

Content-Based Filtering & the Cosine Similarity Metric

Recommend movies that are ~~Similar~~ Similar

genres and similar years to the movies

each user liked

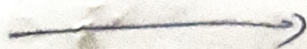
→ recommending items just based on the

attributes of those items, themselves
instead of trying to use aggregate user behavior data.

movie title



X

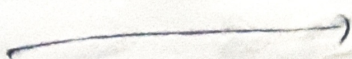


genres



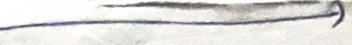
Adventure | Comedy | Fantasy

Y



Animation

Z



Comedy | Romance

P



Drama

→ There are totally 18 possible genres

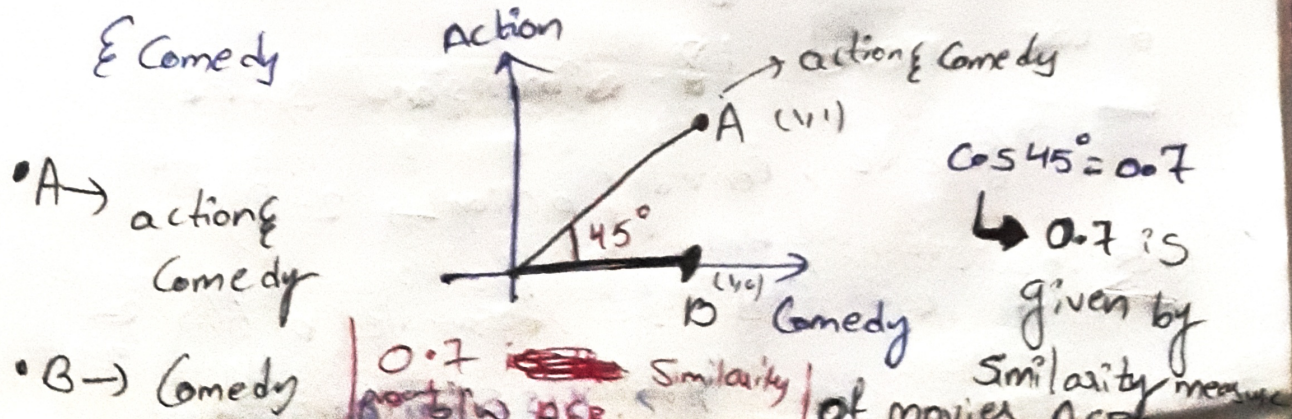
→ To identify the similarity measure ^{between genres} is

given by Cosine Similarity Metric



It's applicable to all sorts of other recommendation approaches too.

→ To understand Cosine Similarity Metric, let's take a we have only 200 possible genres action



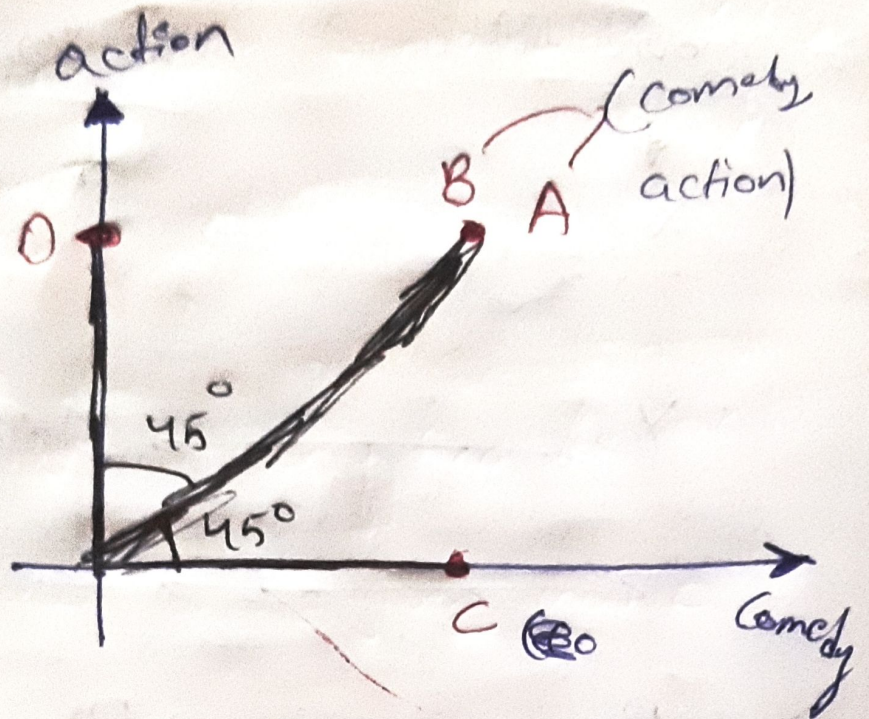
Suppose

$A \rightarrow \text{comedy, Action}$

$B \rightarrow \text{comedy, Action}$

$C \rightarrow \text{Comedy}$

$D \rightarrow \text{Action}$



$$\cos 90^\circ = 0, \quad \cos 45^\circ = 0.7, \quad \cos 0^\circ = 1$$



0 similarity

is present

blw movies

$C \notin D$

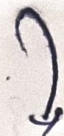


0.7 similarity

is present
blw movies

① $(A/B) \notin D$

② $(A/B) \notin C$



0 similarity

is present
blw movies

$A \notin B$