Learning when you will eat unhealthy

using decision trees for longitudinal data

Gerasimos Spanakis[A], Gerhard Weiss[A], Bastiaan Boh[B], Vincent Kerkhofs[B], Lotte Lemmens[B], Anita Jansen[B], Anne Roefs[B]

[A] Department of Knowledge Engineering, Maastricht University
[B] Faculty of Psychology and Neuroscience, Maastricht University

Background & Introduction

ThinkSlim is an **iphone application** developed to collect real-life data from people and help them detect their unhealthy eating moments before they occur. The application makes use of **Ecological Momentary Assessment (EMA)** which provides us lots of data with a rich longitudinal hierarchical structure.

We combine a decision tree algorithm with the per-user structure of longitudinal data in order to achieve more accurate insights on the conditions that predict unhealthy food intake. The resulting algorithm is applied to data collected by our application and extracts useful rules which are used for providing users with feedback.

Exploratory Data Analysis & Data Preparation

User data is organized per day/time. Based on exploratory analysis statistics, data (numeric & free text) is discretized and the possible values for each attribute are shown below (along with any other categorical attribute) :

Attribute	Discretized values	Details
Craving	LOW, MID, HIGH	
Negative Emotions	NO, YES	sad, bored, stressed, angry
Positive Emotions	LOW, MID, HIGH	happy, relaxed
Location	Home, School, Traveling, Work, Social, Other	
Circumstances	ComputerRelated Eating HighLevelIn HighLevelOut LowLevel WatchingTV Reading Socializing Outdoors Working	Phone / Internet / Computer Eating / Non-social drinking Preparing food, cleaning, sanitary, etc Exercising, hobby, leisure, shopping, etc Relaxing, waiting, lying in bed, etc Studying, thinking, etc Having a drink, etc traveling, etc administration, work activities, etc
Time of day	morning, noon-afternoon, evening	
Weekend	NO, YES	
Specific Craving	N, H, U	Nothing, Healthy, Unhealthy
Specific Eating	N, H, U	Nothing, Healthy, Unhealthy

Each data point is used to predict whether the next data point (provided that they both occur on the same day) will be a healthy or an unhealthy eating moment.

user	Date/time	craving	Neg. emot.	Posit. emot.	Specific craving	Time of day	week end	circumstances	location	Specific eating
pp5	26/01/2015 23:57:15	LOW	NO	LOW	N	evening	NO	LowLevel	Home	N
pp5	27/01/2015 09:32:45	LOW	NO	HIGH	N	morning	NO	LowLevel	Home	N
pp5	27/01/2015 12:17:30	MID	NO	HIGH	H	noon-after	NO	ComputerRelated	Work	U
pp5	27/01/2015 14:43:32.	LOW	YES	MID	N	noon-after	NO	Work	Work	H

user	craving	Neg. emot.	Posit. emot.	Specific craving	Time of day	week end	circumstances	location	Specific eating	NextEating (Class)
pp5	LOW	NO	HIGH	N	morning	NO	LowLevel	Home	N	U
pp5	MID	NO	HIGH	H	noon-after	NO	ComputerRelated	Work	H	H
...										

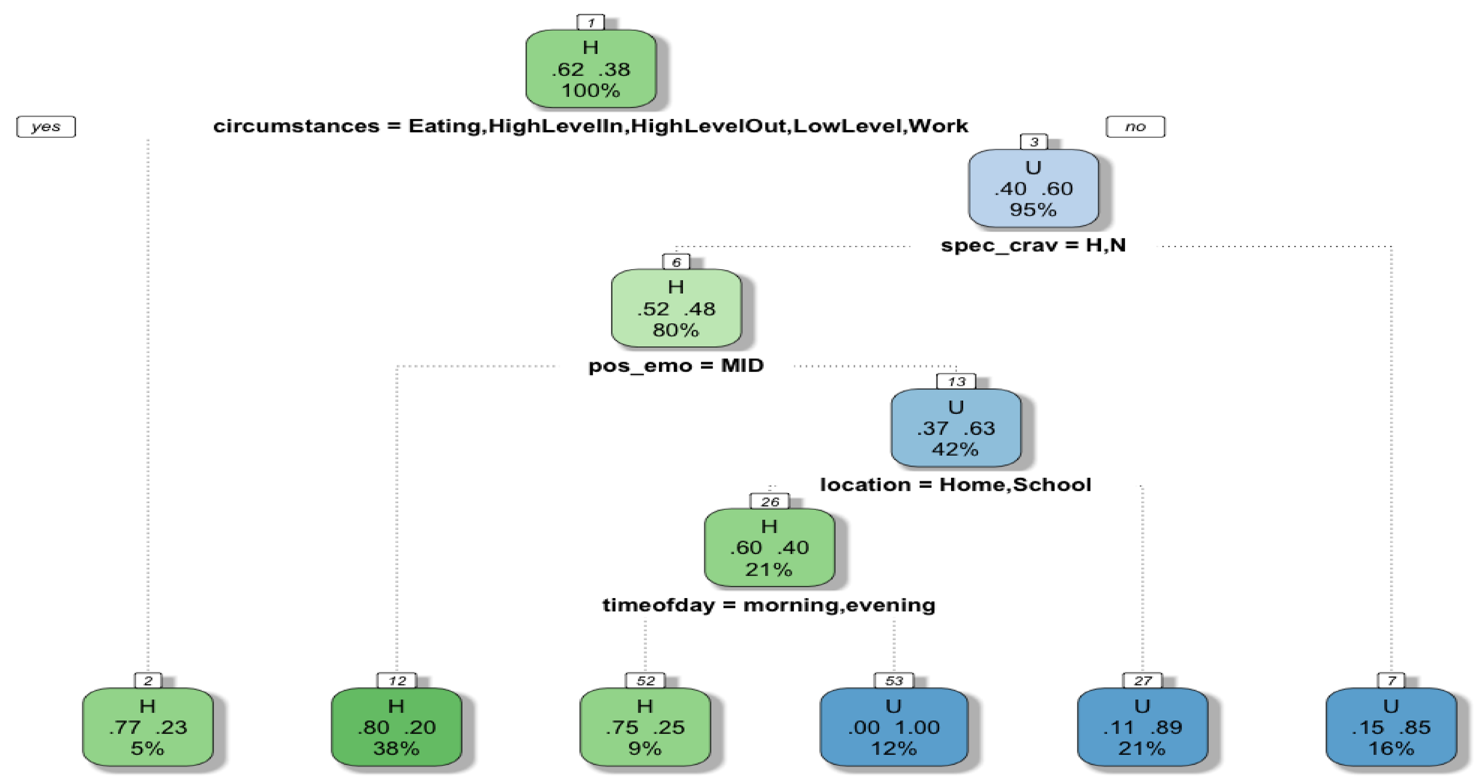
Building the decision tree using a top-to-bottom approach

