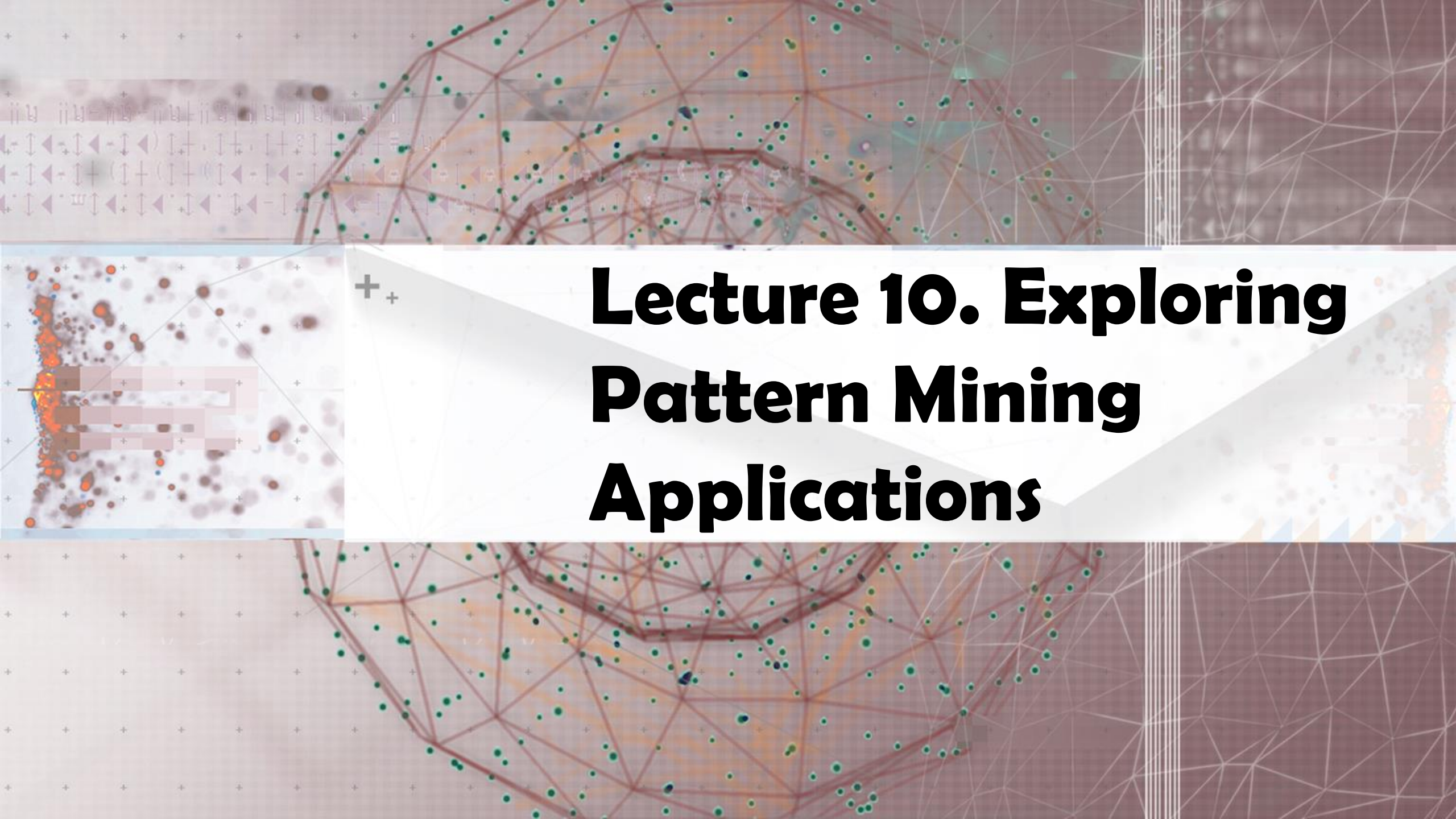

Lecture 10. Exploring Pattern Mining Applications

Lecture 10. Exploring Pattern Mining Applications

- Frequent Pattern Mining for Text Data—Phrase Mining and Topic Modeling
  - Strategy 1: Simultaneously Inferring Phrases and Topics
  - Strategy 2: Post Topic Modeling Phrase Construction
  - Strategy 3: First Phrase Mining then Topic Modeling (ToPMine)

Thanks to Ahmed El-Kishky@UIUC, Chi Wang@MSR and Marina Danilevsky@IBM for their contributions

Note: Only one application is discussed here—Other applications will be discussed in Lecture 11 or have already been scattered in other Lectures

Session 1. Frequent Pattern Mining for Text Data

Frequent Pattern Mining for Text Data: Phrase Mining and Topic Modeling

- Motivation: Unigrams (single words) can be difficult to interpret
- Ex.: The topic that represents the area of Machine Learning

learning
reinforcement
support
machine
vector
selection
feature
random
:

versus

learning
support vector machines
reinforcement learning
feature selection
conditional random fields
classification
decision trees
:

Various Strategies: Phrase-Based Topic Modeling

- Strategy 1: Generate bag-of-words → generate sequence of tokens
  - Bigram topical model [Wallach'06], topical n-gram model [Wang, et al.'07], phrase discovering topic model [Lindsey, et al.'12]
- Strategy 2: Post bag-of-words model inference, visualize topics with n-grams
  - Label topic [Mei et al.'07], TurboTopic [Blei & Lafferty'09], KERT [Danilevsky, et al.'14]
- Strategy 3: Prior bag-of-words model inference, mine phrases and impose on the bag-of-words model
  - ToPMine [El-kishky, et al.'15]