Embeddings: Motivation From Collaborative Filtering

ated Time: 10 minutes

Collaborative filtering is the task of making predictions about the interests of a user based on interests of many other users. As an example, let's look at the task of movie recommendation. Suppose we have 1,000,000 users, and a list of the movies each user has watched (from a catalog of 500,000 movies). Our goal is to recommend movies to users.

To solve this problem some method is needed to determine which movies are similar to each other. We can achieve this goal by embedding the movies into a low-dimensional space created such that similar movies are nearby.

Before describing how we can learn the embedding, we first explore the type of qualities we want the embedding to have, and how we will represent the training data for learning the embedding.
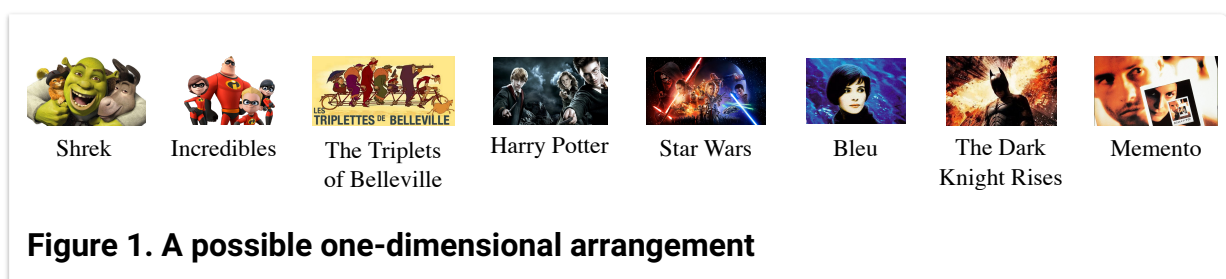
Arrange Movies on a One-Dimensional Number Line

To help develop intuition about embeddings, on a piece of paper, try to arrange the following movies on a one-dimensional number line so that the movies nearest each other are the most closely related:

Movie	Rating (https://wikipedia.org/wiki/Motion_Picture_Association_of_Ame
Bleu (http://www.imdb.com/title/tt0108394/)	R

Movie	Rating (https://wikipedia.org/wiki/Motion_Picture_Association_of_Ame
The Dark Knight Rises (http://www.imdb.com/title/tt1345836)	PG-13
Harry Potter and the Sorcerer's Stone (http://www.imdb.com/title/tt0241527/)	PG
The Incredibles (http://www.imdb.com/title/tt0317705/)	PG
Shrek (http://www.imdb.com/title/tt0126029/)	PG
Star Wars (http://www.imdb.com/title/tt0076759/)	PG
The Triplets of Belleville (http://www.imdb.com/title/tt0286244)	PG-13
Memento (http://www.imdb.com/title/tt0209144/)	R

+ **Click the plus icon for one possible (highly imperfect) solution.**

Shrek   Incredibles   The Triplets of Belleville   Harry Potter   Star Wars   Bleu   The Dark Knight Rises   Memento

Figure 1. A possible one-dimensional arrangement

While this embedding does help capture how much the movie is geared towards children versus adults, there are many more aspects of a movie that one would want to capture when making recommendations. Let's take this example one step further, adding a second embedding dimension.

Arrange Movies in a Two-Dimensional Space

Try the same exercise as before, but this time arrange the same movies in a two-dimensional space.

+ **Click the plus icon for another possible solution.**

Figure 2. A possible two-dimensional arrangement

With this two-dimensional embedding we define a distance between movies such that movies are nearby (and thus inferred to be similar) if they are both alike in the extent to which they are geared towards children versus adults, as well as the extent to which they are blockbuster movies versus arthouse movies. These, of course, are just two of many characteristics of movies that might be important.

More generally, what we've done is mapped these movies into an **embedding space**, where each word is described by a two-dimensional set of coordinates. For example, in this space, "Shrek" maps to (-1.0, 0.95) and "Bleu" maps to (0.65, -0.2). In general, when learning a d-dimensional embedding, each movie is represented by d real-valued numbers, each one giving the coordinate in one dimension.

In this example, we have given a name to each dimension. When learning embeddings, the individual dimensions are not learned with names. Sometimes, we can look at the embeddings and assign semantic meanings to the dimensions, and other times we cannot. Often, each such dimension is called a **latent dimension**, as it represents a feature that is not explicit in the data but rather inferred from it.

Ultimately, it is the distances between movies in the embedding space that are meaningful, rather than a single movie's values along any given dimension.