Topic Modelling Algorithms in Information Retrieval and Filtering

Waterford Institute *of* Technology
INSTITIÚID TEICNEOLAÍOCHTA PHORT LÁIRGE

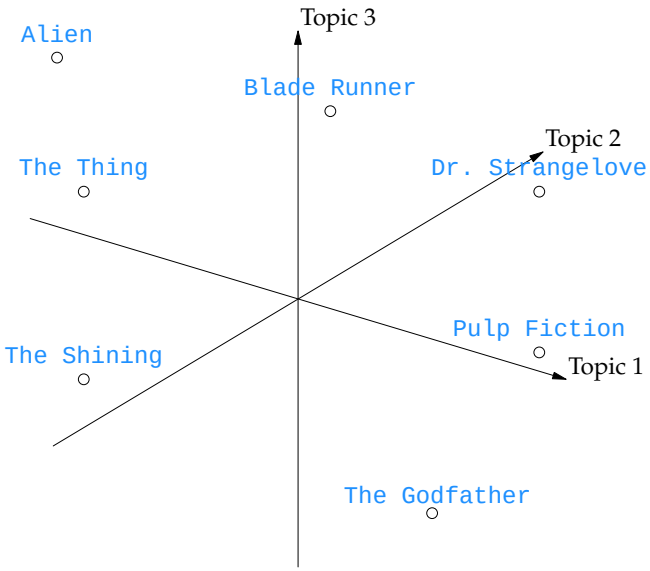
Jack Power — 20080169

BSc (Hons) in Applied Computing, Department of Computing and Mathematics, School of Science and Computing WIT

Topic modelling algorithms use mathematical techniques to express entries in a dataset in terms of the topics present.

It is a technique used in machine learning and natural language processing to extract topics latent to the data, exposing hidden structures which can then be used in searching and filtering.

These topics are learned by the model during training, linking items which share common topics. The resulting vector space can be used in applications such as movie recommenders, based on the topics a user seems to like.



Experimentation

The mechanisms behind different algorithms vary significantly, yielding models which perform differently and possess different properties.

The goal is to explore these algorithms using Amazon SageMaker as an experimental testbed, allowing each model to be trained, tuned, evaluated and compared in a robust and repeatable manner.