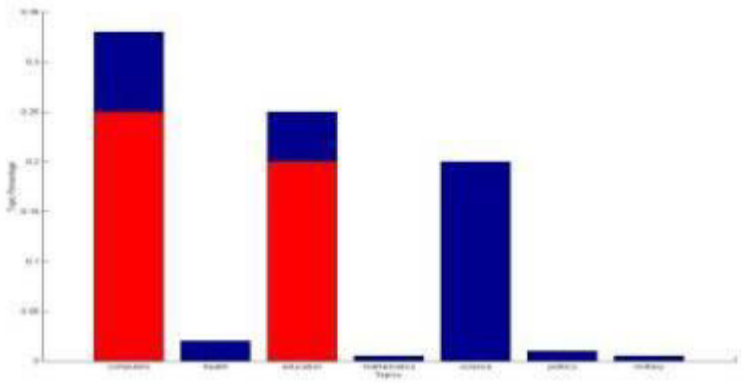
Topic Disparity as a DRL pre-measure

Laboratory for Analytic Sciences

NC STATE UNIVERSITY

General Description

- Documents, topics and words form an information space quantified by distributions
- A measure of disparity of topics-ideas in a corpus of documents may be evaluated relative to *a question*
- Documents are modeled as a weighted distribution of topics:



General Description

- Their distributions are estimated and a divergence (Jensen-Renyi) is evaluated
- Building on existing topic discovery tools, account for the adequate topic mixture of a certain question
- A disparity outcome will reflect a DRL measure: e.g. low disparity implies good DRL

Possible scenario

- Suppose that an analyst has the news articles for each day.
- She wants to answer the question:

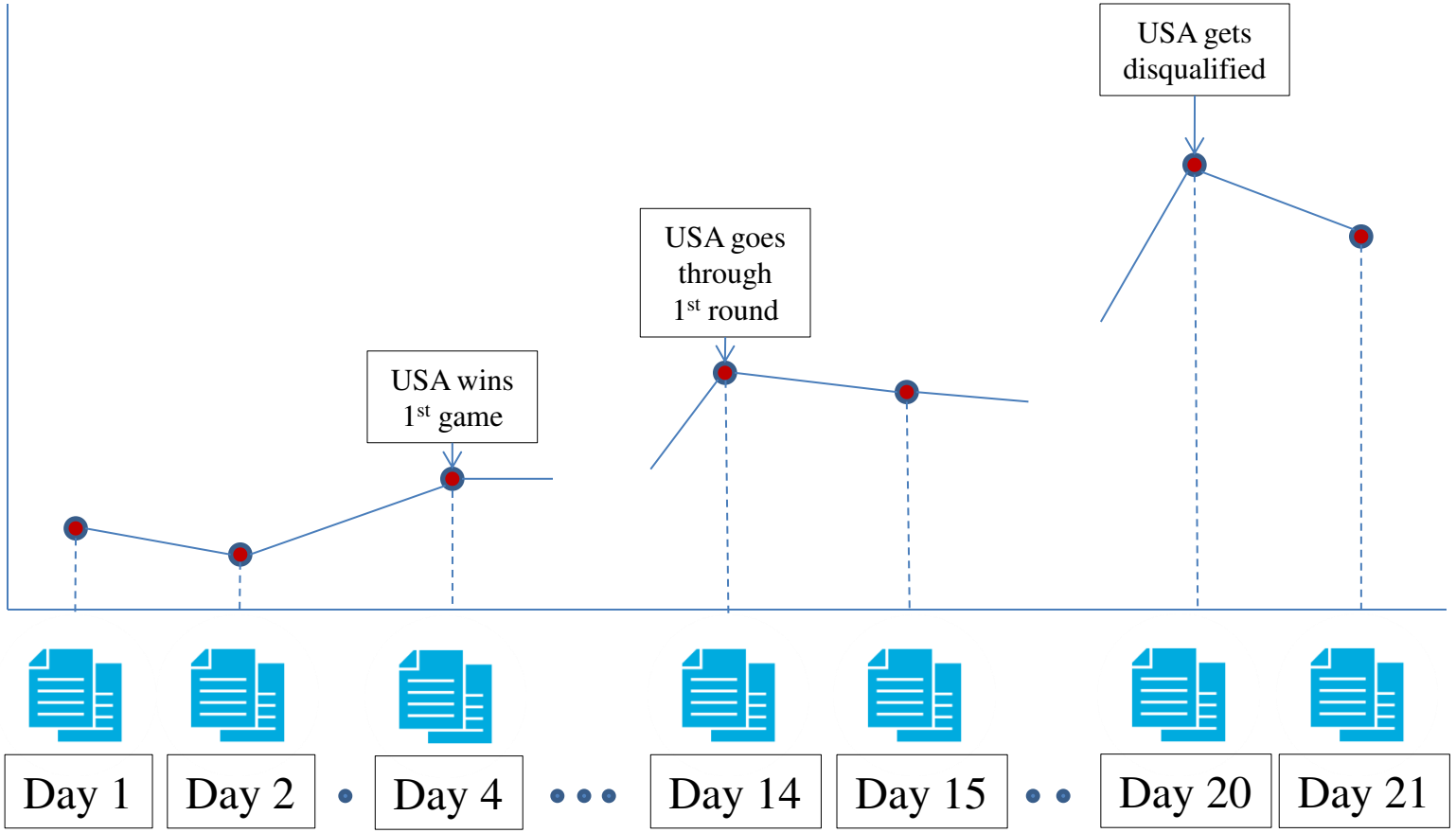
  “Will USA win the world cup?”

- She does not want to look at all of the articles.
- Use the disparity measure to help her choose which day's articles she should read to form an opinion.
- Use the disparity measure to have a measure of the “goodness” of data for any day chosen.

Problem Statement

- The data is: A set of documents-news about the world cup using online sports sites. These corpi are viewed as a time series (daily).
- The question is: “Is USA going to win the world cup?”.
- The quantifier used is document disparity computed daily.

Example of Disparity



Advantages of Disparity

- First measure of relevance of an (information compendium) info-com to the question
- Improved accuracy over time, automated, customizable
- Easy implementation, almost online algorithms

Generalizing Disparity

- The idea of document disparity can be generalized to other modalities (images, data-tables, mixed)
- If the data is an image for example, it is characterized by several attributes which are themselves random (e.g. SNR, Scene-like image characterizer, etc....)
- These on the other hand are chosen with a specific question in mind
- The image integrity/coherence requires a certain amount of information of the various features