

Intro to data mining

There are an incredible amount of **choices** in our modern world! Amazon has millions of books available, Netflix, Hulu, and AmazonPrime have thousands of TV shows and movies between them. There are millions of potential products we can purchase.

Finding relevant stuff

What movie should I watch? What service should I use? What should I eat for lunch?

The problem is sorting out the noise and finding which items will *matter* to a user. It's become **exponentially** more difficult to find relevant content in our world of possibilities.

How to find stuff?

In a small town, long ago, our *friends* used to help us find stuff we would like.

Experts came into our decision process, as popular products were evaluated and reviewed. Expert recommendations and reviews really meant, and still do mean, something.

Also, we use *the thing itself* to aide us in our quest to find useful stuff. We remember what we like, and we use this as a basis for our future decisions.

This stuff is all still used, although we needed some **computational help** to take them into the 21st century.

More than just stuff

If we can predict what people like, and what their values are, we can appeal directly to them.

The government **uses private companies to keep data on us**. Companies have data on most of us, including where we live, what we look like, and the things we buy.

Data mining extends what we normally do

Data mining is focused on *finding patterns in data*.

People can do this, on the smaller scale, but with tremendous amounts of data and options, computation becomes necessary.

Tera-mining isn't something from Starcraft II

First of all, what a fantastic reference.

Datasets have **grown tremendously** in the past decade. It is not uncommon to be doing data mining on **petabytes** of data.

Tera-mining is a word to denote data mining on huge datasets.