Using the data points above as observations, we want to predict under which conditions (i.e. combinations of attributes) we are led to unhealthy eating (class variable, Y). In order to (recursively) build a decision tree, we need to select the "most important" attribute to "split" the data. This is done using two criteria.

First, the attribute with the largest Information Gain (IG) is selected. Then, if C is the dominant class (for each new node) we define Z_k=+ for every user k if the number of observations (in that node) with Y=C is greater or equal than the number with Y != C. Otherwise, Z_k=-. We form a contingency table with the 2^k patterns of Z as columns and the attribute splits as rows and compute the p-value of an independence test (e.g. Chi-Square). If the p-value is less than 0.05/|k|, then the associated variable is selected for splitting and we continue building the tree. If not, then the variable with the second best IG is selected and we repeat the same process.

Rule inference from the decision tree

Given the decision tree structure, we follow every path that leads from root to a leaf and infer one rule per leaf. On each node the split condition can be seen: If it is "true" ("yes") we take the left branch, otherwise we take the right branch.



For example, in the above sample tree, the rule corresponding to the red arrow is the following:

IF CIRCUMSTANCES={ComputerRelated,Outdoors,Reading,Socializing,Watching TV}
AND SPECIFIC_CRAVING={U}
→ NEXT_EATING={U}

Results

Given a sample of N=60 obese people, we extracted 65 significant rules (36 leading to healthy eating and 29 to unhealthy). Some examples of these rules (that lead to unhealthy eating) are the following:

IF SPECIFIC_CRAVING={H,N}
AND SPECIFIC_EATING={H,U}
AND CIRCUMSTANCES={Eating,HighLevelIn,LowLevel,Socializing,Watching TV}
AND TIMEOFDAY={noon-afternoon, evening}
→ NEXT_EATING={U}

IF SPECIFIC_CRAVING={U}
AND TIMEOFDAY={morning}
AND SPECIFIC_EATING={N}
AND LOCATION={Outdoors,School,Social}
→ NEXT_EATING={U}

IF SPECIFIC_CRAVING={U}
AND TIMEOFDAY={noon-afternoon, evening}
AND CIRCUMSTANCES={ComputerRelated,Eating,Socializing,Watching TV}
AND POSITIVE_EMOTIONS={MID,HIGH}
→ NEXT_EATING={U}

IF SPECIFIC_CRAVING={H,N}
AND SPECIFIC_EATING={N}
AND TIMEOFDAY={evening}
AND CRAVING={LOW,MID}
AND CIRCUMSTANCES={ComputerRelated,Reading,Watching TV,Work}
AND LOCATION={Home,Other,Work}
AND POSITIVE_EMOTIONS={LOW}
AND NEGATIVE_EMOTIONS={YES}
→ NEXT_EATING={U}

IF SPECIFIC_CRAVING={H,N}
AND SPECIFIC_EATING={N}
AND TIMEOFDAY={morning, noon-afternoon}
AND LOCATION={Outdoors, Traveling, Other}
AND POSITIVE_EMOTIONS={MID}
AND WEEKEND={YES}
→ NEXT_EATING={U}

Correspondence to:

Gerasimos Spanakis, PhD
jerry.spanakis@maastrichtuniversity.nl
http://dke.maastrichtuniversity.nl/jerry.spanakis/
www.thinkslim.nl

Department of Knowledge Engineering
T +31(0)43 38 83916

P.O. Box 616
6200 MD Maastricht, The Netherlands

STW
Nationaal Initiatief Hersenen & Cognitie
PHILIPS
Maastricht University Leading in Learning!