compound browsing tasks. Network representation of the human browsing behavior resembles a complex network. The complex network traversal topology has a small number of hubs. The hubs contract and disperse navigational pathways. Underlying long tail attributes of complex networks coincide with broad user population and diminish in behaviorally focused user groups.

1. Introduction
Web is becoming increasingly interactive environment. Analytics and models of human-web interactions are rapidly gaining eminent position in web research and data mining. Exploring how we interact and navigate in web environments is indispensable for designing efficient novel personalized web systems [1].

Temporal dynamics of human-web interactions are characterized by periods of activity followed by longer periods of inactivity [2]. This enables segmentation of human-web interactions into parts representing tasks of various complexities. Larger temporal segments indicate tasks that are more complex. They are divided into smaller temporal segments outlining the elemental tasks. The tasks contain the sequences of navigational transitions and interactions with the web environment. The segmentation and task decomposition permits effective analysis of human-web interactions—both at elemental and higher order abstraction levels.

Analysis of interaction sequences revealed numerous findings. Users have a strong tendency to form elemental and complex browsing patterns. They frequently access relatively small number of navigation points. Topological characteristics of human navigation and interactions correspond to a complex network.

2. Approach
The human-web interaction sequences are partitioned according to the detected delays. This leads to the segmentation of long click stream sequences into the finer elements: sessions and subsequences [3]. Larger sequences denote sessions. The sessions are further divided into subsequences. Sessions delineate greater tasks that are accomplished via sub-tasks represented by the subsequences.
Segmented click streams enable identification of pertinent navigation points where users initiate their tasks and the resources they target. The initial points of sessions and subsequences are the starters. The target points are the attractors. Starter-attractor pairs constitute the abstractions of the elemental browsing patterns, thus eliminating the multitudes of navigational pathways between the given starter and attractor. Connecting navigation elements of subsequences are the attractor-starter pairs. They enable us to observe how more complex browsing patterns are being formed—as interconnected sequences of elemental patterns.

3. Case Study and Findings

Case Study: A large-scale organizational web portal has been analyzed for human-web interactions. The portal incorporates six servers in a load-balanced configuration. It provides extensive range of services (>800) and resources (>3 million) vital to the organization. The institution has a number of branches at various locations throughout the country, thus certain services are decentralized. Traffic was substantial and so was the data volume. The analyzed one-year web log set contained over 300 million records.

Findings: We analyzed the navigation point characteristics together with higher-order browsing abstractions. The analysis exposed several pertinent features of human-web interactions. Access statistics of navigation points, higher order abstractions, and interconnecting elements of more sophisticated interactions displayed significant long tails. Topology of human behavioral space clearly corresponds to the complex network. The in and out degrees of nodes (whether navigation points or higher order abstractions) have distinguishing long tails. We have derived a novel log-polynomial-exponential distribution that accurately models it.

Significant finding was that the underlying characteristics generally hold for large user populations having mixed behavioral attributes. They do not hold generally for behaviorally similar micro-groups of users. This implies that the large user populations may exhibit typical long tail features, but the behaviorally similar user group may have completely different characteristics.

4. Conclusions

Significant long tail characteristics have been detected in all analyzed aspects of the user browsing behavior. A network representation of human navigation matches a complex network. It has been observed that the complex network characteristics coincide with large and diverse user populations. They may no longer be observable in narrow and behaviorally specific user groups. Hence, the behavioral models derived from large and diverse user populations may not be suitable for focused user groups.

References