Mapping the Knowledge Space Exploiting Unassisted Machine Learning Tools

**Author:**Florenta Teodoridis, Jino Lu, Jeffrey L. Furman

Understanding factors affecting the direction of innovation is a central aim of research in the economics of innovation. Progress on this topic has been inhibited by difficulties in measuring distance and movement in knowledge space. We describe a methodology that infers the mapping of the knowledge landscape based on text documents. The approach is based on an unassisted machine learning technique, Hierarchical Dirichlet Process (HDP), which flexibly identifies patterns in text corpora. The resulting mapping of the knowledge landscape enables calculations of distance and movement, measures that are valuable in several contexts for research in innovation. We benchmark and demonstrate the benefits of this approach in the context of 44 years of USPTO data.

**Url:**<https://www.nber.org/papers/w30603>