Advanced Concepts of Machine Learning: Representations

Kevin Trebing (i6192338)

November 20, 2018

1 Representation for example 1 (Movie theatre): Relational representation

In this example a relational representation is the most feasible, since a classification for being a good movie can be dependent on many different features. These features can be stored efficiently using multiple tables.

movie	directorName	genre	age-restriction	year	oscars
Interstellar	Christopher Nolan	Adventure	PG-13	2014	1
Inception	Christopher Nolan	Action	PG-13	2010	4
LotR: Return of the King	Peter Jackson	Action	PG-13	2003	11

actorName	oscars
Leonardo DiCaprio	1
Matthew McConaughey	1
Elijah Wood	0
Tom Hardy	0

movie	actor
Interstellar	Matthew McConaughey
Inception	Leonardo DiCaprio
Inception	Tom Hardy
LotR: Return of the King	Elijah Wood

directorName	oscars
Christopher Nolan	0
Peter Jackson	3

Hypothesis language:

- pos(M) :- movieInfo(M, DN, _, _, _, _), movieActor(M, AN), actor(AN, AO), director(DN, DO), sum(AO, DO) > 2

  In natural language: A movie is profitable if the amount of oscars of the director and at least one actor is bigger than 2.

- pos(M) :- movieInfo(M, _, 'comedy', 'R', Y, _, _), movieActor(M, 'Ryan Reynolds'), Y > 2012

  In natural language: A movie is profitable if it is R-rated and has 'Ryan Reynolds' as an actor and the movie came out later than 2012.

- pos(M) :- director(M, 'Christopher Nolan', _)

  This example should be self-explanatory.

This approach is also useful for updating the database. If an actor wins an oscar only the entry in the actor table needs to be updated once. In a multi-instance representation all entries of the actor need to be changed. Furthermore, since the cast of a movie can be mapped from one movie to multiple actors a

multi-instance approach would result in multiple table entries for one movie where only the actor name changes. This would be redundant and can be avoided using a relational representation.

2 Representation for example 4 (Spam or ham): attribute valued representation

For this example we use an attribute valued representation. For this we can use a table for multiple examples:

subject	knownSender	domain	class
supplements	no	other	spam
money	yes	bank	ham
other	yes	gmail	ham
money	no	other	spam
course	no	university	ham
supplements	no	university	ham
arabian prince	no	hotmail	ham

Possible hypthesis:

- ham :- spamorham(_, _, university)

  Meaning: We regard the email as non-spam when the sender domain is a university. (This actually seemed to be the case at my old university. We tried to get sent to the spam folder for a while. Nothing worked.)

- ham :- spamorham(_, yes, _, _)

  Meaning: If we know the sender (e.g. from the contact book) we regard the mail as non-spam.

- spam :- spamorham('arabian prince', no, _)

  Meaning: If the subject is about an Arabian prince and we do not know the sender, it is classified as spam.

We could also use an boolean representation, but then we would have to force all our attributes where we have more than two values into a boolean representation which could lead to a combinatorial explosion. If we reduce the granularity of the attributes to just boolean representations we could lose information.

I sadly do not know what a multi-instance representation would bring us here. But that could be due to the fact that I do not quite get what the major advantages over AV representation are. I also looked for explanations on the internet, but found nothing helpful (maybe I was too focused on this example).