How Mergeflow Uses Machine Learning

04-2019

Machine Learning @Mergeflow

Mergeflow uses machine learning technologies for various aspects of data collection, structuring, and analysis.

1

Collect

Smart Data Collection

Based on the structure of a website and other criteria, Mergeflow learns which pages of a website to crawl.

2

Structure

Signal vs. Noise

Websites are structured differently from R&D papers; patents are different from market news, etc.. For each source, Mergeflow learns how to identify relevant parts of a document vs. noise (e.g. teasers).

3

Analyze

People & Organizations

Machine learning helps Mergeflow find people, organizations, and other entities, including never-before-seen ones.

Financial Information

Mergeflow uses machine learning for discovering market data (segments, sizes, growth) and investment events from text.

Topics

Based on semantic models, Mergeflow assigns contents to topics. We use patent classes as topics.

4

User Interface

Machine-Assisted User Interface

In its user interface, Mergeflow suggests semantically related terms when you start searching. Mergeflow uses its collected data to learn which terms are related.

Some Recommended Readings

We think that a fact-based and de-hyped conversation about machine learning and related technologies is crucial to the success of any operation in this area. This is why we put together a small set of machine learning books, papers, and resources on the web that we have found relevant and interesting.

Books

Manning & Schütze (1999). Foundations of Statistical Natural Language Processing. Great classical textbook on statistical NLP.

Mitchell (1997). Machine Learning. Indispensable. You need this.

Papers

Lei, Barzilay, Jaakkola (2016). Rationalizing Neural Predictions. How does a neural network make decisions?

Mikolov, Chen, Corrado, Dean (2013). Efficient Estimation of Word Representations in Vector Space. Very interesting approach to representing word meanings.

Blogs & Websites

AI Index. https://aiindex.org. Good high-level overviews of the current state of AI.

Chris McCormick's Machine Learning Tutorials and Insights. http://mccormickml.com/. Excellent hands-on tutorials on machine learning.

Rodney Brooks' Blog. https://rodneybrooks.com/blog/. Infusing facts into the conversation about AI.

Andrej Karpathy's Blog. http://karpathy.github.io/2016/05/31/rl/. Good place to start on reinforcement learning.